THE UNIVERSITY
of EDINBURGH

ESTABLISHMENT AND MAINTENANCE OF DNA METHYLATION IN ABSENCE OF THE CHROMATIN REMODELLER LSH

Roberta Ilaria Amendola

Thesis presented for the Degree of Doctor of Philosophy

The University of Edinburgh

2019

THE UNIVERSITY
of EDINBURGH

Declaration

I declare that this thesis was composed by myself. The research in this thesis is my own, unless otherwise stated and has not been submitted for any other degree or professional qualification.

Roberta Ilaria Amendola (June, 2018)

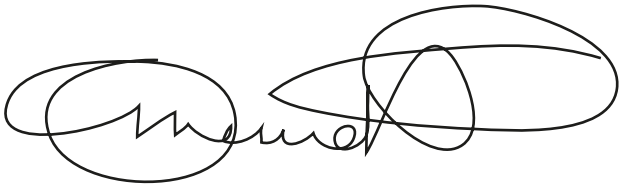


TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to take this space to thank all the people who supported me during these challenging years. First of all, I would like to thank Irina for giving me the position and for challenging me every day to become a better scientist. I would like to thank Philipp for adopting me, when the situation changed, and helping me out to complete my journey. It was hard for me being away from my family, but I made 2 new international ones. First, thanks to my lab family, Burak for being my wise advisor and a bro until the last moment, Natali for helping me surviving stressful moments and making every day easier, Christian for the arguments, Dani for being a great desk neighbour, Simon and Alina for contributing to making the lab gang richer and more fun.

My second family was my Edinburgh family. Thanks Pablo for being the best flatmate I could have asked for, I'll always miss the cooking and sofa nights; Elana for becoming a great wine friend and supporter; Giusy, for being the Italian friend I needed. Natalia, who I felt right to include here too, being my best friend and travelling companion out of the lab; Elena, for creating an amazing group of crazy and creative people; Stefano, for the board game and beer nights, but also for the help with science stuff; Maria, for introducing me to this amazing world that is the public engagement, being a fantastic mentor and a good friend. Thanks Rod, Cata, Gui, the knitting group, for taking part in all the crazy days and nights during these years! Thanks all for helping me not feeling lonely.

Finally, I'd like to thank my friend and family from home and I will do this in Italian. Grazie alla mia famiglia, mamma, papà, Claudio, per avermi supportata anche da così lontano e avermi fatto sentire sempre e comunque la vostra presenza. Grazie alle amiche di sempre, Marta e serena, per il costante supporto, i milioni di messaggi e le chiamate interminabili, che mi hanno aiutata a superare questi anni. Il mio cuore è, e sempre sarà, sparso nel mondo. My heart is, and always will be, spread around the planet.

Thanks. From the bottom of my heart.

ABSTRACT

DNA methylation at cytosine (5meC) is essential for central cellular mechanisms, such as genomic imprinting, tissue-specific gene expression, silencing of retroviral elements and inactivation of the X chromosome in female. Impairment of proteins that either deposit or bind the 5meC leads to human disease, such as imprinting disorders, Rett Syndrome, Immunodeficiency Centromeric instability and facial abnormalities (ICF) syndrome and cancer.

DNA methyl transferases (DNMTs) are directly responsible for the deposition of methyl groups on the DNA and can be divided into de novo and maintenance DNMTs, accordingly to structural and functional features. The de novo DNMTs are responsible for depositing methyl groups on non-methylated DNA, mostly in the form of heterochromatin. The maintenance DNMTs methylate hemimethylated DNA at the replication fork, and are thereby responsible for preventing the loss of 5meC during cell division. These proteins are crucial during embryonic development, since DNA methylation goes through severe reprogramming events. In fact, 5meC is almost completely lost after 3.5 days from the fertilisation and is re-established by day 6.5 of the embryonic development, when the epiblast is formed. From day 3.5 the de novo DNMTs, DNMT3B and DNMT3A, are highly expressed and methylate the genome. The maintenance DNMT1 is then expressed and maintains the newly methylated genomic loci. Mutations of these proteins determine diseases in humans, such as the ICF syndrome. The majority of ICF patients is characterised by mutations in DNMT3B and hypomethylation of DNA. However, a small percentage of patients is characterised by mutations in the Lsh gene.

Lymphoid Specific Helicase (LSH) is an ATP-dependent chromatin remodeler, whose remodeler activity has not been proven in vitro. It is essential for DNA methylation throughout the genome, in mammals and plants and it has been linked to developmentally programmed de novo methylation at unique and repetitive sequences. However, there is

still debate on whether LSH has a role in maintenance of DNA methylation as well. Knockout of Lsh gene in embryos results in ~50% reduction of 5meC in the genome. It is hypothesized that LSH remodels chromatin to enable the access of DNA methyltransferases to DNA during development, when DNA methylation patterns undergo dramatic reprogramming.

In this work, by using an in vitro culturing system, we mimicked the reprogramming of the DNA methylation typical of embryonic development and analysed changes occurring in 5meC deposition and localisation in absence of LSH. To do so, we used mouse ESCs maintained in 2i or serum-containing culturing media. The 2i-containing medium was used to deplete 5meC and ensure a naïve state of pluripotency, resembling the methylation level in the blastocyst state; the serum-containing medium determined an accumulation of DNA methylation, which took the cells to a primed pluripotent state, resembling the epiblast. Using this system, we could analyse how cells lacking LSH responded to DNA methylation reprogramming. We showed that the role of LSH is concomitant with the de novo DNA methylation timing. Furthermore, analyses of DNA demethylation suggested that LSH was not involved in maintenance of DNA methylation. Interestingly, we found that in absence of LSH; loss and gain of methylation were faster. This suggested that the chromatin in cells lacking LSH was more easily accessible to methyltransferases and demethylases, supporting the hypothesis of an involvement of LSH in DNA methylation via its remodelling activity. The same system was advantageous to investigate on a whole genome base the genomic regions requiring LSH activity. Sequencing of the unmethylated fraction of the genome was carried out on cells before and after reprogramming of the DNA methylation and on cells in the early differentiation stage. This experiment confirmed the previous findings that LSH is required both at unique and repetitive regions of the genome, such as IAPs, and gave further insights into the regions that require LSH to be de novo methylated.

In conclusion, this study has demonstrated that LSH is required only concomitantly to the reprogramming of DNA methylation, that occurs during early phases of embryonic development. Furthermore, it has provided with further evidence supporting LSH activity

as chromatin remodeler and facilitator of de novo methylation, but not maintenance. Finally, the high throughput sequencing analysis has deepened the understanding of the loci-specific activity of LSH.

LAY SUMMARY

Life starts as a single cell. All the information that the cell requires to produce a whole organism is encoded in the DNA. A human body is made of over 37 trillion cells. Each one of these cells contain the same DNA, but is different and specialised in a unique job. How is this possible? The cell has to select the information required to accomplish its specific job. To do so, the DNA is modified by the addition of chemical groups. These can allow access to a specific piece of information or, on the contrary, stop the access to it, very much like a GO or STOP signal. The most important stop signal on the DNA is called DNA methylation. When this stop signal is not at its place, the cell has difficulties to specialise in doing its job and, as a result, the adult individual has health problems.

In my work, I studied a protein called LSH. This protein works like a traffic policeman, helping to direct the deposition of DNA methylation. I discovered that LSH is important very early in the development of the embryo, working to ensure that the not accessible regions of the DNA remain so. I also found out that LSH preferentially works at some regions of the DNA. This information could provide the basic knowledge on which to build up to better understand what goes wrong when LSH does not work in humans and eventually discover a therapeutic approach to treat patients with this defect.

1. INTRODUCTION

1.1. Epigenetic gene regulation

1.1.1. History and definition

Life starts as a single cell, which is formed by the maternal and paternal genomes among fertilization. This single cell goes through embryonic development and subsequent divisions and eventually enough cells to form a whole organism are produced. Importantly, this means that, despite the presence of a variety of specialised cell types in the adult, the genome of all these cells is identical. It is therefore clear that an extra layer of regulation exists, which influences the expression of genes and thus the production of specific proteins in the cells and, therefore, the differentiation and specialization processes. It is now known that the regulation of gene expression is controlled by a variety of mechanisms that are grouped under the definition of *epigenetic* gene regulation.

The term epigenetic (literally, *above genetics*) was firstly coined by the developmental biologist Conrad Waddington as a result of his experiments in embryos of *Drosophila*. In 1942, he defined epigenetics as “the branch of biology that studies the casual interactions between genes and their products which bring the phenotype into being” (Waddington, 1942). Waddington’s focus at the time was on decoupling the genotype from the phenotype, only implying the existence of regulation. However, at his time these findings were considered very controversial, being assimilated into the Lamarkian adaptation theory that was rejected at that point. In fact, already in the 19th century, the evolutionary biologist Jean-Baptist Lamarck hypothesised the capability of an organism to acquire specific characteristics during its lifetime, as a response to environmental circumstances. These acquired characteristics could then be transmitted to

the progeny, leading to evolution by adaptation rather than by selection, as later stated by Charles Darwin.

A big step into modern "molecular" epigenetics was taken by Allfrey and Minsky (Allfrey et al., 1964), who confirmed a role for histone in RNA synthesis. In their work, they proposed a mechanism of switching on and off RNA synthesis, therefore silencing and activating gene expression, *via* histone acetylation and methylation.

Since then, the definition of epigenetics has changed to include the further knowledge acquired during the years. In 1994, Holliday offered two definitions for the term, including the concept of inheritance and the idea that these differences were not part of the DNA sequence (Holliday, 1994; Bird, 2002; Jaenisch & Bird, 2003). Russo, in 1996, took Holliday's definition and re-elaborated it into what is now probably the most commonly used definition in the field: "the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states" (Bird, 2007).

Even though the interest in epigenetics has started long ago, it is only in the past 20 years that this had become a hot topic in biology. The genome projects have expanded vastly the knowledge of the DNA sequence for different organisms. This, together with technical advances, enabled delineation of the epigenetics differences on a whole-genome level. It is now commonly agreed that there are 2 main epigenetic mechanisms that regulate gene expression: deposition of epigenetic marks, namely methylation of the DNA on the cytosine bases and modification of histone tails, and RNA-based mechanisms, such as noncoding RNAs and microRNAs.

Although big steps forward have been made, there is still a lot to be uncovered regarding how these modifications affect gene expression, how the modifications are transmitted to the progeny and how the epigenome can be influenced by the environment.

1.1.2. Types of epigenetic modifications: DNA and histone modifications

The DNA contained within each cell nucleus is very long, especially if compared to

the size of a cell, and contains a great amount of information. It is therefore important to properly package it, at the same time avoiding damages, while allowing accessibility to the necessary piece of information. To do so, the DNA is wrapped around histone proteins forming nucleosomes (Figure 1.1) (Kornberg, 1974; Längst & Manelyte, 2015). These are the core unit of chromatin and are formed by an octamer containing two copies of each histone protein, including H2A, H2B, H3 and H4, around which 147bp of DNA are wrapped. Nucleosomes are separated by linker DNA, whose length is species and cell type specific, and which is bound by the linker histone H1 (Bednar et al., 1998). Positioning of nucleosomes on the DNA determine accessibility of the information encoded in a specific genomic region and depends on the interplay between different factors, namely DNA sequence, DNA-binding proteins, nucleosomes remodelling proteins and finally RNA polymerase II machinery (Struhl & Segal, 2013). Depending on the accessibility of the chromatin, this can be classified as euchromatin or heterochromatin. Euchromatic regions are characterised by regularly positioned and interspaced nucleosomes and contain actively transcribed genes. On the contrary, in heterochromatic regions nucleosomes are densely positioned, forming a physical barrier to the DNA for transcription factors and thereby preventing gene expression. Importantly, some heterochromatic regions retain the capability to convert to euchromatin, depending for example on the developmental stage. These regions are named facultative heterochromatin, in contrast with constitutive heterochromatin, which includes regions that must not be expressed, such as repetitive sequences or non-coding centromeric DNA (Trojer & Reinberg, 2007). Deposition of epigenetic marks is crucial for the correct organisation of chromatin and consequently for gene expression regulation.

Epigenetic marks are covalent modification of either the DNA or histones and can play a role of activators and repressors of gene expression. It is now well known that gene expression is not an on/off mechanism. In fact, to finely regulate this process, not only the deposition of a single epigenetic mark is important, but also where this mark is deposited and how it interacts with the marks in the close proximity.

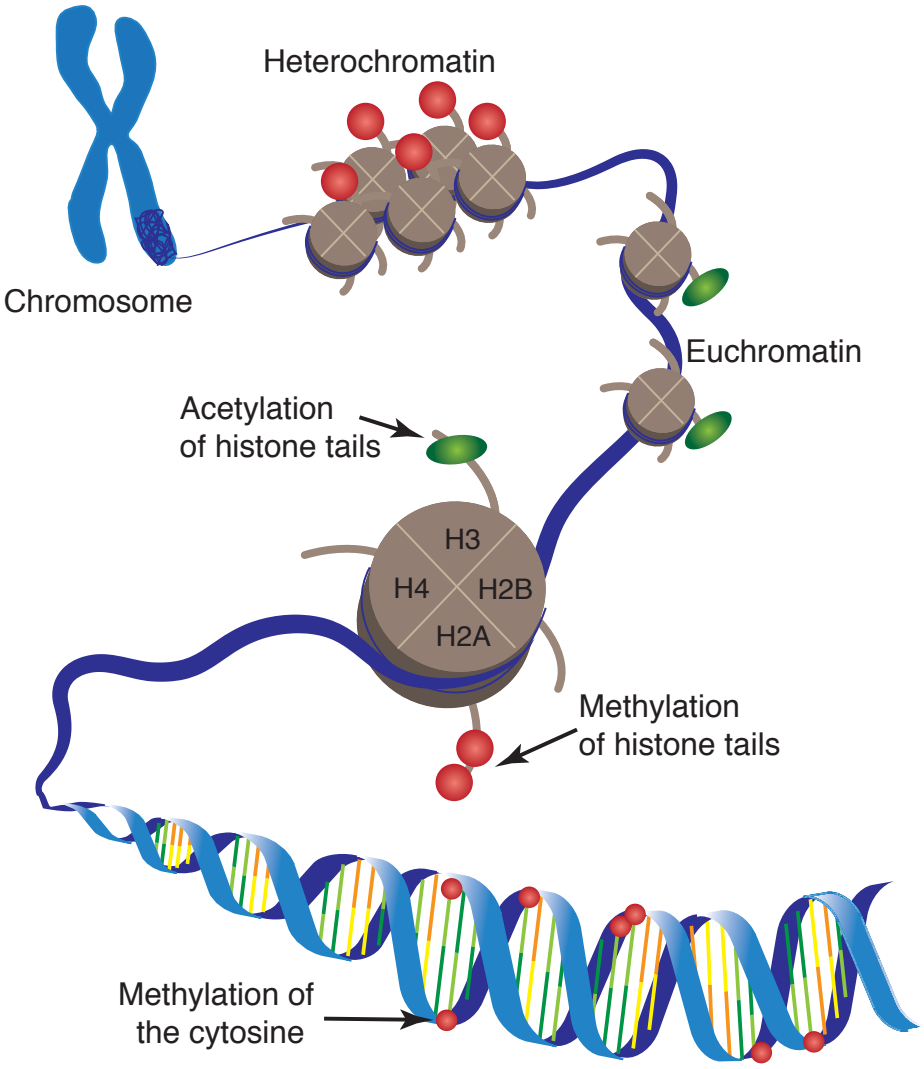


Figure 1.1 Organisation of chromatin in mammalian cells.
Schematic representation of the organisation of chromatin in the nucleus. DNA is wrapped around nucleosomes, which are then compacted into forming chromatin and, during mitosis, chromosomes. The epigenetic modifications contributing to the regulation of gene expression are annotated: methylation of the DNA and modifications of histone tails. Red dots represent methyl groups on the DNA, red elipses represent methyl groups on histone tails, green elipses represent acetylated groups on histone tails. Yellow sticks represents Adenines, orange sticks represents Thymines, light green sticks represents Cytosines, dark green sticks represent Guanines.

The most abundant modification of the DNA is the methylation of the 5th carbon of the cytosine (5meC) (Bird, 2002; Jaenisch & Bird, 2003; Smith & Meissner, 2013). The methylation of the Cs is achieved via transfer of a methyl group from the donor S-adenosylmethionine (SAM) by a DNA methyltransferase (DNMT) enzyme (Figure 1.3A). DNA methylation is mostly abundant in the context of Cytosine-Guanine dinucleotide (CpG), but can also occur at CH, CHH and CHG, where H is any nucleotide but G (Lister et al., 2009). It often acquires a repressive meaning for gene expression. However, it was shown that the effect of 5meC on gene expression varies and is dependent on the localisation of the mark (Jones, 2012). In fact, if methylation at CGIs is repressive, there is increased evidence that DNA methylation at gene bodies can have a positive effect in the regulation of transcription, facilitating elongation and splicing (Jones, 2012; Maunakea et al., 2013; Yearim et al., 2015; Neri et al., 2017). This mark can be removed from the DNA, through the production of an oxidised form of the modification, hydroxy-methylcytosine (5hmeC) (Tan & Manley, 2009). Despite being much less abundant, this modification is now considered a very important epigenetic mark, especially when studying the dynamics of the 5meC in early phases of mammalian development or in diseases.

The second category of epigenetic marks includes modifications of histone N-terminal or C-terminal tails, protruding out of the nucleosome core particle. These posttranslational modifications can involve the addition of either a small chemical group, methylation, phosphorylation or acetylation, or of a peptide, such as sumoylation or ubiquitination, (Figure 1.2) (Rodríguez-Paredes & Esteller, 2011). Some of these modifications can be very abundant in the genome and are crucial not only for the correct regulation of the transcription, but also for the organisation of high level of chromatin compaction within the nucleus. Importantly, histone and DNA modifications are not independent epigenetic modifications, but interact to regulate the expression of genes and compaction of chromatin.

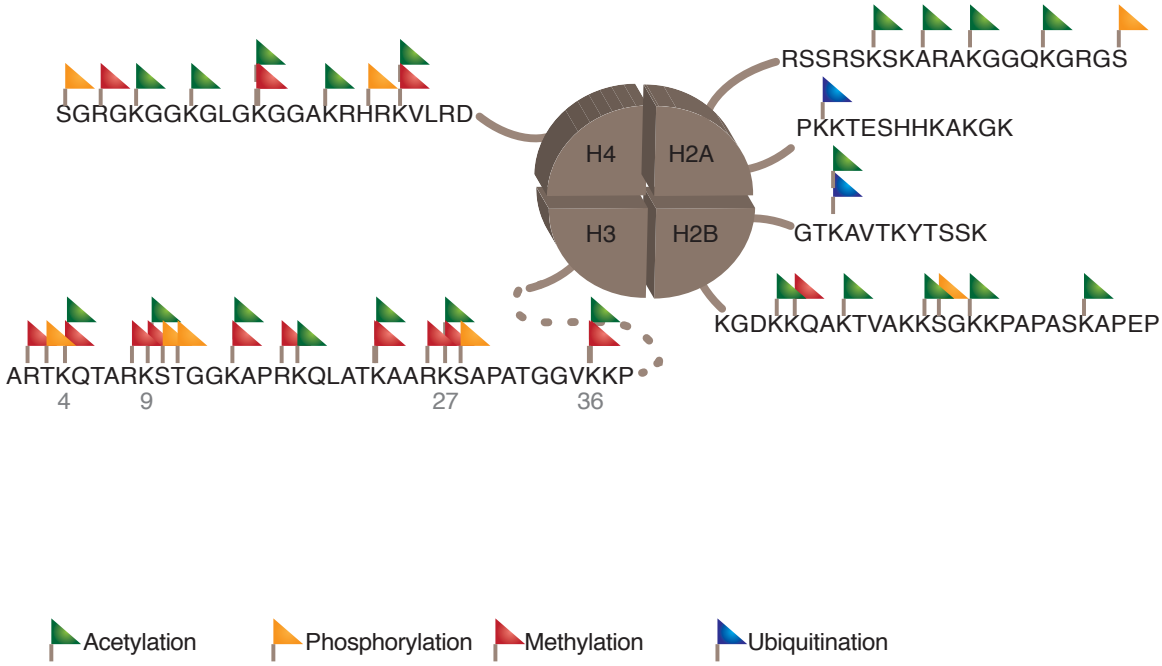


Figure 1.2 Histone tail modifications
A.Schematic representation of the most common modifications found on histone tails. Positions of the modification on specific residues are shown by the presence of a colour-coded flag. Histone modifications relevant to my work are located on Histone 3 (H3). Numbers under the residues indicate the position that is modified, H3K4me/ac, H3K9me, H3K27me/ac, H3K36me/ac.

1.1.3. Histone modifications: acetylation and methylation

It is now well established that the deposition of histone modifications is finely regulated and that the response to these marks depends on the specific residue that is modified, the localisation of the mark in the genome and also on the number of chemical groups that are added to the modified residues. It is therefore very important not only to obtain such information, but also to consider the interplay between different modifications. One of the most striking examples of how antagonistic histone modifications collaborate in regulating gene expression are bivalent domains (Harikumar & Meshorer, 2015; Voigt et

al., 2013). These are genomic regions characterised by the modifications of histone tails in an asymmetric fashion, with a repressive methylation and an active methylation mark that occupy opposite H3 tails. A lot of work has been done to shed some light on this regulatory mechanism and it has been proposed that these apparently contradictorily signalling marks poise genes, that keep a low expression and are reversibly silenced, thus allowing a quick activation or stable silencing in case of a sudden stimulus or during differentiation (Bernstein et al., 2006; Voigt et al., 2013).

Histone acetylation is one of the most studied histone modifications and is maintained between organisms, from humans to yeast and *Drosophila.* It is deposited on lysines by histone acetyltransferases (HATs), that use the acetyl-Coenzyme A (acetyl-CoA) as a cofactor and transfer an acetyl group on the histone. This neutralises the positive charges of the lysine residues and weakens the interaction between histone and DNA. It is a mark of transcription activation, a feature highlighted already in the 1960s by Allfrey (Bannister & Kouzarides, 2011; Kouzarides, 2007). In fact, acetylated lysines are signature of euchromatic regions and specifically found at enhancers and promoters of active genes, while the gene bodies are hypo-acetylated. The equilibrium between acetylated and hypoacetylated state is guaranteed by the activity of HATs and histone deacetylases (HDACs). The oppose action of HDACs restores the positive charge of the lysine and is supposed to stabilise the architecture of the chromatin and lead to repression of gene expression. The balance between HATs and HDACs activity is very important for regulation of transcription, DNA replication and DNA repair. An example of acetylation activity in gene expression was recently published. More specifically, during differentiation of ESCs towards the neural lineage, H3ac is increased at neuronal gene and is vice versa decreased in neuronal-inhibiting loci (Liu et al., 2015).

Histone tails can also be modified by addition of one, two or three methyl groups. Methylation can occur at lysine or arginine residues and can either positively or negatively affect gene expression (Bannister & Kouzarides, 2011; Harikumar & Meshorer, 2015; Martin & Zhang, 2005). Lysine 9 and 27 methylation are a common signature of heterochromatic regions and occurs at inactive genes and repetitive elements. On the

contrary, the methylation of Lysine 4 on histone H3 (H3K4me3) is a mark associated with active transcription. As aforementioned, the silencing histone 3 lysine 27 tri-methylation (H3K27me3) and the activating H3K4me3 marks can coexist at bivalent domains (Harikumar & Meshorer, 2015; Voigt et al., 2013). There is a variety of lysines/arginine methyl transferases (K/RMTs) complexes, catalysing the deposition of a methyl-group from S-adenosyl methionine (SAM) to specific substrates. The first histone KMT to be discovered was SUV39H1, a histone methyltransferase that trimethylates lysine 9 on histone 3, has a SET domain and contributes to formation of constitutive heterochromatin (Aagaard, 1999; Li et al., 2009). Subsequently, more SET and non-SET KMTs were discovered (Bannister & Kouzarides, 2011; Martin & Zhang, 2005). Among these, SUVAR39H1/2, ESET and G9a/GLP are specific for Histone 3 Lysine 9 methylation (H3K9me), while the Polycomb repressive complex 2 (PRC2) is mainly responsible for Histone 3 Lysine 27 tri-methylation (H3K27me3), via its EZH2 subunit. These protein complexes contribute to silencing of gene expression, by enabling “readers” of the epigenetic modification to bind and start repression, such as the heterochromatin protein 1 (HP1) or the Polycomb protein EED (Yun et al., 2011). On the other hand, mixed lineage leukaemia (MLLs) and SET1A and B are specific for H3K4me3 and are therefore considered facilitators of transcriptional activation. Histone methylation can be erased by demethylases enzymes, such as LSD1 (lysine (K)-specific demethylase 1A), JMJDs (Jumonji domain-containing proteins) and JHDMs (Jumonji C domain-containing histone demethylases).

It is not hard to believe that, in many instances, histone methylation and DNA methylation have been shown to be indispensable for each other, especially when organising the compaction of the chromatin and therefore establishing transcriptional silencing. While in organisms such as *Neurospora Crassa*, a mould, this link is unidirectional from histones to DNA modification, through HP1 (Du et al., 2015), in mammals it is bidirectional. In mammals, this interaction can be bridged by proteins which recognise DNA methylation through their Methyl-CpG Binding Domain (MBD1-4 and MeCP2). For example, MBD1 can recruit ESET to newly synthesised DNA and thereby

participate in the maintenance of H3K9 methylation during cell division (Rose & Klose, 2014; Sarraf & Stancheva, 2004). MDB1 was also shown to interact with Suv39h1 and HP1, which recruit HDACs, and coordinate transcriptional repression (Fujita et al., 2003; Rose & Klose, 2014). On the contrary, evidence of DNA methylation dependent on histone methylation is given for example by the recruitment of the *de novo* DNMTs by G9a/GLP complex. This cascade is important to antagonise demethylation of the DNA at imprinted loci (Zhang et al., 2016).

Due to the very complicated and interlinked relationship between DNA and histone methylation, it remains unclear the order of events, that leads to gene transcriptional repression and change in chromatin architecture and could be applicable to all instances. While in *N. crassa*, as aforementioned, DNA methylation always depends on histone methylation (Du et al., 2015), in mammals, DNMTs seems to recognise heterochromatin through direct interaction with histone methyltransferases, such as Suv39h1, Eset and G9a/GLP (Rose & Klose, 2014). However, knock out in cancer cell lines of *Dnmt1* gene was shown to lead to loss of H3K9me, suggesting the dependence of histone methylation on DNA methylation (Espada et al., 2004; Rose & Klose, 2014).

Role	Proteins	Function	References
Writer	DNMT3A, DNMT3B, DNMT3C, DNMT3L DNMT1, G9a	DNA methyltransferases: use SAM to transfer a methyl group on the 5th carbon of the cytosine	Figure 1.3 and Goll & Bestor, 2005;
Reader	MBD UHR Zinc-finger proteins		Meng et al., 2015
Eraser	TET, TDG	Remove methylation from the 5th carbon of the cytosine	Figure 1.4 and Rasmussen et al., 2016

Table 1.Control of DNA methylation.
Brief summary of the main proteins that participate into DNA methylation control, as writers, readers or erasers. In the last column, references of main works on the proteins and figures in this thesis work.

1.2. DNA methylation

1.2.1. Characterization of DNA methylation in the mammalian genome

One of the main regulatory marks for gene repression is the methylation of the fifth carbon of the cytosine (5meC). This was first detected in calf thymus DNA by paper chromatography in 1948 (Hotchkiss, 1948). However, it took almost two decades to understand the importance of this modification (Compere & Palmiter, 1981; Holliday & Pugh, 1975).

DNA methylation on the cytosine is found in fungi, plants and animals, and can be inherited by the next generation of individuals. However, pattern and even presence of DNA methylation varies among different species (Table 2). As for its presence, it can be stated that 5meC is found in all groups of eukaryotes, namely protists, fungi, plants and animals, but it is absent in some species, such as the common laboratory animal models *Caenorhabditis elegans and Drosophila melanogaster.* Furthermore, two main patterns of 5meC in the genome have been identified: mosaic or whole genome (Aliaga et al., 2019; Colot & Rossignol, 1999). In invertebrates, 5meC is mainly spread in a mosaic-fashion, where regions with high methylation are alternated with regions completely depleted of the modification. In contrast, in invertebrates, DNA methylation is mostly distributed throughout the whole genome and depleted only at regulatory elements (Aliaga et al., 2019; Capuano et al., 2014; Colot & Rossignol, 1999). Importantly, DNA methylation is very abundant in the mammalian genome. In fact, 70 to 80% of cytosines in the context of CpG dinucleotide are methylated in mammals (Law & Jacobsen, 2011). Most CpG islands (CGIs), CpG rich regions, of maximum 1kb length and commonly found at transcription start sites, but also in gene bodies and intergenic regions, are normally depleted of methylation. If CGIs become methylated, this leads to long-term silencing of the associated genes (Jones, 2012). Methylation outside of CGIs is more common and usually more dynamic and tissue-specific (Jones, 2012).

Group	DNA methylation	Exceptions
Plants	Yes, on Cs in CpG and non-CpG context. Abundant at repetitive elements.	• *Chlamydomonas reinhardtii* (Lopez et al., 2015): low level of methylation in the nuclear genome, abundant 5meC in the chloroplast genome. Hypermethylation in gametes and zygote.
Fungi	Yes, on Cs in CpG and non-CpG context. Abundant at repetitive elements.	• *Saccharomyces cerevisiae* (Capuano et al., 2014)*:* liquid chromatography analyses show no DNA methylation in commercially and laboratory yeast strains.
Animals	Yes, mainly on Cs in CpG context. Abundant at repetitive elements.	• *Caenorhabditis elegans* (Greer et al., 2015): no 5meC, but 6mA (methylation of the 6th N of Adenine). • *Drosophila melanogaster*(G. Zhang et al., 2015) : extremely low level of 5meC, predominant role of 6mA.

Table 2.DNA methylation in different living organisms.
Brief summary of the characteristics of DNA methylation in different groups of eukaryotes. 5meC patterns are very variable among different species, therefore in the last column some species are listed as examples of exceptions to the characteristics of their group.

While non-CpG methylation is widespread in plants (Law & Jacobsen, 2011), and has been observed in mammals (Ramsahoye et al., 2000), CpG methylation remains the most important modification, since it is more abundant and is involved in a variety of cellular mechanism. More specifically, 5meC is important during development to regulate gene expression in a tissue-specific manner and for genomic imprinting, by regulating the parent-of-origin expression of specific alleles. Furthermore, it is crucial for preserving genome stability and integrity by silencing endogenous retrotransposons. Finally, it is involved in inactivation of X-chromosome in female mammals, on which silenced genes are DNA methylated (Bird, 2002; Jaenisch & Bird, 2003). As a consequence of its involvement in these processes essential for cells and organisms, misregulation of this mark, *via* mutations in proteins involved in the methyl group deposition or in the binding to 5meC, results in diseases. Some examples are imprinting disorders; Immunodeficiency,

Centromeric instability and Facial abnormalities (ICF) syndrome, characterised by mutations in DNMT3B, LSH and CDCA7; Rett Syndrome, determined by mutations in MeCP2; Lupus and cancer (Reik, 2007; De Sario, 2009; Smith & Meissner, 2013).

Methylation of the DNA is catalysed by a group of enzymes called DNA methyl transferases (DNMTs), also defined as “writers” of the DNA methylation (Goll & Bestor, 2005). These enzymes use SAM as a donor of the methyl group, which is then transferred to the 5th carbon group of the cytosine leading to the formation of 5meC and the release of SAH (Figure 1.3A). The deposition can occur on unmodified DNA, mechanism defined as *de novo* methylation, by the *de novo* methyl transferases DNMT3A and DNMT3B. On the other hand, when DNA is duplicated, the newly synthesised strand is unmodified, while the complement strand keeps the modification. DNMT1 is the maintenance methyl transferase, capable of methylating hemimethylated DNA at heterochromatic regions and at the replication fork (Figure 1.3B).

Despite a lot of work that has been done to elucidate the mechanisms by which the 5meC is involved in the regulation of gene expression, there are yet questions to be answered. In particular, it is very important to fully understand how the methylation is deposited and regulated during the early phases of embryonic development. In fact, this would allow a better understanding of what goes wrong in human diseases and would possibly open new routes in the treatment of these patients.

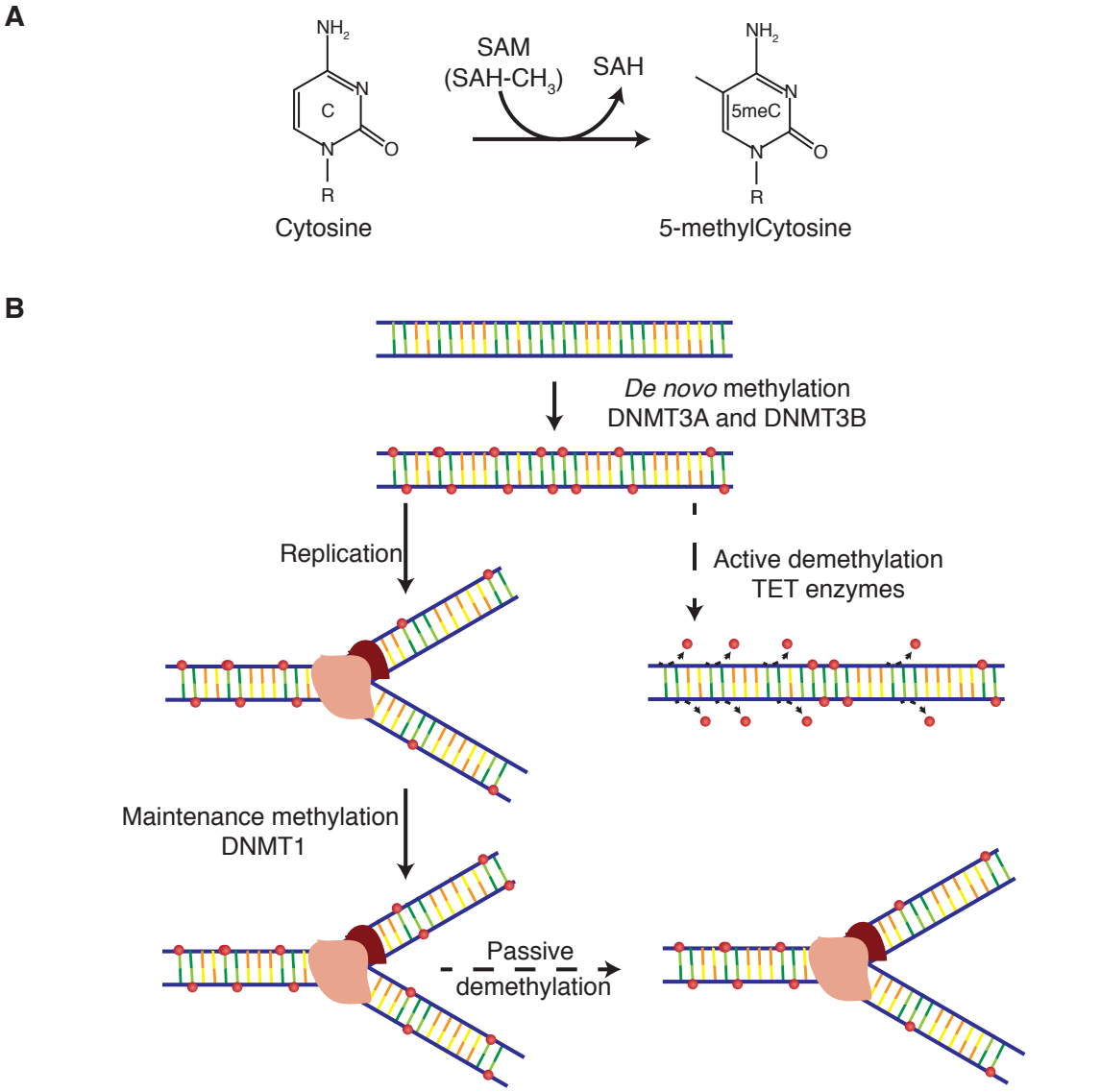


Figure 1.3 DNA methylation
A.Schematic representation of the chemical reaction occuring at cytosine when a methyl group is deposit. SAM is the methyl group donor. The CH3 is then transfered on the 5th carbon of the cytosine ring, and SAH is released.
B.Naked DNA is methylated by the *de novo* methyltransferases DNMT3A and DNMT3B. At replication, the maintenance DNA methyltransferase DNMT1 symmetrically methylates the newly synthesised DNA strand. Removal of methyl groups occurs actively by the TET enzymes or passively, by lack of maintenance activity at the replication fork.

1.2.2. De Novo DNA methylation and DNA Methyltransferases

DNMTs are capable of depositing a methyl group onto the cytosine. However, the two groups of methyltransferases, *de novo* and maintenance, differ in their substrate specificity. In fact, while DNMT1 propagates the methylation concomitantly with cell division and has a specific activity for hemimethylated DNA, DNMT3s have affinity for hemi and non-methylated DNA (Okano et al., 1999).

The main methyltransferases involved in *de novo* methylation are DNMT3A, DNMT3B and DNMT3L (Okano et al., 1999). The DNMT3A and DNMT3B have a very similar multidomain structure: a C-terminal catalytic domain; the ADD domain, responsible for DNA binding through a conserved amino acid sequence; a variable N-terminal domain, containing regulatory regions and a PWWP domain, responsible for the recognition of histone 3 lysine 36 tri-methylated (H3K36me3) (Goll & Bestor, 2005). DNMT3L is a much smaller protein, completely lacking the PWWP domain and is thereby considered an inactive member of this family of protein. In fact, DNMT3L is important in guiding DNMT3A during methylation in gametes, but has no direct methyl transferase activity (He et al., 2011).

Despite the similar structure, DNMT3A and DNMT3B have distinct biological roles and targets in the genome. Evidence of this are found in studies in knock out models for *Dnmt3a* and *Dnmt3b*, that showed different phenotypes in mice (Li et al., 1992; Smith & Meissner, 2013). More specifically, mice lacking DNMT3A can normally develop to birth, but die shortly after. However, absence of DNMT3B is embryonic lethal and determines severe defects in the development of the embryo.

DNMT3A is expressed earlier in the zygote and is fundamental for the methylation of maternal and paternal imprints, with the help of DNMT3L (Figure 1.5A) (Law & Jacobsen, 2011; Smith & Meissner, 2013). On the other hand, DNMT3B has the biggest contribution to the *de novo* methylation later in the embryonic development, that occurs during implantation following the almost complete erasure of 5meC at the blastocyst stage

(methylation dynamics in the early development will be discussed later in this chapter).

A newly discovered *de novo* methyltransferase is DNMT3C. This enzyme seems to be crucial for the regulation of DNA methylation at retrotransposons in male germ cells (Barau et al., 2016).

While *de novo* methylation occurs mostly during embryonic development, gene silencing via DNA methylation deposition is important for lineage commitment and differentiation. In fact, conditional knock outs of the *de novo* methyltransferases in mouse Embryonic Stem Cells (mESCs) do not lead to cell death. However, when cells are pushed towards lineage commitment, these are not capable of keeping the pluripotency genes silenced and die (Ambrosi et al., 2017).

DNMT3s act on DNA that does not present any DNA methylation. However, this does not mean that the DNA that gets methylated is completely naked. In fact, although it is not clear yet, it is now commonly agreed that DNA methylation arrives only after the deposition of other epigenetic marks, such as histone methylation. The PWWP domain of DNMT3s can in fact recognise H3K36 methylation, which directs the protein to the gene body (Baubec et al., 2015; Neri et al., 2017; Rondelet et al., 2016). This co-occurrence not only indicates the importance of the interplay between epigenetic marks, but also suggests that the *de novo* methyltransferases have to deposit methyl groups on chromatin that is already tightly packed around histones. Interestingly, studies *in vitro* have shown that the activity of the DNMT3s for compacted chromatin is lower if compared to the activity on naked DNA (Felle et al., 2011; Robertson et al., 2004). Therefore, before methylating the DNA, these proteins require the activity of a different category of enzymes, the chromatin modifiers and chromatin remodelers. The latter support DNA methylation by facilitating the access to the chromatin for the DNMTs, by sliding or unwrapping a nucleosome or evicting a histone octamer. Among these proteins, the Lymphoid Specific Helicase (LSH) will be the focus of my thesis.

1.2.3. Maintenance of DNA methylation and maintenance machinery

During mitosis, cells duplicate their DNA by using the original strands as templates. It is important at this stage to properly modify the newly synthesized DNA strand in order to preserve the epigenetic state of the loci in the new cell. In order to do so, the DNA methylation maintenance machinery is used. The DNA methyl transferase involved in this process is DNMT1, the first to be discovered and most abundant in the cells among all DNMTs. Despite being involved in *de novo* DNA methylation, via interaction with DNMT3A (Ambrosi et al., 2017; Law & Jacobsen, 2011), DNMT1 has a high specificity for hemimethylated DNA, guaranteed by auto-inhibition among binding of unmethylated DNA, and is capable to then methylate the newly synthesised and un-modified strand. Importantly, it was shown that *Dnmt1*$^{-/-}$ ESCs undergo apoptosis when induced to differentiate and *Dnmt1*$^{-/-}$ embryos die at day 9.5 of embryonic development (Ambrosi et al., 2017; Law & Jacobsen, 2011; Li et al., 1992), indicating the importance of this enzyme for the survival of the embryo.

Other enzymes are also crucial for maintaining the DNA methylation. First of all, it was shown that in mESC lacking the *de novo* methyltransferases, DNMT1 is not able to maintain on its own the DNA methylation, suggesting a role for DNMT3s into maintenance and that DNMT1 is necessary but not sufficient for its role (Ambrosi et al., 2017; Chen et al., 2003; Liang et al., 2002). Furthermore, DNMT1 is recruited to the replication forks by UHRF1 (Ubiquitin Like With PHD And Ring Finger Domains 1, called Np95 in mouse). The methylated H3K9-like present at the N-terminal domain of the DNA ligase 1 (LIG1) binds UHRF1 and recruits DNMT1, ensuring maintenance of DNA methylation (Ferry et al., 2017). This newly highlighted interaction between LIG1 and UHRF1 also showed a new role for lysine methylation in maintenance of DNA methylation. In fact, the N-terminal domain of LIG1 is a histone mimicking site, methylated by a histone methyl transferases, G9a, in complex with GLP.

G9a and GLP are highly related and are lysine methyl transferases (KMTs) containing a SET domain. These belong to the same family of KMTs as SUV39H1 and H2,

SETDB1/ESET and PRDM3 and 16 (Dong et al., 2008; Jurkowska et al., 2011) and can methylate non-histone proteins as well. However, G9a and GLP are specifically responsible for the deposition of mono and di-methylation on H3 lysine 9. G9a is most commonly found as either a homodimer or heterodimer with GLP (G9a Like Protein), but in ESCs the most abundant form of the complex is the heterodimer of G9a with GLP formed via their C-terminal catalytic domains (Shinkai & Tachibana, 2011). The dimer is directed to DNA by the WIZ protein, which also acts as a stabiliser of the complex (Bian et al., 2015). The knock outs of *G9a* and *Glp* in mice result in embryonic lethality (Shinkai & Tachibana, 2011). Mice carrying a catalytically inactive form of G9a have a similar phenotype to the KOs for *G9a/Glp*, with embryonic lethality. This highlights the impossibility to proceed through the development in absence of G9a mediated H3K9me2.

Importantly, the interplay between DNA methylation and histone methylation by G9a/GLP has been extensively studied. As mentioned above, it was very recently highlighted a role of this complex in methylating LIG1 and thereby recruiting UHRF1, which is a crucial step for the correct maintenance of DNA methylation (Ferry et al., 2017). A role for G9a in DNA methylation was firstly shown by the direct interaction of G9a with DNMT3A and 3B *via* its ankyrin domain (Epsztejn-Litman et al., 2008). The same domain is important for G9a role in protecting imprinting control regions, for which its catalytic activity is not required (Zhang et al., 2016). This activity in maintenance was shown at other loci in the genome, such as promoters (Myant et al., 2011) and retrotransposons (Dong et al., 2008). On the other hand, the deposition of H3K9me1/2 by G9a/GLP is crucial for gene repression and silencing of enhancers in the inner cell mass (ICM) during mouse development (Zylicz et al., 2015).

1.2.4. Removal of DNA methylation: DNA hydroxyl-methylation and TET enzymes

The final important way cells control DNA methylation is via its removal. It is in fact

crucial to specifically remove methyl groups from the DNA in order to reactivate gene expression, for instance during germ cell development, transition to pluripotency and during differentiation. Removal of DNA methylation occurs mainly via two routes: passive demethylation and active demethylation (Bhutani et al., 2011; Moore et al., 2013).

Passive demethylation is mainly linked to DNA replication. In fact, when cells duplicate the DNA, as discussed earlier, the newly synthesised DNA strand is unmodified (Figure 1.3B). Therefore, the DNA maintenance machinery acts to establish the symmetrical modification. However, if this machinery is inefficient or impaired, the resulting DNA will stay hemimethylated (Figure 1.3B). At the next round of cell division, the methylation will be completely lost on the DNA synthesised using the unmethylated strand as a template, and so on and so forth. This mechanism is mainly responsible for the loss of methylation mainly of the maternal genome from the fertilisation to the 3-pronulcei phase, approximately 15 hours after, when DNMT1 expression is very low and DNA methylation is almost completely lost (Amouroux et al., 2016; von Meyenn et al., 2016).

Active demethylation implies the process of methyl groups removal from the genome that occurs not concomitantly with DNA duplication. This active removal is operated by two main families of "erasers" in mammals: AID/APOBEC (activation-induced cytidine deaminase/apolipopotein B mRNA-editing enzyme complex) and Ten Eleven Translocation (TET) enzymes (Figure 1.4) (Moore et al., 2013). The first route is a deamination reaction of the amine to a carbonyl group, leading to the formation of a thymine and thus a mismatch. The first AID step is then followed by the correction of the mismatch by the Base Excision Repair (BER) pathway, through the activation of the thymidine DNA glycosylase (TDG). The second route is made possible by the activity of the TET enzymes. These is a family of 3 oxygenases (TET1, TET2, TET3), that use O_2 to oxydise 5meC to 5 hydroxymethyl-cytosine (5hmeC) (Tahiliani et al., 2009). The subsequent reactions are further oxidations of 5hmeC to 5-formyl-cytosine (5fC) and 5-carboxy-cytosine (5CaC), which is then finally removed by the intervention of the BER pathway, leading to the complete removal of the methyl group and the re-establishment of an unmodified cytosine. Alternatively, 5hmeC can be deaminated by AID to 5

hydroxymethyl-uracil (5hmeU) leading again to the activation of the BER pathway (Figure 1.4).

The discovery of these enzymes finally proved the existence of an active removal of the DNA methylation in metazoan species, including mammals, that was until then only hypothesised. However, there are still many open questions. For instance, the role of the TET enzymes during the early phases of development is yet controversial. In fact, while 5hmeC was initially detected in the paternal genome, more recently mass spectrometry analyses have shown that rate of paternal 5hmeC does not change significantly in absence of TET3, suggesting that DNMT3A and DNMT1 might be accounted for the demethylation (Amouroux et al., 2016). Nevertheless, 5hmeC is now considered an epigenetic mark in its own right, but its biological meaning is yet to be fully uncovered.

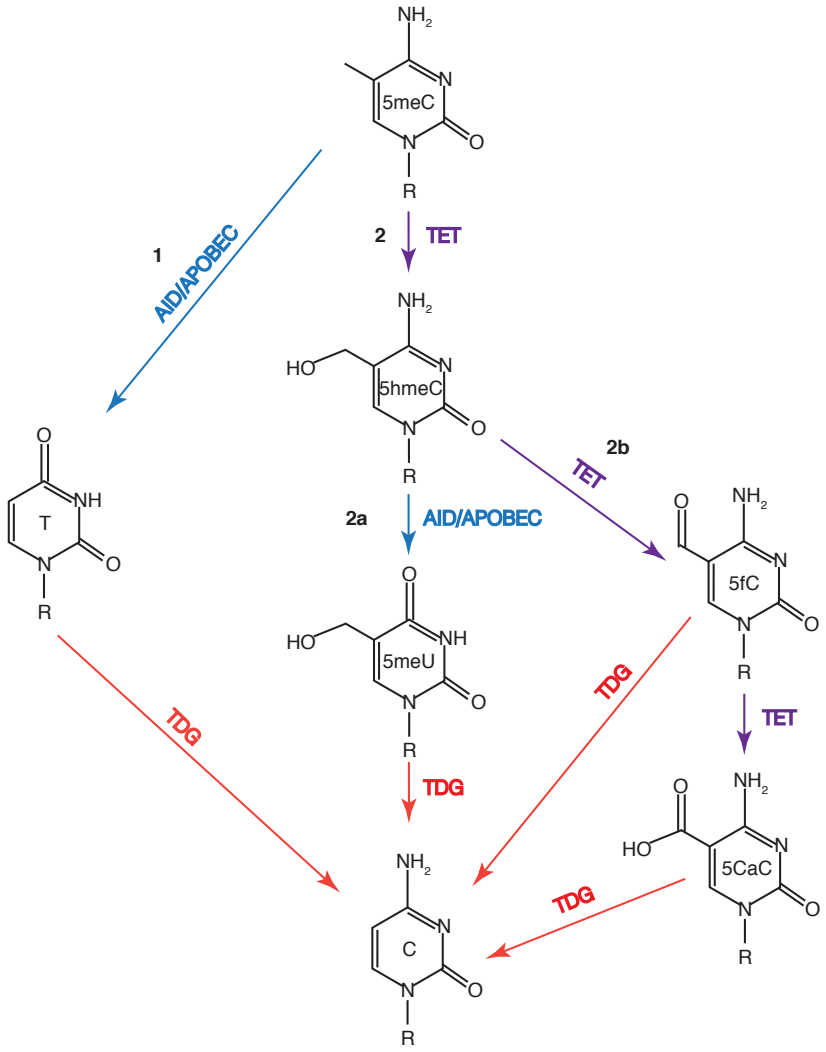


Figure 1.4 Active DNA demethylation
Different routes for active DNA demethylation. 5meC can be deaminated by the AID/APOBEC complex (1) to Thymidine (T) and then the mismatch repaired by the BER pathway, by thymidine DNA glycosylase (TDG). Alternatevely, the 5meC can be oxydised by the Ten Eleven Translocation enzymes (TET) to 5hmeC (2), which can be either subsequently deaminated by the AID/APOBEC complex (2a) to 5hmU or further oxydised by TET once to 5fC or twice to 5caC (2b). The mismatch is repaired by TDG in all cases, leading to the formation of Cytosine (C).

1.3. DNA methylation in early embryonic development

1.3.1. Dynamics of DNA methylation in early embryonic development

It has been long known that some traits can be inherited by the progeny, despite not being encoded in the genome. It is now better understood how some of these epigenetic modifications are passed onto the new generation and their importance in the correct formation of the embryo. These loci are imprinted genes (Ferguson-Smith, 2011). The term genomic imprinting describes a phenomenon by which a group of genes is monoallelically expressed, accordingly to the methylation status of the parent of origin. The DNA methylation of the differentially methylated regions (DMRs) of the parental genome regulates imprinting. However, not all DNA methylation is inherited from the parental genomes. In fact, this modification goes through heavy reprogramming in germ cells and then during the early phases of the development (Figure 1.5B) (Ambrosi et al., 2017; Smith et al., 2012; Wang et al., 2014).

The maternal and paternal gametes have a very well established and different DNA methylation profiles. More specifically, the paternal genome has higher methylation levels compared to the maternal genome (Figure 1.5B). Before the parental nuclei fuse, after fertilisation, a demethylation wave is started with the almost complete erasure of 5meC from the genome by the time the inner cell mass forms in the pre-implantation blastocyst. However, the depletion of 5meC occurs with different dynamics in the two fusing genomes (Ambrosi et al., 2017; Smith et al., 2012; Wang et al., 2014).

The sperm genome undergoes a fast demethylation process (Ambrosi et al., 2017; Howlett & Reik, 1991), resulting in an almost complete loss of 5mC prior the beginning of the cell divisions. Studies have shown that the loss of methylation in sperm is active and most probably accomplished by the TET3 enzymes (Ambrosi et al., 2017; Iqbal et al., 2011; Wossidlo et al., 2011). In fact, accumulation of 5hmeC was detected in the time frame between fertilisation and cell division. This supports the hypothesis that the

demethylation occurring is a consequence of the activity of the TET enzymes. Surprisingly, a recent study (Amouroux et al., 2016) showed that the demethylation of the paternal genome in the zygote is not affected by the loss of TET3, suggesting that other mechanisms might be crucial for this loss.

On the other hand, the oocyte has a lower DNA methylation compared to sperm. However, loss of methylation is slower and the minimum is reached at the blastocyst stage (Ambrosi et al., 2017; Howlett & Reik, 1991). In this case, the loss is passive and concomitant with the cell divisions. Due to the difference in rate of loss, the methylation detected between the PN1 and PN2 phases of the embryonic development is to attribute mainly to the maternal genome. This pattern is then slowly lost by dilution.

Some loci are excluded from this demethylation (Wang et al., 2014), such as differentially methylated imprinting control regions and some repetitive elements, such as Intercisternal A particles (IAPs). This indicates the existence of a maintenance mechanism that stop the demethylation from happening and therefore protects specifically these loci. Furthermore, this also suggests that DNA methylation might not be the only mark that suppresses the transcription of these elements. This was proven true for the IAP elements (Lane et al., 2003): DNMT1O, a shorter isoform of DNMT1, is the maintenance DNA methyltransferase responsible for protecting these loci in the early embryo. DNMT1O is maternally inherited and, despite DNMT1 being expressed at all stages of development, it is crucial for keeping these repetitive elements methylated. DNMT1s becomes important only later, after the blastocyst stage (Ambrosi et al., 2017).

After embryonic day 3.5 (E3.5), the blastocyst starts developing into the epiblast and concomitantly with the implantation of the embryo, a *de novo* methylation wave starts (Figure 1.5B). The *de novo* methyltransferases are expressed and methyl groups are deposited onto the genome. Despite both *de novo* methyltransferase being expressed, it was shown that DNMT3B is the main responsible for the wave of methylation occurring at this stage (Ambrosi et al., 2017; Borgel et al., 2010). DNMT3a is more important in later stages of development, at the onset of differentiation.

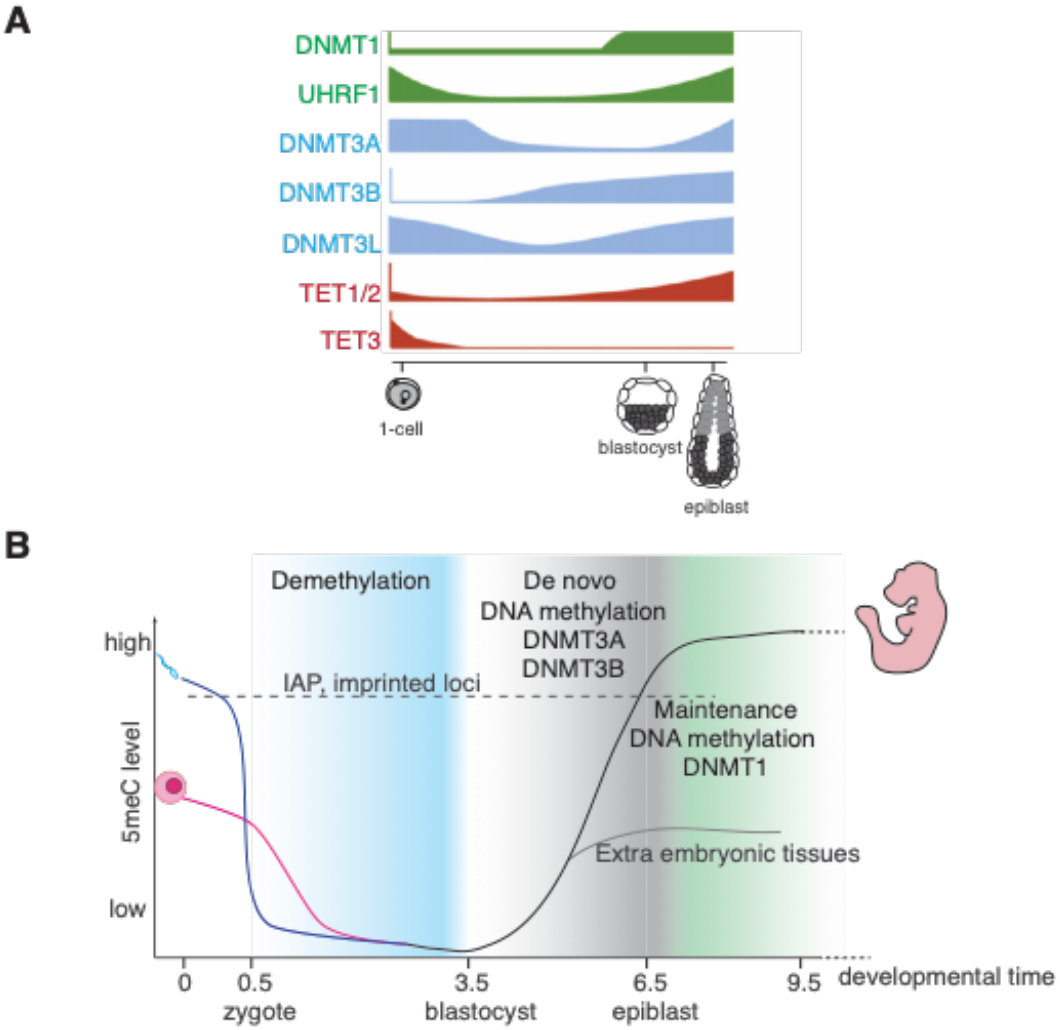


Figure 1.5 DNA methylation reprogramming during the early mammalian development.
A. Expression patterns of the DNMTs, UHRF1 and TET enzymes during development. Colors are associated with the mechanism the enzymes are mainly involved in: green is maintenance, light blue is *de novo* DNA methylation and red corresponds to demethylation.While DNMT1 is activated later during development, DNMT3A, helped by DNMT3L, is active since the initial phases. Interestingly, the TET enzymes are not very highly expressed early in development, indicating that these are not the main actors in the rapid demethylation of the sperm genome occurring after fertilization. Figure adapted from Ambrosi et al, 2017
B. Dynamics of DNA methylation during early embryonic development in mouse.
On the *y* axis global 5meC is indicated and on the *x* axis the enbryonic day. DNA methylation levels of the oocyte and the sperm are high. After fertilisation and the subsequent cleavege stages of the zygote, the 5meC is lost from the genome (blue and pink lines), but remains at some specific regions, namely the imprinted loci and repetitive elements such as IAPs (dashed line).
The lowest level of DNA methylation is reached by day 3.5 of embryonic development, by which time the blastocyst is formed. Expression of *de novo* methyl transferases re-establishes methylation throughout the genome. Activation of the expression of DNMT1 assures maintenance of the DNA methylation during the following cell divitions, that will take to the formation of the embryo.
An exeption to this fast increase in 5meC are the extraembryonic tissues, developing at this moment of the gestation. In fact, 5meC is not as abundant as it is found in the embryo.

1.3.2. Misregulation of DNA methylation in early development

Due to the biological importance of DNA methylation, any dysfunction of the modification, of one of its readers, writers or erasers (Table 1), or of a chromatin remodeler, can lead to disease or even embryonic lethality, due to the severe misregulation of gene expression and consequently cell differentiation.

These phenotypes can be linked to hypomethylation or hypermethylation of the genome. Overall, the consequences of the absence of DNA methylation are linked to changes in the conformation of chromatin. In fact, DNA methylation itself limits the accessibility of chromatin by making protein and transcription binding site not available (Keshet et al., 1986) and, by doing so, participates in the protection of the genome. Therefore, hypomethylation could be linked to open chromatin, increased accessibility and thereby affect genome stability and integrity (Meng et al., 2015).

Knock outs of DNMT1 in mice lead to embryonic lethality by E9.5, whereas KOs in mESCs determine reduction of the global methylation level and has no effect on cell proliferation or viability (Li et al., 1992; Liao et al., 2015). This suggests that maintenance of DNA methylation does not only depend on DNMT1 but has probably redundant effectors, that can compensate for absence of DNMT1, when the cells do not go through reprogramming. Mutations in *Dnmt1* in humans are also associated with cancer. Similarly, in mice mutations of DNMT3B can be either embryonic lethal or lead to growth retardation and hypomethylation at minor satellite DNA (Meng et al., 2015; Okano et al., 1999). In adult mice and humans, mutations in *Dnmt3b* are common in cancer (Zhang et al., 2017). Importantly, mutations of this *de novo* methyltransferase were found in patients affected by a rare autosomal recessive disease, the Immunodeficiency Centromeric Instability Facial Abnormalities Syndrome (ICF). Knock outs of *Dnmt3a* in mice are not embryonic lethal, indicating the importance of this protein later in embryonic development (Meng et al., 2015; Okano et al., 1999). However, these mice die shortly after birth. Finally, triple KOs of the DNMTs are embryonic lethal.

Experiments in mice have shown that both single and double KOs of *Tet1* and *Tet2* are viable (Rasmussen & Helin, 2016). Single KOs showed specific phenotypes: while TET1 depletion leads to reduced litter size, lack of TET2 results in higher incidence of leukaemia (Ambrosi et al., 2017). Interestingly, *Tet3* KO mice die shortly after birth. As TET3 is the main demethylase expresses at the very first stages of development (Figure 1.5A), these results suggest that it might be the most developmentally important demethylase.

Chromatin remodelling enzymes are crucial for providing methyl transferases with access to chromatin. When thinking of DNA methylation as a keeper of genome stability and integrity, chromatin remodelling enzymes could be considered crucial for maintaining genome stability. Therefore, any mutation or deletions of remodelers involved in DNA methylation subsequently impairs this mechanism. Mutations of chromatin remodelers were in fact found in many cancer types (Längst & Manelyte, 2015; Narlikar et al., 2013), showing that these proteins could be possible good candidates for new therapeutic approaches.

1.4. Chromatin remodelling

1.1.1. ATP dependent chromatin remodelers families

Packaging of DNA into cell nuclei is essential for the survival of any organism. This is achieved because of a very well organized structure, the chromatin, whose unit is the nucleosome core. As discussed above, a nucleosome is composed by 147bp of DNA tightly wrapped 1 and three quarter turns around a histone octamer. Nucleosomes are then separated by a linker DNA, which has a variable length between 20 and 90 bp. This

structure is highly flexible and dynamic and has been thoroughly studied in the past decades by employing variety of approaches including new technologies that enabled whole genome nucleosome mapping. It has been shown that the positioning of the nucleosomes on the chromatin depends on DNA sequence and thus is determined by the chemical characteristics of the DNA (such as the preference for GC rich sequences over AT rich ones) (Blossey & Schiessel, 2018). However, the variability of nucleosome positioning in differentiating cells or during development suggests that the sequence cannot be the only regulation mechanism for the positioning of nucleosome.

Importantly, in order for changes in positioning to occur, many hydrogens bonds between nucleosomes and DNA have to be disrupted. Enzymes responsible for this are the chromatin remodelers. These are typically DNA translocases that use ATP hydrolysis to rearrange nucleosomes on chromatin. Chromatin remodelers are very abundant. In fact, it has been estimated that there is about one remodelling complex per 10 nucleosomes in the nucleus (Erdel et al., 2011; Erdel & Rippe, 2011; Rippe et al., 2007).

All the members of this class of enzymes are characterised structurally by the presence of a catalytic ATPase domain of the Snf2 subfamily, consisting of a DExx and a Helicase C domains (Figure 1.6) (Längst & Manelyte, 2015), a superfamily of proteins. Furthermore, there is a large variety of proteins and protein complexes classified as chromatin remodelers, all characterised by some basic properties: affinity for DNA and nucleosomes, reader domain for binding to epigenetic modifications, domains regulating the ATPase activity and domains for the interaction with other proteins (Clapier & Cairns, 2009). These conserved domains are accompanied by unique domains, such as the Bromo or Chromo domains, that are the “reader” domains of the protein, capable of recognising epigenetic marks and determining the recruitment of the remodeler onto the chromatin. Based on the sequence similarity of the proteins, 4 conserved families have been classified: SWI/SNF, CHD, ISWI, INO80 (Figure 1.6).

The SWI/SNF (switching defective/sucrose nonfermenting) family was the first to be identified in yeast and is characterised by the presence of a C-terminal Bromo domain,

that recognises acetylated lysines in histone tails. Furthermore, an HSA (Helicase SANT) N terminal domain is present and is predicted to bind DNA. Two main complexes belong to this family, in human BAF and PBAF, organized around the BRG1 and BRM subunits.

The CHD (chromodomain, helicase, DNA binding) family has a characteristic N terminal domain, containing two tandem Chromo domains, responsible for the binding to methylated lysines at histone tails. It could be further divided into subfamilies, based on other structural domains (Clapier & Cairns, 2009; Längst & Manelyte, 2015). The main complex in mammalian cells is NuRD (nucleosome remodelling and deacetylase), a large remodelling complex containing a number of auxiliary proteins including histone deacetylases HDAC1 and HDAC2 and MBD3, a member of methyl CpG-binding domain protein family.

The ISWI (imitation switch) family is characterised by a SANT and a SLIDE domain, adjacent to each other on the C-terminal of the protein. It contains 2 to 4 subunits and forms the human NURF, CHRAC and ACF complexes.

The INO80 (inositol requiring 80) family is peculiar for the presence of an insertion, forming a split ATPase domain. This insertion is required for the binding of a helicase-related protein and actin related proteins (Längst & Manelyte, 2015; Narlikar et al., 2013).

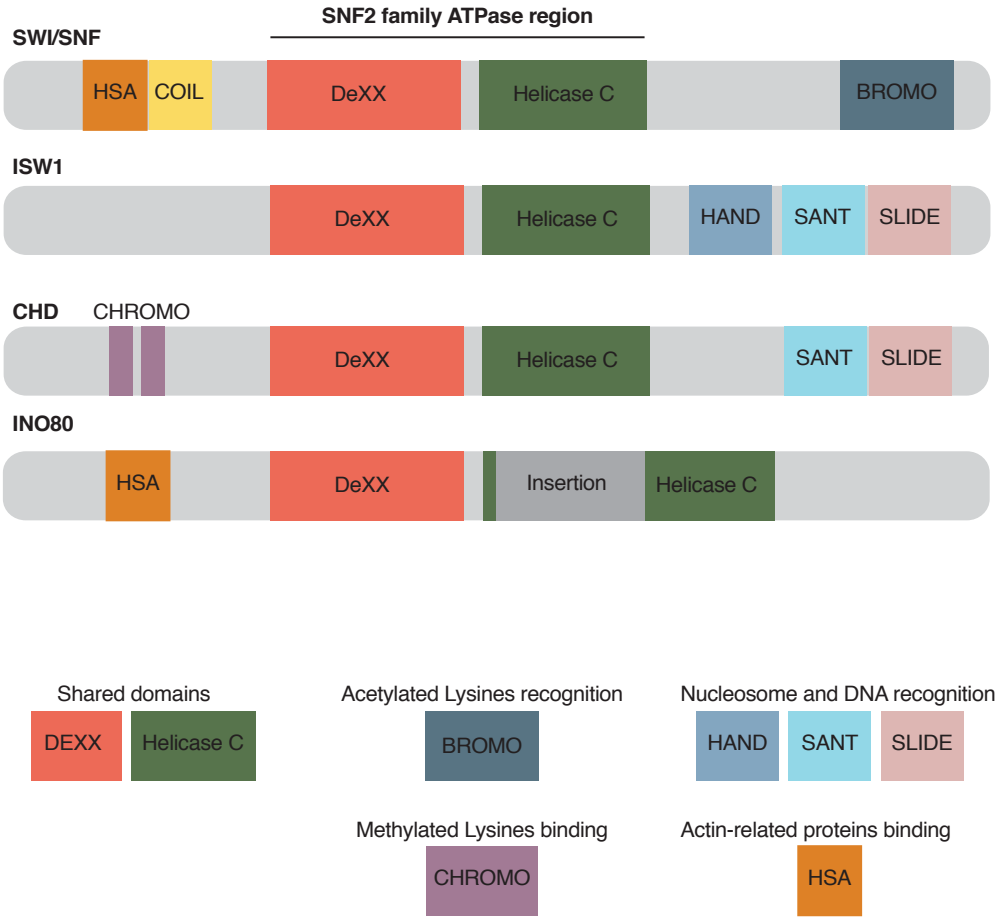


Figure 1.6 Chromatin remodellers.
Simplified representation of the domain organisation for the main families of chromatin remodeling ATPases. The HelicaseC and the DEXX domains are shared between all the classes. Some domains are specific for each class. The Bromo domain recognises acetylated lysines and is characteristic of the SWI/SNF family. Methylated lysines recognition is assured by the Chromo domain in the CHD family. HAND, SANT and SLIDE can recognise nucleosomes and DNA, and are characteristic of the ISWI family. Finally, the HSA domain is found in the INO80 family. Adapted from Moore et al, 2015

1.1.2. Function of chromatin remodelers

As briefly described above, the chromatin remodelling family of enzymes comprise a large variety of remodelling complexes and proteins. Despite having different specific roles, it is correct to say that these enzymes are responsible for modifying the state of the chromatin. The perturbation of the chromatin state is mainly aimed to expose a hidden site on the DNA or alter the composition of chromatin itself. To expose an underlying sequence of DNA, nucleosomes can either slide and thereby be repositioned to a neighbouring DNA, or evicted from the chromatin. Alternatively, unwrapping of the DNA around the histone octamer can expose a shorter sequence or allow inclusion of a new nucleosome (Maier et al., 2008; Morrison & Shen, 2009). Remodelers are also capable of modifying the composition of the chromatin by exchange of histone variants, such as the H3.3 at telomeres or ejection of a dimer (Clapier & Cairns, 2009; Längst & Manelyte, 2015).

Despite being difficult to classify the remodelers depending on their activity, it is possible to distinguish them depending on the presence of specific domains (Clapier & Cairns, 2009; Längst & Manelyte, 2015). In fact, SWI/SNF are mainly responsible for sliding and eviction of nucleosomes from the DNA, lacking any assembly properties. On the other hand, the NuRD complex (CHD family) is important for repression of transcription, whereas other complexes of the family are reported to be responsible for sliding of the nucleosome to promote transcription. Most of the ISWI complexes catalyse nucleosome spacing, compaction and formation of higher order structures, concomitantly with repression of the DNA sequence. Finally, the INO80 family is the only one showing helicase activity and INO80 and SWR1 were shown to be responsible for removing H2A-H2B dimers and replacing with the histone variants H2A.Z-H2B.

The existence of many remodelling complexes and the specificity of these proteins for any DNA suggest the existence of a more complex regulatory mechanism, resulting from the interplay of the DNA sequence, the context the DNA sequence is in, the

epigenetic marks occurring on the histones and, nevertheless, the specificity of the single remodeler. Not surprisingly, in fact, a study comparing NURD, (P)BAP, INO80 and ISWI genome-wide showed that each enzyme has a unique set of targets (Moshkin et al., 2012). We are yet far from fully understand how this specificity is determined and how the remodelers are capable of complementing each other's work to maintain genome stability and integrity.

An exception to these canonical remodelers is the protein LSH. Despite being classified as a member of the Sfn2-like group of remodelers, its chromatin remodeler activity in complex with CDC7A has only been recently proven (Jenness et al., 2018). Interestingly, however, it was shown to be crucial for normal levels and patterns of DNA methylation in plant and mammals.

1.5. Lymphoid Specific Helicase (LSH)

1.1.3. LSH, characteristics and structure

LSH, also known as HELLS, PASG and SMARCA6, is an ATP-dependent chromatin remodeler, member of the SNF2 superfamily of remodelers is DDM1 (decrease in DNA methylation 1), the plant homologue of mammalian LSH, was identified first (Vongs et al., 1993). In a screening in *Arabidopsis thaliana* aimed to identify mutants causing genome hypomethylation, deficiency in DDM1 in the plant was shown to cause 70% loss of DNA methylation (Vongs et al., 1993).

Due to its sequence similarity with the SNF2 remodelers (Figure 1.7), DDM1 was hypothesized to be a novel chromatin remodeler and its remodelling activity was demonstrated in a later study (Brzeski & Jerzmanowski, 2003; Jeddeloh et al., 1999).

The mammalian LSH was identified later, firstly in activated lymphocytes (Jarvis et

al., 1996). In that study, LSH was discovered while looking for a helicase responsible for the somatic recombination of the VJ locus, thereby the descriptive name given. However, the helicase activity of LSH has never demonstrated. It was then found in mouse thymus and testis and shown to be highly expressed in all dividing cells in the mouse embryo (Raabe et al., 2001). It is now well known that LSH is expressed in the adult mice as well.

In mouse, the *Lsh* gene is located on chromosome 19 and contains 24 exons. The gene is transcribed into an 821 amino acids protein. The amino acid sequence is well conserved between species and differs between mouse and humans for 16 additional amino acids at the N-terminus of the human protein (Briones & Muegge, 2012). A coiled-coil motif characterises the N-terminus of the protein, which contains the nuclear localisation signal (NLS) (Figure 1.7). The catalytic ATPase domain is organised as two lobes, both containing an ATP binding sequence, and contains 7 helicase motifs, characteristics of this superfamily of remodelers. The first lobe is characterised by a DeXX, or SNF2_N, motif and the second by an HELIC_C motif (Figure 1.7).

Most of the chromatin remodelling proteins form big protein complexes. On the contrary, it was shown that LSH is mainly found as a monomer (Myant & Stancheva, 2008). Interestingly, it is indeed capable of binding other proteins. More specifically, pull-downs experiments have proven a direct interaction between LSH and DNMT3B, via the CoiledCoil N-terminal domain of LSH (Myant & Stancheva, 2008). This interaction seems to work as a bridge for the recruitment of DNMT1, that can finally recruit the HDAC1 and HDAC2. These enzymes remove acetyl groups form lysines, allowing compaction of chromatin and therefore transcriptional repression of the region. Furthermore, the coiled-coil motif binds the E2F3 transcription factor, cooperating with its role in tumour progression (Von Eyss et al., 2012). More recently, it was shown that LSH and the zinc finger protein CDCA7 form a chromatin remodelling complex (Jenness et al., 2018). Altogether, these findings highlight that LSH requires other proteins, part of its interactome, to efficiently carry out its function.

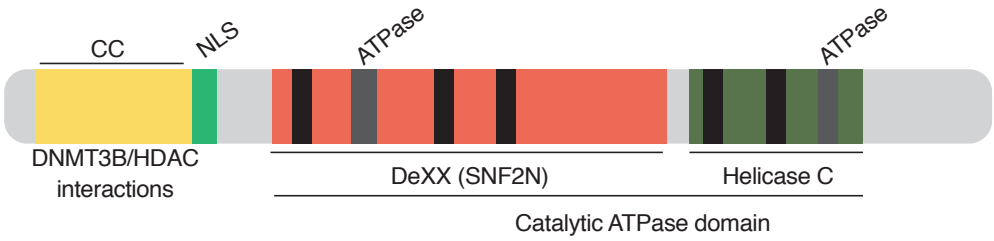


Figure 1.7 Domain organisation of chromatin remodeling ATPase LSH.
Schematic representation of the domain organisation of murine LSH. The N-terminal domain of the protein includes a Coiled-Coil (CC) motif (yellow), required for interactions with DNMT3B and recruitment of HDACs, and the nuclear localisation signal (NLS - green). The catalytic ATPase domain includes DeXX, or SNF2_N (red) and helicase motifs (black and grey), and Helicase C (green) and helicase motifs (black and grey).The two ATP binding sites (grey) bind ATP simultaneously.

1.1.4. Function of LSH, from plants to mammals

In plants, DDM1 was shown to reposition nucleosomes *in vitro* and its absence in *Arabidopsis* leads to hypomethylation at CpG and CpNpG sites, together with misexpression of retrotransposons, that disrupts expression of protein coding genes (Brzeski & Jerzmanowski, 2003; Vongs et al., 1993). Knock-outs in plants are characterised by a rearrangement of histone modifications, more specifically by the replacement of H3K9me2/3 by H3K4me, as a sign of euchromatin formation and eventually activation of gene expression. Moreover, defects in mitosis and in DNA damage response and morphological abnormalities in leaves shape and number have been described in *ddm1* mutant *A.thaliana* (Jeddeloh et al., 1999; Kakutani et al., 1995).

Despite having been named in mammals Lymphoid specific helicase, it has been ruled out any helicase activity of the protein. However, it was shown necessary for DNA methylation throughout the genome in mammals as well (Dennis, 2001; Zhu et al., 2006).

In fact, depletion of LSH in mouse leads to up to 50% loss of 5meC on a whole genome level. The hypomethylation is accompanied by rearrangement of histone modification. More specifically, increase in acetylation at H3 and H4 and local increased of H3K4ac were detected in LSH depleted cells (Dennis et al., 2001; Tao et al., 2011; Termanis et al., 2016). Studies have shown that cells lacking LSH show hypomethylation at both repetitive and non-repetitive sequences. Many studies have analysed methylation of specific loci in the genome and identified few that require LSH for correct methylation. More specifically, Yu and colleagues showed that the knockout (KO) of *Lsh* causes hypomethylation of long terminal repeats (LTR) of many retrotransposons, including satellites sequences, endogenous retroviral elements (ERVs) and IAPs. The hypomethylation leads to expression of these elements, thereby highlighting a role of LSH in protecting genome integrity (Huang et al., 2004; Yu et al., 2014a). Furthermore, a 2007 study showed that *Lsh* knockout induces expression of *HoxA6* and *HoxA7* genes, possibly because LSH is required at these loci to regulate DNMT3B binding and silencing of gene expression (Xi et al., 2007). However, misexpression of *Hox* genes is known for leading to homeotic defects and specifically *HoxA6* and *HoxA7* are involved in hematopoietic and ovarian epithelial cell cancers, defects that have not been detected in $Lsh^{-/-}$ mice. This suggests that the defects found in $Lsh^{-/-}$ MEFs in this study might be reversible and overcome *in vivo* during differentiation. More studies have shown importance of LSH in methylation of specific genes, such as the *Rhox* cluster (Donohoe et al., 2009; Fan, 2005; Myant et al., 2011; Ren et al., 2017). Furthermore, it was shown that reintroduction of a catalytically active LSH protein in KO MEFs can partially recover the loss of DNA methylation and determine re-silencing of misexpressed genes (Termanis et al., 2016). However, no whole genome analysis has been performed in the early development mouse, to determine all the loci in the genome that require LSH for correct methylation and thereby are found hypomethylated.

An interesting debate is whether LSH is required for *de novo* DNA methylation or maintenance, or maybe for both these processes. Transfection of an unmodified episomal DNA in $Lsh^{-/-}$ mouse embryonic fibroblasts (MEFs) was used to prove the role of LSH *in*

vitro in *de novo* methylation, but not maintenance (Lungu et al., 2015; Zhu et al., 2006). However, several studies have been published supporting the importance of LSH in DNA methylation maintenance, for instance in cancer (Dennis et al., 2001; Samuelsson et al., 2016) or specifically at repetitive elements in mice (Suzuki et al., 2008).

Similarly to DDM1, LSH depletion has an effect on histone modifications. More specifically, it determines perturbation of H3K4me3 and increase in acetylation of the histone tails (Yan et al., 2003). This correlates with the loss of DNA methylation and therefore of heterochromatin. Furthermore, pull downs experiments proving the co-immunoprecipitation of LSH and DNMT3B (Myant & Stancheva, 2008) suggested that the activity of LSH might be linked to that of the *de novo* methyltransferase. It is proposed that LSH recruits DNMT3B and determines an increase of its concentration locally, supporting the methyl transferase activity of the protein. However, it is important to highlight that LSH activity is not exclusively associated with DNMT3B. In fact, several studies have shown that LSH has a role independent of the methyl transferase in methylation of repetitive elements, in association with G9a/GLP HMT, but also in cancer progression (Dunican et al., 2013; X. He et al., 2016; Jiang et al., 2017; Joshi et al., 2011; Myant et al., 2011).

Despite the ATP-dependent chromatin remodelling activity of LSH being considered the main function of the protein (Ren et al., 2015), the most evolutionary conserved function of LSH is in DNA repair. Previous studies in budding yeast (Alvaro et al., 2007; Costanzo et al., 2010), which have no DNA methylation, but carry an LSH homologue, and in *A.thaliana* (Costanzo et al., 2010; Shaked et al., 2006) have proposed a role of LSH independent of its DNA methylation activity. A subsequent study in human fibroblasts and mouse showed that *Lsh*$^{-/-}$ cells are more sensitive to ionizing radiation (IR) and repair double strand breaks (DSB) less efficiently (Burrage et al., 2012). This impairment is mainly due to less stable and permanent phosphorylation of the H2AX histone variant in response to IR, which impairs the recruitment of the DSB repair machinery, finally determining absence of phosphorylation of the checkpoint CHK2 protein. In addition, the study proved that the catalytic activity of LSH is responsible for its role in DNA damage repair. Altogether these data showed that LSH can have a role independent of its activity

in DNA methylation, but it remains unclear how this could be linked to its role as a chromatin remodeler.

1.1.5. Characterisation of LSH Knockouts in mouse

To understand LSH function, generation of LSH-null cell lines and knockout mouse was necessary. Currently, there are two mouse models published in literature, carrying this deletion.

The first *Lsh*$^{-/-}$ mouse was generated by targeted deletion of exons 6 and 7 via homologous recombination (Geiman et al., 2001). Helicase domain I and II were replaced by a cassette containing neomycin resistance (neo) and thereby premature perturbation of transcription and translation prevented production of full length protein. Heterozygous mice for the deletion looked indistinguishable from their wild type littermates. However, the *Lsh*$^{-/-}$ mice, despite surviving through development, die few hours after birth. Only one out of 480 mice survived up to 5 days postnatally. Moreover, these mice showed 22% reduction in weight compared to the littermates and severe kidney failure. Studies performed on these mice, showed instability of the mitotic spindle and reduced proliferation (Tao Fan et al., 2003) and accumulation of H3K4 methylation (Yan et al., 2003). Due to the short survival time, no further conclusions could be drawn.

Subsequently, Sun and others generated a different KO mice, by replacing exons 10, 11 and 12, together with helicase domain II, III and IV, with a neo cassette (Sun et al., 2004). These deletions resulted in the production of an hypomorphic mutation. Once again, the heterozygous mice looked phenotypically similar to the wild type littermates and no hypomethylation was detected on a whole genome level. This approach resulted in higher percentage of survival after birth. However, only 40% of *Lsh*$^{-/-}$ mice survived up to few weeks postnatally. Nevertheless, the mice were severely affected by the absence of LSH. More specifically, they showed 25% weigh reduction at birth comparted to the wild

type littermates, growth retardation and a severe aging phenotype, manifested by hair loss, cachexia and reduced fat accumulation. Moreover, these mice exhibited lymphoid depletion, respiratory and kidney failure, osteoporosis and osteopenia. Analyses of *Lsh*$^{-/-}$ MEFs isolated from this strain showed increased expression of senescence-associated genes. Altogether, these data suggest the importance of LSH during early phases of development, despite not determining the exact time-frame of activity of the chromatin remodeler.

Recently, in the Stancheva lab, a new mouse model was generated (unpublished data, Chao Li). This new strain carries a conditionally reversible stop cassette inserted into intron 3 of the *Lsh* gene (Figure 1.8). This cassette contains a splicing acceptor (SA), a EGFP-neomycin resistance gene followed by a polyadenylation site at the 3' end. This cassette is flanked by inverted LoxP and Frt sites, that can be used to conditionally flip the cassette orientation and thereby re-establish normal expression of the locus. The cassette was integrated either into one or into both alleles in ES cells to produce lines with *Lsh*$^{off/+}$ or *Lsh*$^{off/off}$ genotype. The *Lsh*$^{off/+}$ or *Lsh*$^{off/off}$ ESC cell lines were further used for *in vitro* experiments. Protein expression analysis showed that LSH proteins was undetectable both in *Lsh*$^{off/off}$ ESCs and E12.5 embryos, compared to the heterozygous control without affecting the expression of the DNMTs (unpublished data – Chao Li). Mice generated from *Lsh*$^{off/+}$ ESCs develop and survive postnatally for up to 16 weeks. Nevertheless, they experience severe weight loss and develop neurological problems shortly after birth. Furthermore, these mice show DNA methylation loss comparable to previously generated *Lsh*$^{off/off}$ mice, including hypomethylation of centromeric repeats, IAPs and unique loci (unpublished data – Chao Li). Altogether, these experiments validate the mouse knockout generated in the Stancheva lab. The *Lsh*$^{off/off}$ mice generated in the lab constitute a great model to study *in vivo* the consequences of LSH deficiency after birth: they are efficiently depleted of LSH, and, compared to the other KOs mice, they also carry fewer perturbations in the *Lsh* loci, generated by the depletion.

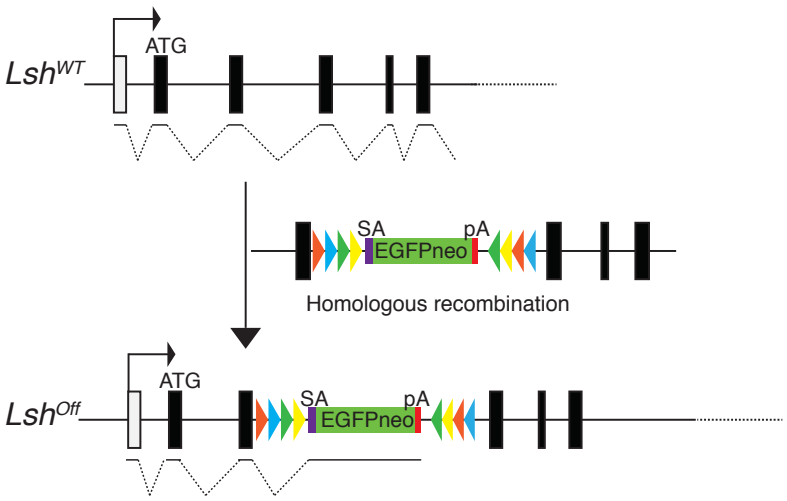


Figure 1.8 Domain organisation of the knock-out cassette.
Representation of the cassette introduced in the *Lsh* locus to produce the off allele. The off cassette was introduced in intron 3 by homologous recombination. Black boxes represents exons, SA = splicing acceptor; pA = polyadenylation site; coloured triangles represent (in order) LoxP, lox511, Frt and F3 sites adjacent to the cassette.

For this work, we used $Lsh^{off/+}$ or $Lsh^{off/off}$ ESC cells, that were not derived from mouse blastocysts, but were produced by homologous recombination from wild type ESCs. These cells provided a big advantage for the study of changes occurring during reprogramming events. In fact, DNA hypomethylation at unique and repetitive sequences, such as IAPs, occurs in absence of LSH during embryonic development and more specifically between E6.5 and E9.5. Since the $Lsh^{off/+}$ or $Lsh^{off/off}$ ESC we used in this study had not gone through embryonic development and reprogramming, it was possible to analyse the effects of LSH on DNA methylation maintenance and establishment, in the absence of developmental reprogramming that occurs early in mammalian development. Then, by inducing reprogramming of the DNA methylation, we could study how absence of LSH interferes in time in *de novo* and maintenance of DNA methylation and in differentiation.

Authors	Knock-out	Protein mutation	Phenotype
Geiman et al., 2001	Deletion of exons 6 and 7, via homologous recombination; replacement of helicase domain I and II by a neo resistance cassette.	Prevention of production of a full-length protein.	*Lsh*$^{+/-}$ mice: indistinguishable from wild type littermates. *Lsh*$^{-/-}$ mice: 1 out of 480 mice survived more than few hours after birth. Mitotic spindle instability and reduced cell proliferation (Fan et al., 2003), accumulation of H3K4me (Yan et al., 2003).
Sun et al., 2004	Replacement of exons 10, 11, 12 and replacement of helicase domain II, III, IV with a neo resistance cassette.	Hypomorphic mutation.	*Lsh*$^{+/-}$ mice: indistinguishable from wild type littermates. *Lsh*$^{-/-}$ mice: 40% of mice survived few weeks after birth; weight reduction at birth, growth retardation and aging phenotype.
Chao Li (unpublished)	Insertion of a conditionally reversible stop cassette in intron 3.	Premature stop of transcription and translation.	*Lsh*$^{off/+}$ mice: indistinguishable from wild type littermates. *Lsh*$^{off/off}$ mice: survive postnatally for up to 16 weeks; weight loss and neurological problems.

Table 3.LSH knock-outs.
Brief summary of the main knock-outs models for LSH in mice. The first works regarding the mouse model are listed in the first column. The approaches used to produce the KOs are indicated in the second and third columns. In the last column a brief description of the phenotipe of the mouse model is reported.

1.1.6. LSH misregulation and disease

DNA methylation is crucial to assure stability and integrity of the genome, thereby allowing normal function of the cell and, eventually of the whole organism. When removing LSH, mice cannot survive after birth, showing severe phenotypes, or, when surviving,

eventually develop disease. This suggest the importance of the protein for the development of the organism and the need for better understanding its role and mode of action.

Interestingly, not only depletion, but also mutations or overexpression of LSH were linked to disease, such as ICF, leukaemia and glioma (He et al., 2016; Jenness et al., 2018; Joshi et al., 2011; Keyes et al., 2011; Thijssen et al., 2015; Velasco et al., 2018; Xiao et al., 2017). These mutations are generally determinants of hypomethylation and subsequently misexpression of protein coding genes and repetitive elements, which plays an important role in determining the pathological phenotype.

An important example of disease linked to hypomethylation is the ICF syndrome. This is a rare recessive autosomal disorder characterised by severe immunodeficiency, which make patients affected more susceptible to infections of the respiratory and gastrointestinal tract, associated in most cases with facial abnormalities. One of the hallmarks of ICF is the hypomethylation of Centromeric regions, leading to instable and fragile chromosomes 1, 9 and 16 (Hulten, 1978). Only around 50% of patients affected by ICF syndrome are characterised by mutations in the methyl transferases and less than 30% in ZBTB24. In more recent years, mutations associated to this syndrome have been identified in CDCA7 and LSH/HELLS genes (Thijssen et al., 2015). This led to the characterizations of subgroups of patients, namely ICF1 (mutations in DNMT3B), ICF2 (mutations in ZBTB24), ICF3 (mutations in CDCA7) and ICF4 (mutations in LSH/HELLS). Importantly, Velasco and colleagues have shown that the hypomethylated sequences in ICF2,3,4 patients are different from the hypomethylated regions of patients with ICF1 (Velasco et al., 2018). This further supports the hypothesis of a partially separate role of LSH from DNMT3B, with LSH supporting DNMT3B for methylating only some genomic regions. Recently, the Funabiki lab has shown that CDCA7 and LSH work together as a remodelling complex (Jenness et al., 2018). In this work, they showed that CDCA7 facilitated LSH binding to chromatin and stimulates LSH remodelling activity, independently of its DNA-binding activity. Despite not been able to fully characterize the process, they showed the importance of the formation of this complex. Furthermore, they

suggest that mutations in the complex in ICF patients are most likely responsible for inability of recruitment of LSH to chromatin and therefore impossibility to methylate the DNA. However, further molecular studies are necessary to elucidate the mechanisms behind this disorder.

1.6. Open questions and aims of the project

In this brief introduction on chromatin modifications, modifiers and LSH, I tried to show the importance of epigenetic mechanisms for development and normal functions of cells. There are still many questions to be answered. In the Stancheva lab, we focused our attention on the chromatin remodeler LSH. Despite many studies of this protein have shown its importance in DNA methylation and involvement in other cell pathways, such as DNA repair, there is still a lot to clarify. The focus of my work was to address the following questions:

- Does the lack of LSH impair the establishment or the maintenance of DNA methylation? When is LSH required during mammalian development?
- Is DNA methylation at specific loci LSH dependent? If so, which loci require LSH activity?
- How does the absence of LSH lead to hypomethylation of the genome?

2. MATERIALS AND METHODS

2.1. MATERIALS

2.1.1. DNA analysis buffers

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

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

SSC buffer: 175.3 g NaCl2, 88.2 g $Na_3C_6H_5O_7$ were used for 1 L of solution. pH was adjusted to pH 7 with HCl.

Denaturing buffer: 1M NaOH, 0.5 EDTA

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

PCR buffer IV (10X): 200 mM $(NH_4)_2SO_4$, 750 mM (Tris-HCl pH 8.8), 0.1% (v/v) Tween® 20, 15 mM $MgCl_2$.

Special PCR buffer: 166 mM $(NH_4)_2SO_4$, 670 mM Tris pH 8.0 and 100 mM ß-mercaptoethanol.

Sodium Acetate (3M): 40.8g of sodium acetate in 80ml of H_2O. pH was adjust to 5.2 with acetic acid.

Sequencing buffer (2.5X): 20 mM Tris-HCl pH 8.0 and 5 mM $MgCl_2$.

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

2.1.2. Protein analysis buffers

Nuclear extraction buffer (NE1): 20mM HEPES pH7.5, 10mM KCl, 5mM $MgCl_2$, 0.1% (v/v) Triton X-100, 20% (v/v) glycerol, 0.5mM DTT and Protease Inhibitor Cocktail (1X) (SIGMA-P8340), added prior to use.
Sample buffer (4X): 200 mM Tris-HCL pH 6.8, 8% SDS, 20% Glycerol, bromophenol blue. 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.
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.
Running buffer for Western blotting: 25 mM Tris, 250 mM Glycine and 0.1 % (w/v) SDS.
Transfer buffer for Western blotting: 25 mM Tris and 250 mM Glycine.
Ponceau S staining solution: 1% (v/v) glacial acetic acid, 0.5% (w/v) Ponceau S.

2.1.3. HIS6-CxxC-AVI protein production

Lysis Buffer: 50 mM Tris pH 8.0, 300 mM NaCl. 1 mM PMSF, 1/2 tablet Roche protease inhibitor cocktail and 10U/ml benzonase (12 µl) were added prior to the use.
Ni-NTA Binding Buffer: 50 mM Tris pH 8.0, 300 mM NaCl and 1 mM Imidazole.
Ni-NTA Low Salt Wash Buffer: 50 mM Tris pH 8.0, 300 mM NaCl, 10 mM Imidazole, 0.1% Triton-X100. 1 mM PMSF was added prior to the use.
Ni-NTA High Salt Wash Buffer: 50 mM Tris pH 8.0, 1M NaCl. 1 mM PMSF was added prior to the use.
Ni-NTA Elution Buffer: 50 mM Tris pH 8.0, 300 mM NaCl, 300 mM Imidazole. 1 mM PMSF was added prior to the use.
Dialysis Buffer: 50 mM Tris pH 8.0, 300 mM NaCl 0.5 mM PMSF was added prior to the use.
Biotinylation Buffer: 20 mM Tris pH 8.0 and 250 mM Potassium glutamate.

Biomix (5X): 50 mM ATP, 50 mM MgOAc and 250 mM D-biotin.

BC150: 20 mM HEPES pH 7.9, 150 mM KCl, 0.5 mM DTT and 10% v/v glycerol.

2.1.4. Biotinilated CxxC affinity purification (BioCAP) sequencing buffers

CAP100 Buffer: 100 mM NaCl, 0.1% Triton X-100, 20 mM HEPES-KOH pH 7.9 and 12.5% Glycerol.

CAP300 Buffer: 300 mM NaCl, 0.1% Triton X-100, 20 mM HEPES-KOH pH 7.9 and 12.5% Glycerol.

CAP500 Buffer: 500 mM NaCl, 0.1% Triton X-100, 20 mM HEPES-KOH pH 7.9 and 12.5% Glycerol.

CAP700 Buffer: 700 mM NaCl, 0.1% Triton X-100, 20 mM HEPES-KOH pH 7.9 and 12.5% Glycerol.

CAP1000 Buffer: 1000 mM NaCl, 0.1% Triton X-100, 20 mM HEPES-KOH pH 7.9 and 12.5% Glycerol.

2.1.5. Antibodies

The antibodies used in this thesis work were the following:

Antibody anti	Species	Use	Producer
DNMT3A	Mouse	Western Blotting	Abcam
DNMT3B	Rabbit	Western Blotting	ThermoFisher
DNMT1	Rabbit	Western Blotting	Santa Cruz
LSH	Mouse	Western Blotting	Santa Cruz
HDAC1	Rabbit	Western Blotting	Santa Cruz
Secondary ab	Mouse or Rabbit	Western Blotting	LI-COR Biosciences
5meC	Mouse	Dot Blot	Active Motif
ssDNA	Mouse	Dot Blot	Enzo Life Sciences

2.1.6. Primers

The DNASTAR Lasergene software was used for primer designing. These were then purchased from Eurofins. The lyophilised primers were then diluted with ddH2O for stocking to 100 μM and stored at -20°C.

The primers used in this thesis work were the following:

PCR

Primers for BioCAP-PCR analyses span CG rich promoter regions of the genes. Alignment maps are included in chapter 7. Appendix

Fabp7	5' - ATTGGCTTTTTGCCCGCTTC - 3' 5' - AACTGGAGGAACTCGGGTCT - 3'
Msn	5' - GGGGTTTGTAAAGTCGTGGC - 3' 5' - TTTGGCTGGAAACTGTCGGG - 3'
ActinB	5' - AAGTGGCCTTGGAGTGTG - 3' 5' - CAAGGAGTGCAAGAACACAGC - 3'

qRT-PCR

Primers for qRTP-PCR analyses span exon-exon junctions. Alignment maps are included in chapter 7. Appendix

GAPDH	5' - TGGTATCGTGGAAGGACTCATGAC - 3' 5' - ATGCCAGTGAGCTTCCCGTTCAGC - 3'
IAPs	5' - ACTAACTCCTGCTGACTGG - 3' 5' - TGTGGCTTGCTCATAGATTAG - 3'
DNMT3A	5' - ACGGCAGAATAGCCAAGTTCA - 3' 5' - TCAGTGCACCACAGGATGTC - 3'
DNMT3B	5' - TGTGCCAGACCTTGGAAACC - 3' 5' - TGTCTCCCTTCATTGTTTCCTGA - 3'
G9A	5' - AGTGTAACCAGGCATGCTCC - 3' 5' - TGCAGTAAACCTCGCCATCC - 3'
PRDM14	5' - AAGCACCTGAAGTACACGCC - 3' 5' - AGGTTGAACACAGGTAGGGC - 3'

Bisulfite-converted DNA

Primers for Bisulfite-PCR analysis span 5'UTR of IAP elements.

IAPs (5' UTR)	5' - TTGATAGTTGTGTTTTAAGTGGTAAATAAA - 3' 5' - AAAACACCACAAACCAAAATCTTCTAC - 3'
IAPs nested	5' - TTGTGTTTTAAGTGGTAAATAAATAATTTG - 3' 5' - CAAAAAAAACACACAAACCAAAAT - 3'

Sequencing

Primers for Bisulfite-converted DNA sequencing analysis align to the commercially available cloning vector.

pJET Seq-pJET-F	5' - GCCTGAACACCATATCCATCC - 3'
IAPs nested	5' - TTGTGTTTTAAGTGGTAAATAAATAATTTG - 3' 5' - CAAAAAAAACACACAAACCAAAAT - 3'

2.2. METHODS

2.2.1. CELL CULTURE

2.2.1.1. Embryonic stem cells

Lsh$^{off/+}$ and *Lsh*$^{off/off}$ ESCs were cultured either in serum-medium or 2i-medium. The formulations were as follow:

- Serum medium: Minimum Essential Media with high glucose, 10% v/v Fetal Bovine Serum (Sigma), 0.1 mM nonessential aminoacids (Thermo Fisher Scientific), 1 mM sodium pyruvate (Thermo Fisher Scientific), 1μM β-mercaptoethanol (Thermo Fisher Scientific), 1% v/v Penicillin-Streptomycin-Glutamine (Thermo Fisher Scientific) and homemade LIF;

- 2i medium: Neurobasal media (Gibco), DMEM/F12 medium (Gibco), 0.5X N2 supplement, 0.5X B27 supplement, 7.5% BSA (Sigma), 1% v/v Penicillin-Streptomycin-Glutamine (Thermo Fisher Scientific), 1mM PD0325901 (System Biosciences), 3mM CHIR99021 (System Biosciences), 1.5X 10-4M monothioglycerol and homemade LIF.

For the facultative heterochromatin depletion experiment, *Lsh*$^{off/+}$ and *Lsh*$^{off/off}$ ESCs were maintained in the same media stated above, with the addition of UNC0638 (Sigma, U4885) at 500 nM final concentration. The drug was added at every time the medium was changed or cell were passaged.

When grown in serum medium, cells were grown on plates or flasks, coated with 0.1% gelatine and maintained at 85%-90% confluency. Every other day, cells were either harvested or passaged, as follow: after washing the cells with warm 1X PBS, cells were detached with 0.1% warm trypsin (Thermo Fisher Scientific); trypsin was then inactivated with medium and seeded in a new plate or flask. When not passaged, or harvested, the medium of the cells was changed. The same protocol was applied when cells were treated with UNC0638.

When grown in 2i medium, cells were grown on plates or flasks, coated with 0.1% gelatine and maintained at 85%-90% confluency. Every other day, cells were either harvested or passaged, as follow: after washing with warm 1X PBS, cells were detached with 1X TrypLE Express (Gibco); the enzyme was deactivated with BSA and the cell suspension was transferred to a sterile tube for centrifugation; finally, the supernatant was discarded and the cell pellet was resuspended in warm medium, to then be seeded to a new plate or flask. When not passaged or harvested, the medium of the cells was changed.

Following harvesting of the cells, these were resuspended in 2ml of 1X PBS and centrifuged in a table top centrifuge. The supernatant was then removed and the cell pellets were either temporary stored at -70ºC or processed immediately.

2.2.1.2. Embryonic bodies

Differentiation of ESCs into EBs was done as follow: ESCs were harvested and counted; $4*10^6$ cells were then seeded in Petri dishes in a volume of 14 ml of serum medium, without LIF. At day 4, the medium was changed and replaced with EBs medium, with N2 supplement. At day 6, the medium was changed and replaced with EBs medium with N2 and retinoic acid.

After 2 days, ESCs were already forming EBs and the medium was changed by transferring the cells suspension to a sterile tube and waiting until the cells settled at the bottom of the tube; the medium was then removed using a 3 ml sterile Pasteur pipette and cells were washed with 1X PBS. The PBS was removed after the cells settled once again at the bottom of the tube and replaced by 14 ml of warm medium. The same procedure was followed at day 4 and 6 of culture.

At day 4 and 8, collection of the EBs was necessary: these were resuspended in 2 ml of 1X PBS and centrifuged in table top centrifuge. The supernatant was then removed and the cell pellets were either temporary stored at -70ºC or processed immediately.

2.2.1.3. Bacterial cells

Bacterial cells used (*E.coli* DH5α) were grown in Liquid Lysogeny Broth (LB) medium at 37ºC shaking or on LB agar in Petri dishes at 37ºC.

2.2.2. NUCLEIC ACIDS ANALYSIS

2.2.2.1. Genomic DNA extraction

Cell pellets were digested overnight at 55°C in 1X TE buffer, with 0.1% SDS and 200 μg/ml Proteinase K (Thermo Scientific). Nucleic acids were then extracted with phenol:chlorophorm:isoamyl alcohol (25:24:1) and subsequently with chlorophorm. 0.3M NaOAc (pH 5.5) in 1 time volume of 100% isopropanol or 3 times volume 100% EtOH was used to precipitate the genomic DNA. The DNA pellet was washed with 70% cold EtOH and finally resuspended in 1X TE buffer. Treatment of the DNA suspension with RNAse A for 30 minutes at 37°C assured digestion of left over RNA. Samples were stored at 4°C.

2.2.2.2. Detection of 5meC - DNA immuno dot blot

Genomic DNA was sonicated to 250bp and denatured for 10 minutes at 100°C. Cold 2M ammonium acetate was used to stop the denaturing reaction. A sheet of filter paper and a nitrocellulose membrane were placed in a 96-wells dot blot apparatus (Bethesda Research Laboratories), the denatured DNA was loaded on the membrane in duplicate and then UV cross-linked. Membrane was blocked with 4% skimmed milk, 0.1% Tween20 in 1X TBS buffer. Twin membranes were probed with primary antibody against either 5meC or ssDNA (diluted 1:1000 in milk) and incubated shaking overnight at 4°C. The membranes were washed three times in 1X PBS with 0.1% Tween20, blocked and incubated with secondary antibody (donkey IR anti-mouse 800 or donkey anti-rabbit IR 700, Li-COR Biosciences) at room temperature for minimum 1 hour. After a final wash with 1X PBS the membranes were scanned with the LI-COR Odyssey Scanner.

Intensity of the 5meC dots was calculated using the Image Studio software and normalised to the intensity of the ssDNA dots detected on the twin membrane.

2.2.2.3. Detection of 5meC – High Pressure Liquid Chromatography (HPLC)

After extraction and purification of gDNA, HPLC analyses were carried out by Dr. Ramsahoye as previously described (Ramsahoye, 2002). Triplicate measurement for each sample was performed and quantifications were calculated as area under each peak.

2.2.2.4. Detection of 5meC - Bisulfite sequencing

2 μg of DNA were converted using the EpiTect Bisulfite Kit (QIAGEN), following the manufacturer's instructions. After conversion and amplification of the converted DNA, this was cleaned and the eluate was used for PCR for IAPs sequence, using the following conditions:

95°C for 5 min

95°C for 30sec

55°C for 30sec

72°C for 30sec → 2x22 cycles

72°C for 10min

2 μl of PCR product were then used in a nested PCR reaction, using the following conditions:

95°C for 5min

95°C for 30sec

50°C for 30sec

72°C for 30sec → 2x26 cycles

72°C for 10min

The whole PCR product was run on a 1% agarose gel and purified using the GeneJet Gel extraction kit (Thermo), following the manufacturer's instructions. 12.5ng of purified PCR product were ligated O/N in a blunt vector, using the CloneJet PCR cloning

kit (Thermo), following manufacturer's instructions. 100 μl of *E.coli* DH5α were transformed with half of the ligation reaction and incubated at 37°C O/N. To test the successful transformation of the bacteria, a colony PCR was performed, using the following conditions:

95°C for 5min

95°C for 30sec

60°C for 30sec

72°C for 45sec → 2x26 cycles

72°C for 10min.

PCR products were then run on a 1%agarose gel to screen for positive clones. These were then cleaned of the PCR reaction solution using 10 μl of a clean-up solution. Samples were incubated for 20 min at 37°C and then for 15 min at 80°C. Finally, a sequencing reaction was performed, using the following conditions:

95°C for 1min

95°C for 30sec

50°C for 30sec

60°C for 4min → 2x25 cycles

16°C for 120min.

Sequencing results were aligned using the DNA Lasergene software and the bisulfite converted sites were analysed using the BISMA software for repetitive sequences.

2.2.2.5. Detection of 5meC - Biotinilated CxxC Affinity Purification (BioCAP) sequencing

BioCAP was performed as previously described by (Blackledge et al., 2012) and

optimised with the help of Dani Wicaksono (Buonomo's lab) (Figure 5.1, page 106). In brief: *E.coli* BL21 were transformed with pHis-CxxC-Avi plasmid and 18kDa expressed protein was purified using Lysis Buffer and French® Pressure Cell Press. Protein was purified from the protein lysate using His-Bind Ni-NTA beads. His6-tag was removed by cleavage with His6-TEV protease O/N at 4 °C and the CxxC-Avi was biotinylated using Biotinylation Buffer (20 mM Tris pH 8.0 and 250 mM Potassium glutamate) and 5X Biomix (50 mM ATP, 50 mM MgOAc and 250 mM D-biotin). Once the CxxC-Bio was ready, the BioCAP reaction could be performed. 80 ng/µl of genomic DNA was sonicated in a Diagenode Bioruptor Twin to 100 – 300 bp long fragments. DNA was than quantified with Qubit BR dye on a Qubit fluorometer and dissolved in CAP100 buffer to a concentration of 17.5ng/µl.

25 µl Neutravidin beads (Sera-Mag Speedbeads – GE Life Science) were incubated with 25 µg of total Bio-CxxC. The beads-Bio-CxxC was then used for precipitation of sonicated gDNA at a maximum concentration of 17.5 ng/µl. After 1 hour of incubation at 4 °C of gDNA and beads-Bio-CxxC, these were washed twice in CAP100 buffer and eluted in CAP300 buffer. This last step was repeated for elution in CAP500, CAP700 and CAP1000 buffers. The gDNA was purified from the beads by spin-column based DNA/PCR product purification kit (Invitrogen Purelink), following manufacturer instructions. Low salt (CAP300 and CAP500 eluate) and high salt (CAP700 and CAP1000 eluate) were pulled and only high salt fraction was used for sequencing analysis.

Libraries for each sample to be sequenced were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina, starting with 0.5ng of eluted DNA.

Analysis of the sequencing results were performed in collaboration with Shaun Webb, as follow: 75b paired-end reads were aligned to the mouse mm10 genome with bwa mem version 0.7.5. We filtered the alignments to only include properly paired reads and to remove those that map to blacklisted regions identified by the ENCODE project. We further removed duplicate reads with Picard MarkDuplicates version 1.107. To account for varying read depths we used deepTools version 2.5.1 to create bigWig files normalised

by RPKM (reads per kilobase per million reads) for visualisation and further analysis. We employed Macs2 callpeak version 2.1.1 to characterise regions of non-methylated DNA within each sample and used the Bioconductor R package diffBind to identify regions of differential methylation between samples.

Differential BioCAP-seq peak co-ordinates were intersected with regions of repeat families classified by the repeatMasker website. The significance of the number of overlaps was determined by permutation analysis using the peakPermTest function in ChIPPeakAnno with the total set of BIOCAP-seq peaks from all samples used as the background model. The background was randomly sampled 1000 times to select the same number of peak regions as the test set and to build a distribution of expected overlaps with repeat families. A p-value was calculated by comparing the mean of the expected distribution against the observed value and peak sets with a p-value <= 0.05 were deemed to have a significant increase in overlap.

2.2.2.6. RNA extraction, reverse transcription and quantitative RT-PCR

When at 90 to 95% confluency, cells were harvested and collected in 2ml Eppendorf tubes. 1 ml of TRIzol Reagent (Thermo Fisher) was then used to resuspend the cell pellet. This was finally stored at -70°C. The RNA was extracted accordingly to manufacturer instruction. Briefly, aqueous phase was extracted with chloroform and nucleic acids were precipitated with 100% isopropanol. The RNA was then washed twice in cold 70% EtOH and air dried. The RNA pellet was then diluted in RNAse free water and treated with Ribolock RNAse inhibitor and DNAse I (Thermo Fisher) for 20min at 37°C, followed by heath inactivation for 5min at 95°C. To test for the quality of the RNA, 300ng of RNA were run on a 1.5% agarose gel, 10min at 90V.

For the reverse transcription reaction, 2 μg of RNA were used, with 1μl of Oligo(dT) primers, 0.1M DDT, 10mM dNTPs, Ribolock RNAse inhibitors and 1μl of Superscript III

reverse Transcriptase (Thermo Fisher). The produced cDNA was finally diluted in RNase free water and stored at -20°C.

Analysis of the gene expression was performed by qRT-PCR in a Roche 480 Lightcycler. SYBR Green master mix (Roche) and 2.5 μM primers were used according to manufacturer's instructions. To ensure the efficiency and specificity of the primers, melting curves for each primer pair was analysed. Gene expression was then analysed using the double Delta Ct analysis, relative to the housekeeping gene *Gapdh* and then normalised to Day 0.

2.2.2.7. Microccal Nuclease (MNAse) Assay

20 million cells were harvested, trypsinised and washed in PBS 1X. Cells were then crosslinked for 10 minutes using 1/10 volume of 10X crosslinking buffer and 36.5% formaldehyde to 1% final concentration. The reaction was stopped by addition of 2.5M glycine and cell suspensions were centrifuged for collecting cell pellet.

Cell pellet was homogenised in 500 µl of douncing buffer using a dounce homogeniser. 25 units of enzyme (300U/µl, Thermo Fisher) were used for each sample, which was incubated for 0, 3, 6, 9 and 12 minutes. The reaction was quenched using EDTA to 5mM final concentration. Samples were reverse crosslinked O/N at 65°C, with Proteinase K and 1%SDS.

Digested and decrosslinked samples were purified using a spin-column based DNA/PCR product purification kit (Invitrogen Purelink) and eluted in 35 µl of TE buffer. The digestion was then analysed using 1.5% agarose gel.

2.1.1. PROTEIN ANALYSIS

2.1.1.1. Protein extraction

Cells were harvested and cell pellet was resuspended in 1ml Hypotonic NE1 buffer and disrupted in a Dounce Homogeniser. Nuclei were precipitated by centrifugation at 4°C for 5 min at 4500rpm and resuspended in NE1 buffer with benzonase. NaCl was added to a final concentration between 200 and 500mM. Nuclear matrix and pellet were separated by centrifugation at 4°C for 15 min at 13000rpm.

Protein concentration was measured using the Bicinchoninic acid assay (BCA) and protein extract was then stored at -70°C.

2.1.1.2. Western blotting

Nuclear extract was loaded on an 8% separating gel and then transferred on a nitrocellulose membrane for 1 hour at 400mA. Membranes were blocked with 4% milk, 0.1% Tween20 in TBS buffer and probed over night with primary antibody against LSH (Santa Cruz), DNMT3B (Abcam), DNMT3A (Abcam), DNMT1 (Santa Cruz) and HDAC1 (Santa Cruz). After three washes in PBST (0.1% Tween20), membranes were blocked and incubated with secondary antibody (donkey IR anti-mouse 800 or donkey anti-rabbit IR 700, Li-COR Biosciences) at room temperature for at least 1 hour. Membranes were finally washed with PBS and scanned at the LI-COR Odyssey Scanner. For histone analyses, PVDF membranes were preferred. Ethanol was used for activating PVDF membranes and was added to the transfer buffer, to facilitate the process.

3. DNA methylation dynamics *in vitro*

3.1. Introduction

Mammalian development is a complex and not fully understood process. The formation of the new individual is linked to changes occurring on a genome-wide scale and epigenome level. In fact, the epigenome is dramatically reprogrammed: immediately after the formation of the zygote, DNA methylation is erased, to then be re-established after 3.5 days of development (He et al., 2011; Howlett & Reik, 1991; Li, 2002; Reik et al., 2001). Although a lot is known, it is still not completely clear how these changes occur and all the players involved in the process. Moreover, some loci escape reprogramming and do not lose and gain methylation at the same pace as the rest of the genome. These are imprinted regions and repetitive elements such as the type-II intracisternal A-particle (IAP) elements (Howlett & Reik, 1991; Smith et al., 2014). The difficulties in understanding this process are increased by technical limitations. In fact, obtaining enough material to study the first steps of development is challenging and time consuming. On the other hand, *in vitro* models are not very reliable for such studies and, although single cell analyses are now a possibility, this is not yet a routine approach in every lab.

To overcome these limitations, new *in vitro* approaches have been developed. Embryonic stem cells (ESCs) derived from E3.5 blastocyst stage embryos have for long been used for *in vitro* studies to investigate the naïve state of pluripotency, characteristic of this point in development. ESCs are maintained in culture using foetal bovine serum and leukaemia inhibitory factor (LIF), which stops them from differentiating (Smith et al., 1988; Williams et al., 1988) by activating the JAK-STAT signalling pathway (Niwa et al., 1998). Although these cells are capable of self-renewal and differentiation, high levels of DNA

methylation characteristic of much later developmental stages highlighted that these cells do not properly represent the stage of development that they are derived from. Mainly, in 2008 Ying et al showed that ESCs maintained in culture using serum are characterised by a very variable methylation profile. This is due to the oscillating pluripotency state of the cells: within the same cell culture, the cells show different levels of expression of pluripotency genes, such as Oct4, driven by variable response to the Fibroblast Growth Factor (FGF) signalling, which is contained in the serum used to maintain the mESCs. As a result, this culturing system imposes a high, on average, level of methylation in the cell population, when compared to the embryonic stage the cells derived from. The same group suggested the use of a different culture system enabling to maintain the naïve pluripotent state of the cells. This alternative system involves the use of supplements to substitute for nutrients in the serum and the addition of 3 small molecule inhibitors (3i), CHIR99021, PD184352 and SU5402, targeting respectively GSk3ß in the WNT cascade, MEK1/2 in the FGF cascade and finally the FGF receptor tyrosine kinase (Ying et al., 2008). Cells cultured in 3i medium were shown to be more homogeneous within the population, expressing stable levels of pluripotency factors and finally shown to have a much lower DNA methylation level. In the same year, it was proposed that the use of only two of these inhibitors, CHIR99021, PD184352, (2i) was sufficient to drive the ESCs to naïve pluripotency (Silva et al., 2009)**.** Although mESCs maintained in both conditions, 3i and 2i, can be considered pluripotent, it is now commonly stated that the mESCs in 2i medium can be categorised as naïve pluripotent, whereas mESCs in serum medium are known as primed pluripotent ESCs.

The mechanism leading to hypomethylation in presence of the 2 small molecules inhibitors is yet unclear. However, it is hypothesised that it could be determined by an interplay between passive and active demethylation and a reduced *de novo* DNA methylation activity, that overall lead to depletion of the mark. The altered signalling pathways have been shown to lead to the upregulation of transcriptional and epigenetic regulator PR Domain-Containing Protein 14 (PRDM14) (Okashita et al., 2014). On one hand, PRDM14 downregulates the *de novo* methyl transferases *Dnmt3a* and *Dnmt3b*

(Leitch et al., 2013), leading to impairment of the *de novo* DNA methylation. On the other hand, PRDM14 also targets TET1 and TET2 and enhances their recruitment to the designated loci, therefore inducing active DNA demethylation (Okashita et al., 2015). Finally, it was shown that GLP is repressed by PRDM14 (Leitch et al., 2013). This protein, in complex with G9a, is crucial to direct UHRF1 to the replication fork, via methylation of the DNA ligase 1 (Ferry et al., 2017), and consequently regulating maintenance of DNA methylation. Hence, when PRDM14 is upregulated in 2i, this could lead to an impairment of the maintenance of the DNA methylation, and therefore a passive demethylation (Leitch et al., 2016) (Figure 3.1).

Importantly, it was shown that in the ESCs genome some regions are relatively resistant to hypomethylation, namely the Imprinting Control Regions (ICR) and the IAPs, while partial loss is reported for the major satellites (Ficz et al., 2013).

We therefore decided to use the 2i culturing system to investigate LSH-dependent changes in DNA methylation that normally occur during the early stages of embryonic development. In particular, by keeping the cells first in serum medium and then switching them to 2i medium, the effects of lack of LSH on the demethylation and maintenance of methylation could be studied. Then, by changing the culturing conditions to serum-containing medium, the role of LSH in *de novo* DNA methylation could be explored. Finally, taking the cells to the first stages of the differentiation by triggering the formation of Embryonic Bodies (EBs), allowed to better understand the involvement, if any, of LSH in DNA methylation changes accompanying differentiation.

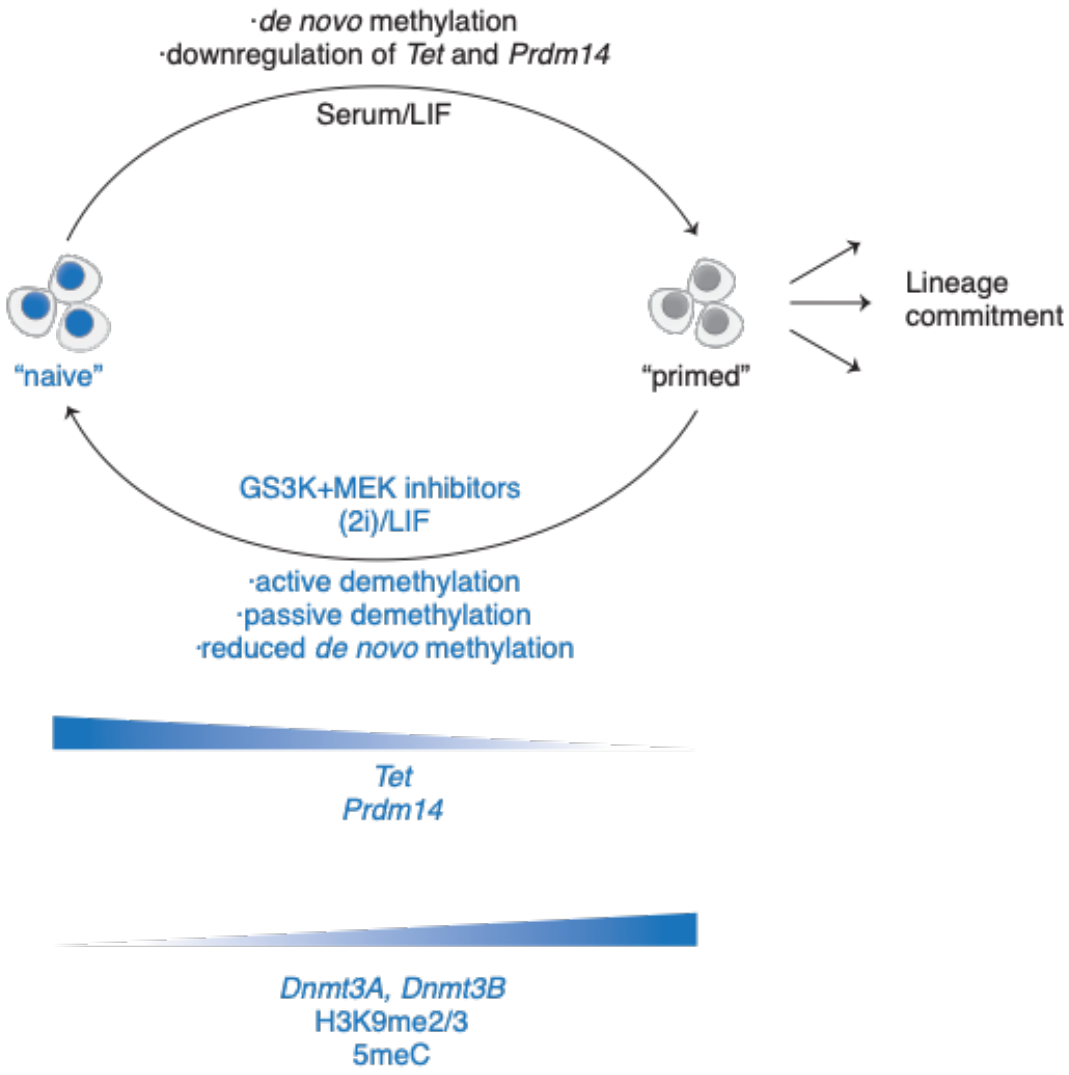


Figure 3.1 Naive and primed mESCs represent different states of pluripotency *in vitro*.
Schematic representation of the main differences occurring between mESCs maintained in culture with GS3Ki and MEKi (2i), or "naive" pluripotent, and ESCs maintained in serum, or "primed" pluripotent.

3.2. Establishment of a cell culture system to mimic DNA methylation dynamics of early embryonic development

First of all, the 2i culturing system had to be established in the Stancheva lab and tested. Heterozygous ESCs for *Lsh,* carrying the cassette on only one allele (*Lsh*$^{off/+}$ ESCs), were used to optimise the culturing conditions. Cells were first maintained in serum-containing medium. Then, the cells were switched to the 2i-medium and kept for up to 16 days in this medium, determining the establishment of naïve state of pluripotency. After that, the naïve pluripotent mESCs where once again switched to serum-containing medium, determining the subsequent activation of the *de novo* methylation machinery, and maintained for a maximum of 16 days in this medium. This time was later decreased to 4 days, in order to more closely mimic the short time existing *in vivo* before the beginning of the differentiation stage of development. Finally, the primed mESCs were differentiated into EBs, by LIF withdrawal and addition of retinoic acid, to start the commitment to neuronal lineage (Figure 3.2). DNA methylation, RNA expression and protein expression analyses were performed every two days of culture, corresponding to the passaging of the cells.

The cells reacted well to the medium containing the 2 inhibitors, showing no unexpected death or severe growth arrest. Immediately after switching to the 2i-medium, the cells were dividing slightly slower compared to the mESCs in serum-containing medium, but eventually the difference was minimised. As previously shown (Ying et al., 2008), cells were forming tight colonies on the plate, rather than spreading as typically ESCs do when growing in serum supplemented medium (Figure 3.2 B and C). Furthermore, while some differentiated cells were visible in the serum-containing culture, all the cells in 2i-medium showed typical features of pluripotent cells, such as rounded shape, pigmented cytoplasm or growth in small colony-shaped aggregates. This was an indication of the naïve ground state of pluripotency of the cells in 2i-medium. When cells were switched back to serum-containing medium, the initial morphology was re-established already after 2 days, showing that changes happening in the cells were easily

and quickly reversible (Figure 3.2 B).

Since the culturing system worked well with these cells, we decided to proceed and analyse the DNA methylation and gene expression of the main players involved in this regulation.

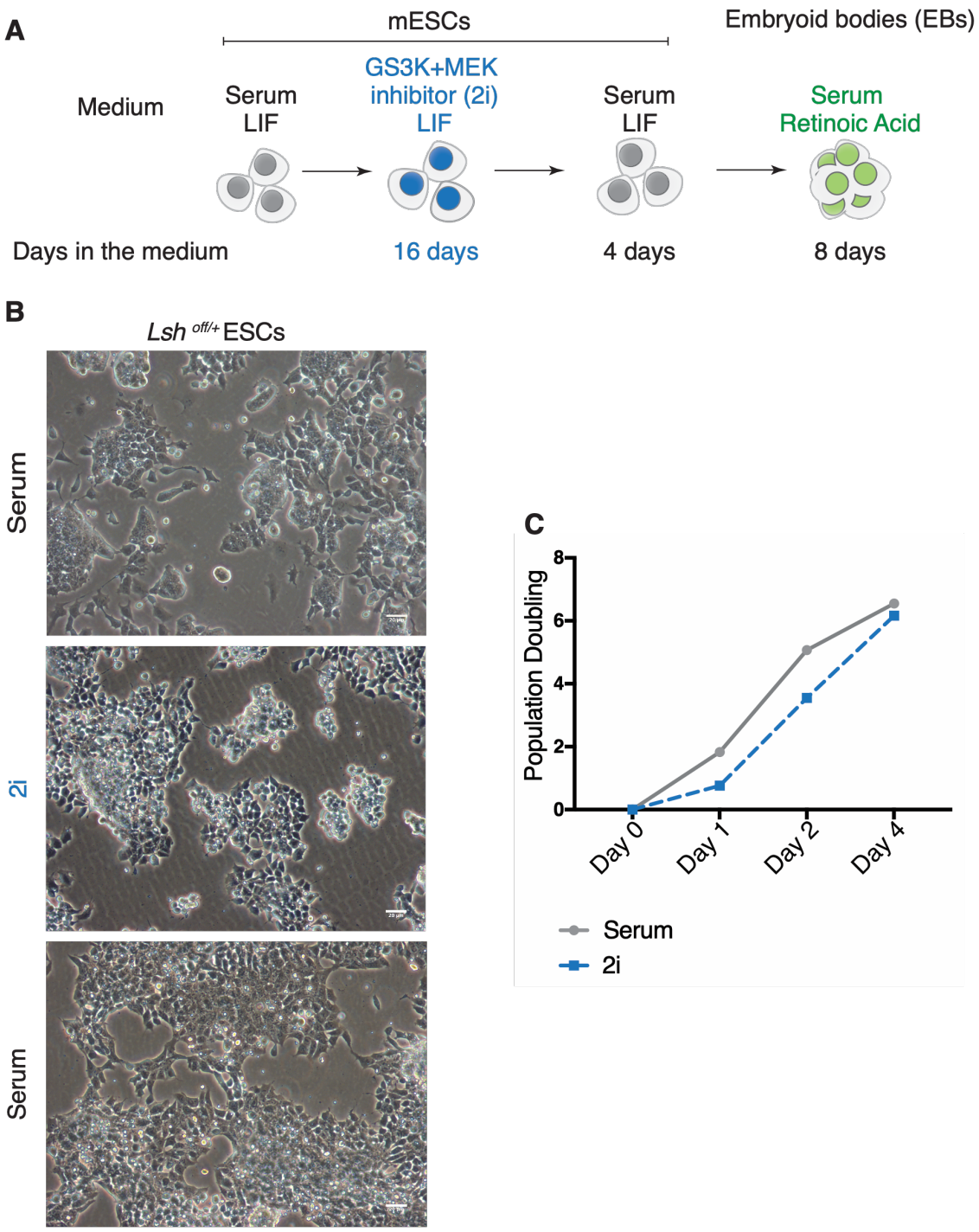


Figure 3.2 Establishment of a cell culture system mimicking DNA methylation dynamics of early embryonic development

A. Schematic representation of the experimental design.

B. Representative images of *Lsh* $^{off/+}$ ESCs in serum-, 2i- and again serum-containing medium.

C. Population doublings of *Lsh* $^{off/+}$ ESCs in serum compared to *Lsh* $^{off/+}$ ESCs in 2i. Despite an initial difference, after 4 days in culture the cells divide with similar rate.

3.3. Total DNA methylation of ESCs is lost in 2i and re-established in serum culturing condition

Since the cells were showing to be responsive to the different culturing conditions, global DNA methylation levels were analysed initially by using an antibody against 5-methyl cytosine (5meC) on denatured DNA dot blots. Quantification of the amount of 5meC was performed over the amount of ssDNA present on the blot.

It was previously reported that the DNA methylation of ESCs maintained in 2i-medium decreases rapidly and reaches a steady state at around 30% of the initial value (Ficz et al., 2011). As shown in Figure 3.3, DNA methylation in *Lsh*$^{off/+}$ ESCs dropped by more than half of the initial methylation level at 4-day time point in culture supplemented with the 2 inhibitors. After 8 days in 2i-medium, global DNA methylation finally reached 30% of the initial 5meC level, as expected, and stayed mostly stable. Interestingly, the dot-blot analysis showed that the restoration of DNA methylation did not require similar time to that required for demethylation. In fact, the methylation was restored in 4 days of culture in serum-medium after the time in 2i (Figure 3.3 A and B).

These results showed that global levels of 5meC decrease in *Lsh*$^{off/+}$ ESCs when cultured in 2i-medium and that this hypomethylation is quickly recovered. Furthermore, a time window was established, to optimally study the dynamics of these changes. More specifically, at least 12 days of culture in 2i-medium are necessary to deplete the global DNA methylation by 70% and, on the other hand, 4 days in serum-containing medium are enough to restore the global methylation. However, the dot-blot approach has big technical limitations, such as the quantification limit of the antibody-based detection and reproducibility of the technique. In addition, RNA contaminations are not detectable and in fact affect the quantification of the 5meC. Therefore, the quantification of global DNA methylation was subsequently performed by High Pressure Liquid Chromatography (HPLC).

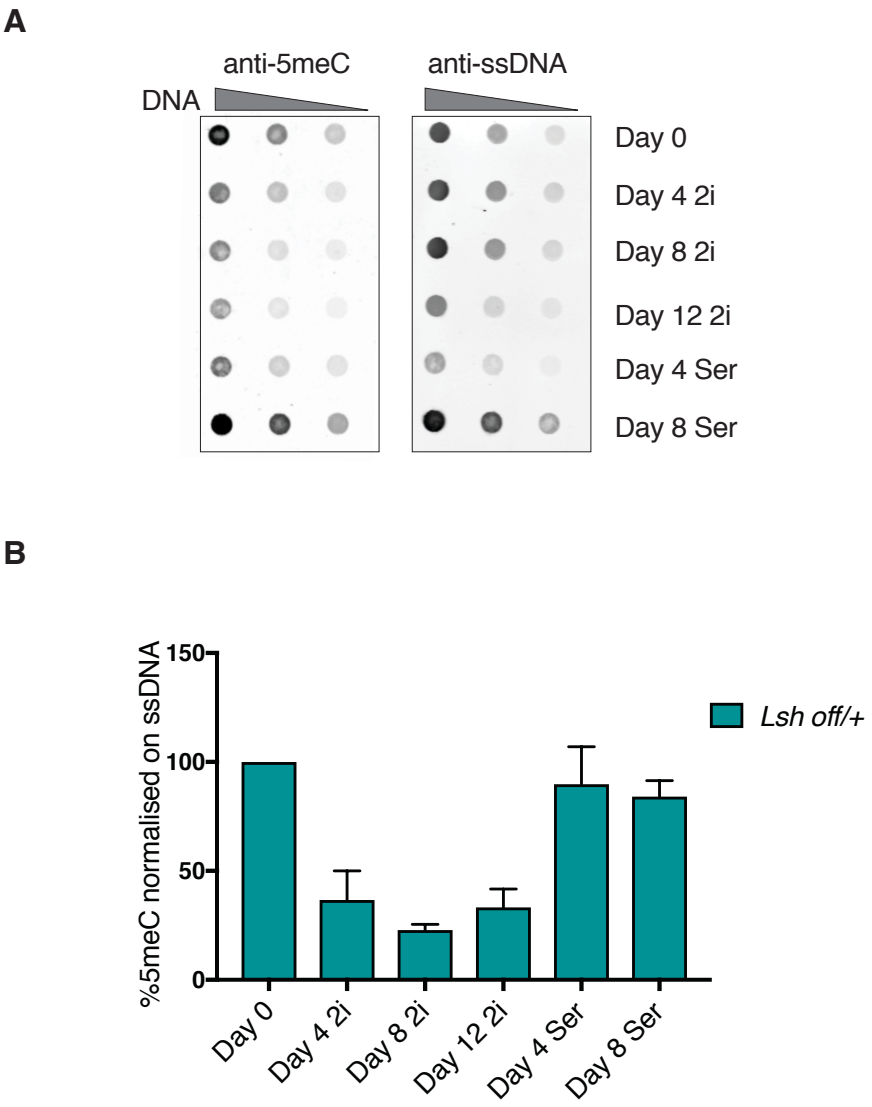


Figure 3.3 Total DNA methylation of ESCs is lost in 2i and re-established in serum-containing culture conditions

A. In the left panel, global DNA methlation was detected by an anti 5meC antibody on a DNA dot blot. In the right panel, DNA loading was quantificatied by an anti ssDNA antibody on an identical dot blot. Serial dilutons of DNA were used, with the highest concentration beeing 1μg.

B. Quantification of the 5meC, relative to the loaded ssDNA. The values are shown relative to Day 0 and expressed as a percentage. The methylation drops in 2i-medium and is restored quickly in serum-containing culture conditions.

Error bars indicate standard deviation between biological replicas.

Altogether, these experiments indicated that this system could be efficiently used with *Lsh*$^{off/+}$ mESCs to study DNA methylation changes. Testing the expression of DNMTs was then necessary to understand if the regulation of the system could be analysed within this time-frame in culture.

3.4. Protein and RNA levels of DNMTs and PRDM14 are differentially regulated in 2i- and serum-containing cultures

The hypomethylation typical of ESCs cultured in 2i-medium is accompanied and supported by regulation of the DNA methyltransferase enzymes. More specifically, Western blot analysis has previously shown that when the cells were cultured in this medium increased expression of PRDM14 was observed, which then led to downregulation of the *de novo* methyltransferases DNMT3A and DNMT3B (Ficz et al., 2011). However, DNMT1 was not affected and appeared stably expressed. This regulation can be detected at the mRNA level as well, indicating that the mechanism regulating the depletion of 5meC occurs on a gene expression more than on protein level.

Protein abundance analysis by Western blotting of extracts obtained from *Lsh*$^{off/+}$ ESCs maintained first in serum-containing medium and then in 2i-medium showed that DNMT3B and DNMT3A are down regulated when the cells are cultured with the inhibitors for 12 days. The methyl transferases were then upregulated again when the cells were cultured in serum-containing medium, showing that the downregulation is only transient and can be quickly reversed. In contrast to what was previously reported (Leitch et al., 2013), Western blot analysis showed that both DNMT3A isoforms were downregulated in 2i-medium (Figure 3.4 A).

Importantly for our study, the expression of LSH protein in 2i-medium maintained cells had never been analysed before. Therefore, we decided to investigate the LSH protein levels by Western blotting and found that LSH was not downregulated in this

system (Figure 3.4A). This indicates that the signalling pathway that leads to the hypomethylation does not involve LSH regulation on either RNA or protein level and will therefore be possible to discern the effect of lack of LSH from the effects induced by the 2i-medium. Interestingly, it appeared that LSH was upregulated once the cells were moved back to the serum-containing medium. It was previously hypothesized that LSH is expressed especially in highly proliferating cells (Lee et al., 2000; Raabe et al., 2001). This might suggest that the transiently increased expression of LSH in serum, after culturing in 2i-medium, could be linked to the proliferative state of the mESCs.

Analysis of the mRNA levels in the same system confirmed what was detected at protein level. In fact, *Dnmt3a* and *Dnmt3b* were downregulated in naïve pluripotent cells, and upregulated in primed cells. On the other hand, *Prdm14* was upregulated in 2i medium-maintained cells and then downregulated in serum, as expected. Finally, to verify the hypothesis that upregulation of *Prdm14* might lead to an effect on the amount of G9a protein, via its target activity on GLP (Leitch et al., 2013), qRT-PCR was performed. G9a expression does not change between culturing conditions, suggesting that its expression is not affected by the 2 inhibitors (Figure 3.4B).

These experiments show that the regulation of methyltransferases occurs as expected in 2i-medium and either 12 or 16 days in 2i are sufficient to establish these changes. Furthermore, the mRNA and protein expression profiles changed rapidly in serum and were discernible after 8 days of culture in serum-containing medium on a protein level. Therefore, it was decided to culture the cells for a shorter time with serum after 2i, in order to mimic more closely the short time that occurs *in vivo* between the beginning of the *de novo* methylation wave and the differentiation (He et al., 2011; Howlett & Reik, 1991; Li, 2002; Reik et al., 2001).

Importantly, the 2i culture conditions apparently had no effect on either the expression or stability of LSH. Therefore, the system can be used reliably to study the consequences of the absence of LSH, concomitantly with changes in DNA methylation.

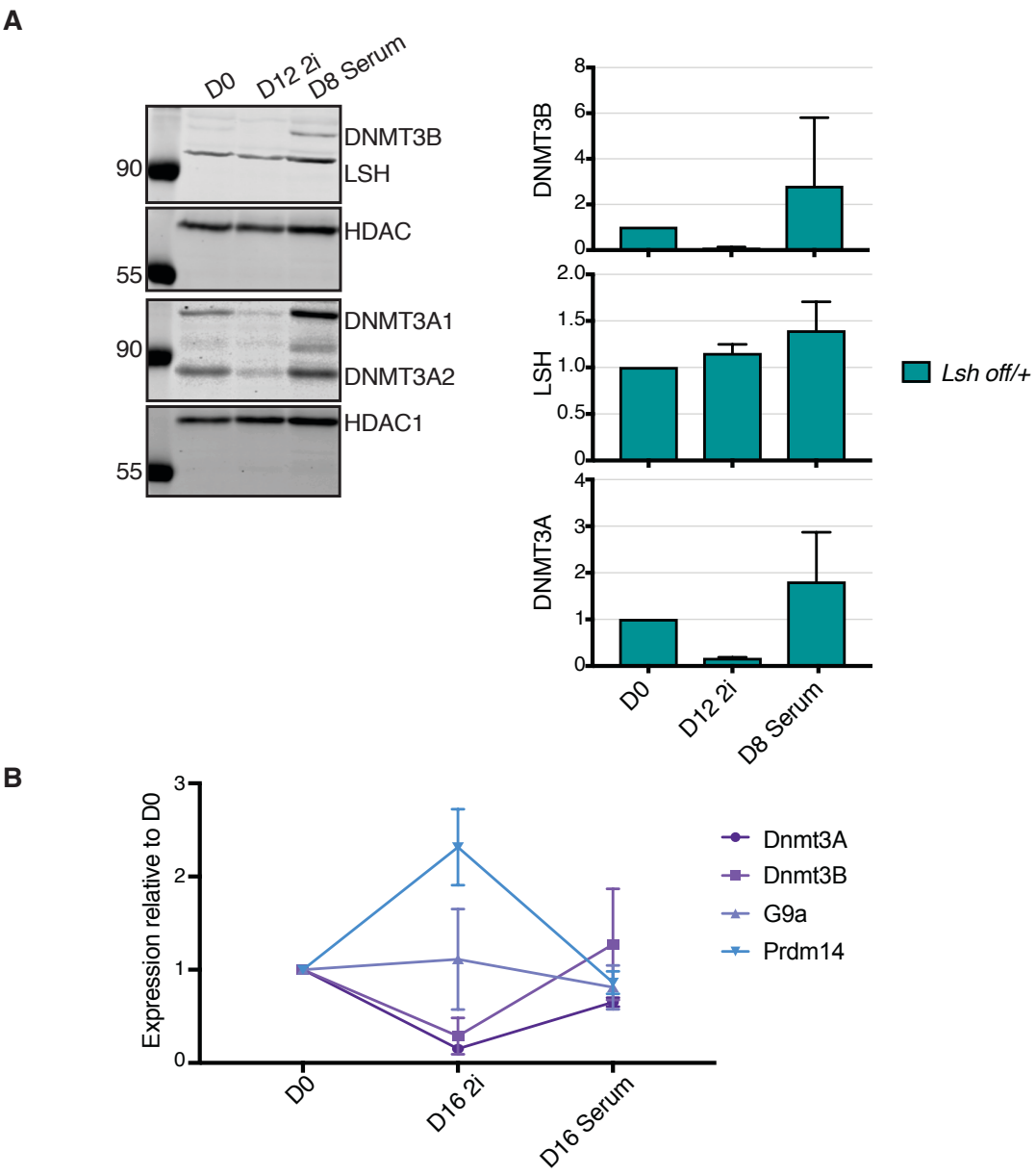


Figure 3.4 Protein and RNA levels of DNMTs and PRDM14 are differentially regulated in 2i-medium.
A. On the left, Western blots of nuclear extracts from cells grown with serum, 2i and returned to serum-containing medium are shown. The blots were probed with antibodies against LSH, DNMT3B, DNMT3A and HDAC. While DNMT3B and DNMT3A are downregulated in 2i-medium, LSH is expressed at all points of the experiment. HDAC was used as a loading control. On the right, quantification of protein expression shown as average of 3 technical replicas, normalised to HDAC and relative to D0.
B. RNA expression analysis by quantitative PCR of DNMT3B, DNMT3A, G9a and PRDM14 in cells grown in serum, 2i and serum again. GAPDH was used as an internal control. Quantification of the RNA expression is shown as average of 3 technical replicas, normalised to GAPDH and relative to D0. Error bars indicate standard deviation between biological replicas.

3.5. Conclusions

Understanding the role of LSH in early mammalian development is crucial to shed some light on its importance in the healthy adult, as well as individuals affected by disease, such as ICF. We decided to overcome the difficulties in the study of initial phases of mammalian development by taking advantage of different culture conditions established for ESCs, namely serum-containing or 2i-containing cultures. However, before exploring the effects of the absence of LSH in these cells, it was important to assess that the system performed as expected. To do so, heterozygous *Lsh*$^{off/+}$ ESCs were used to set up the system.

These preliminary experiments showed that culturing *Lsh*$^{off/+}$ ESCs in 2i-medium does not lead to either growth arrest or cell death. Also, the cells showed very similar features to what was previously reported in literature. In particular, DNA methylation level decreased after 8 days in 2i-containing medium, showing that the inhibition of WNT and FGF signal cascades is efficient in these cells. The methylation loss was accompanied by downregulation of the *de novo* methyltransferases, and upregulation of Prdm14, as expected. Furthermore, switching the cells back to serum-containing medium efficiently restored the global DNA methylation levels. The restoration of the DNA methylation was associated with upregulation of the *de novo* methyl-transferases and downregulation of transcriptional regulator, Prdm14.

Interestingly, LSH protein expression level did not change when the cells were cultured with the addition of the inhibitors for 4 days. This indicates that, even though LSH is hypothesized to be involved in the *de novo* DNA methylation, its protein and mRNA levels are not regulated by the same signalling pathways as those of the *de novo* DNA methyltransferases and therefore not affected by CHIR99021 and PD184352 in the culture.

Finally, adjustments of the time in culture proved to be necessary in order to achieve the optimal time-frame allowing demethylation and remethylation to take place, while

analysing in parallel the effects of LSH absence.

From these preliminary data, the system overall proved efficient and suitable to study DNA methylation dynamics in the absence of LSH.

4. DNA methylation dynamics in the absence of LSH and heterochromatin

4.1. Introduction

Epigenetic regulation is essential for the correct function of the cell. This regulation can occur via different mechanisms, namely either via RNA-based regulations or histone modifications and DNA methylation. It has been previously shown that the protein LSH is central for regulation of DNA methylation in mammals and plants. Its plant homologue, Decrease in DNA methylation-1 (DDM1), was firstly found to be essential for DNA methylation in *Arabidopsis thaliana.* In fact, absence of the protein leads to loss of 70% of the DNA methylation in *ddm1* mutant plants (Kakutani et al., 1995; Vongs et al., 1993). Later studies in mice have shown a role of LSH in methylation in mammals as well (Geiman et al., 2001). In fact, LSH-null mice die shortly after birth and display kidney failure and reduced birth weight, which are accompanied by hypomethylation of the genome. It is therefore clear that one of LSH main functions is regulating DNA methylation and this is conserved between species. Since these publications, a lot of work has been done to elucidate the role of LSH in DNA methylation.

Recently, the Stancheva lab produced mice from $Lsh^{off/+}$ ESCs (data not published). These mice, contrary to the previous models, can develop and live after birth, despite showing reduced weight and a shacking phenotype from early age. Importantly, this mouse model, like its predecessors, is also characterised by reduced DNA methylation. More specifically, experiments at day 12.5 of the embryonic development (E12.5) in the $Lsh^{off/off}$ and $Lsh^{off/+}$ mice, showed that absence of LSH results in 40% decrease of DNA methylation in the $Lsh^{off/off}$ embryos and mice, while the $Lsh^{off/+}$ are not affected.

Although a role of LSH in DNA methylation is clear, it is only hypothesized, but not directly proven, that it plays a role early in development, most probably before E12.5. During these first phases of the mammalian development, both the *de novo* DNA methylation and maintenance of the newly deposited methyl groups are important. It still remains unclear whether LSH is involved in one or both these processes, as shown by the numerous contradicting studies trying to address this question.

After establishing a system that could allow to recapitulate *in vitro* the reprogramming of the epigenome *in vivo*, DNA methylation and hydroxyl-methylation were analysed in absence of LSH. These experiments would help shedding some light on the role of LSH in the early phases of mammalian development.

4.2. The absence of LSH affects the loss and gain of DNA methylation

In order to investigate the DNA methylation dynamics in $Lsh^{off/off}$ ESCs compared to $Lsh^{off/+}$ ESCs, both cell lines were cultured in the culturing system described in the previous chapter. After a short period in serum-containing medium, the cells were grown in 2i-supplemented medium. Finally, 16 days of culturing in this medium were chosen as the most representative time point for the experiment. Subsequently, an optimization step was carried out to determine for how long the cells needed to be grown in serum-containing medium. Aiming to have a representation of the short period of time that elapses between the initial *de novo* methylation and the starting of the differentiation, 4 days of culturing in serum-containing medium after 2i were chosen. After 4 days, the cells were taken through the initial stages of differentiation by withdrawal of LIF in non-adherent culture conditions for 4 days. Addition of retinoic acid then induced the cells to committing to neuronal cells lineage and a final time point at 8 days of culture was established to detect the changes in DNA methylation occurring following the initial phases of differentiation (Mohn et al., 2008). Analysis of the population doubling of $Lsh^{off/off}$ ESCs compared to $Lsh^{off/+}$ ESCs

showed no significant difference between cell lines, when grown in the same conditions (Figure 4.1A). To ensure that *Lsh*$^{off/off}$ ESCs were not expressing LSH and that this depletion was constant in the different culturing systems, western blot analysis was performed. As expected, no protein expression was detected in cells maintained in serum nor in 2i-containing medium (Figure 4.1B).

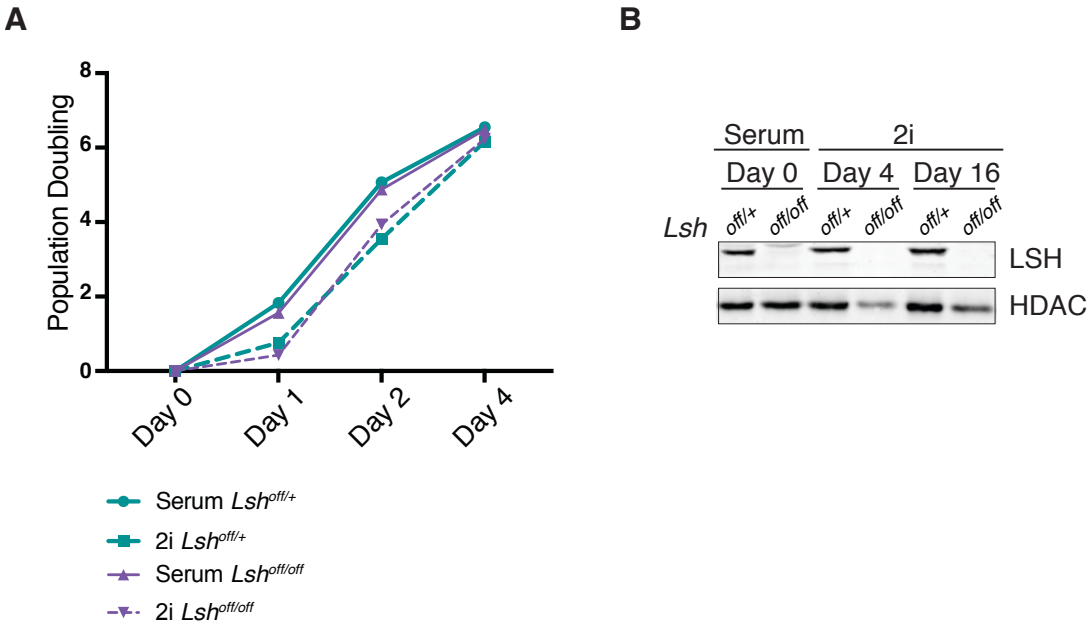


Figure 4.1.*Lsh*$^{off/off}$ ESCs grow similarly to *Lsh*$^{off/off}$ ESCs and do not express LSH protein.
A.Population doublings of *Lsh* $^{off/+}$ ESCs in serum and 2i compared to *Lsh* $^{off/off}$ ESCs in serum and 2i-containing medium. The two cell lines behave almost identically in both co-colturing conditions.
B.Western blots of nuclear extracts from ESCs grown with serum or 2i-containing medium are shown. The blots were probed with antibodies against LSH and HDAC. While LSH is expressed in *Lsh* $^{off/+}$ ESCs at all stages of the time course experiment, as expected *Lsh*$^{off/off}$ ESCs do not show any expression of the LSH protein. HDAC was used as a loading control.

Genomic DNA was extracted at each time point and analysed by High-performance Liquid Chromatography (HPLC), as previously described (Ramsahoye, 2002). This is a very accurate technique to quantify the total amount of 5meC present in the sample, as well as the total h5meC.

At the start of the experiment (day 0), the *Lsh*$^{off/off}$ ESCs displayed lower DNA methylation level compared to the *Lsh*$^{off/+}$ ESCs (Figure 4.2A). Surprisingly, however, the difference in methylation detected between the cell lines was smaller than what was previously shown for mouse embryonic fibroblasts (MEFs) and E12.5 embryos. However, this could be explained with the characteristics of the samples. In fact, these cells were genetically modified *in vitro* for the introduction of two copies of the cassette, kept in culture and frozen. This means that any difference in the methylation could have only occurred *in vitro*, whereas the previous data was obtained from cells that had gone through the formation of a mice embryo.

Once the cells were moved to the 2i-containing medium and the demethylation process started, both cell lines lost DNA methylation, but not equally. More specifically, the *Lsh*$^{off/+}$ ESCs lost half of their methylation after only 4 days in 2i and then reached a steady state of approximately 1.2% of total 5meC, which is approximately 30% of the initial value. This is consistent with what we previously detected by dot-blot analysis (Figure 3.3). On the other hand, the *Lsh*$^{off/off}$ ESCs in 4 days lost half of their initial DNA methylation to then continue losing 5meC in the following days, dropping it to 0.5%, which is approximately 13% of the initial DNA methylation level. These results indicate that the global loss of DNA methylation is exacerbated in the absence of LSH. However, at this stage the mechanism by which the loss is increased is not clear. In fact, this significant difference might occur as a result of either a more open chromatin state in cells lacking LSH, which therefore might render the genome more accessible to demethylases, or be caused by increased activity of the demethylases, such as the TET enzymes, or finally stem from impaired maintenance of DNA methylation at replication foci, which would point to a role of LSH in maintenance of DNA methylation.

When the cells were switched to serum-containing medium, the DNA methylation was expected to be restored, as a consequence of the expression of the *de novo* methyl-transferases, DNMT3A and DNMT3B. Indeed, DNA methylation levels were restored in both cells types, but after 4 days in serum-containing medium, they were still significantly lower in *Lsh*$^{off/off}$ ESCs (Figure 4.2A). This might indicate a slower rate of recovery or an

impairment of the *de novo* methylation in cells lacking LSH.

Early phases of differentiation are characterised by silencing of pluripotency genes and activation of lineage specific ones. Also, cells grown as embryoid bodies stop dividing at day 4 due to addition of retinoic acid and dramatically decrease expression of the *de novo* methyl-transferases. In our culture system, at day 4 of the EBs differentiation the 5meC, interestingly, reaches higher levels if comparing to the DNA methylation detected at the beginning of the experiment. Then, after addition of the RA, the EBs start expressing lineage specific genes and this seemed to be accompanied by a decrease in the levels of DNA methylation after 8 days of culture (Figure 4.2A). The changes at this stage were similar in both the $Lsh^{off/off}$ and $Lsh^{off/+}$ EBs. However, the decrease in 5meC at this stage was significantly more substantial in the $Lsh^{off/off}$ EBs. Despite the observation that *de novo* DNA methylation occurred efficiently even in the absence of LSH, the difference detected in EBs at day 8 suggests an impairment of this process. Furthermore, the inability of the $Lsh^{off/off}$ ESCs to completely compensate for the difference in global DNA methylation established after the demethylation phase in 2i might indicate that a proportion of the genome cannot be methylated efficiently in the absence of LSH. Altogether, these data suggest that activity of LSH is required to correctly complete the *de novo* methylation.

The representation of the data as a histogram gave a good understanding of the 5meC level at each time point and of any difference existing between the $Lsh^{off/off}$ and $Lsh^{off/+}$ ESCs. However, it was difficult to compare the overall dynamics of the changes and to highlight any difference in removal and accumulation of 5mC on the DNA over time. This is especially true due to the difference in the *de novo* methylation, as the recovery starts from very different levels in the $Lsh^{off/off}$ and $Lsh^{off/+}$ ESCs. Therefore, we decided to display these data by dividing them into two plots, one representing the demethylation and one the *de novo* methylation rate normalizing to the starting point of each process, for each cell line. As a result, in each plot, the value of the first time point (day 0 or day 16 2i, respectively) was plotted as 1 for $Lsh^{off/off}$ and $Lsh^{off/+}$ ESCs and the following values presented relative to 1 (Figure 4.2B and C).

The rate of demethylation was similar for the two genotypes (Figure 4.2B). Normalisation to 1 clearly showed that the 5meC decreases by half after 4 days and by 25% more after 8 days in both of *Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ ESCs. As obvious from the histograms, while *Lsh*$^{off/+}$ ESCs reach a steady state in the DNA methylation levels, *Lsh*$^{off/off}$ ESCs continue losing 5meC. The reasons behind this further loss is however not clear, as discussed above.

Surprisingly, when the rate of DNA methylation gain upon recovery from 2i was displayed relative to day 16 in 2i, the accumulation of 5meC was very different between the two cell lines (Figure 4.2C). In fact, the *Lsh*$^{off/+}$ ESCs displayed a 4-fold increase of 5meC at 4 days as EBs, after which 5meC decreases, as expected due to the initiation of the differentiation process. Despite the similar trend in *Lsh*$^{off/off}$ ESCs, there was a 8-fold increase in the methylation at day 4 of the differentiation process. Since the absolute amount of methylation on the DNA was smaller, this suggests that in the same time period, more methyl groups accumulate on DNA in LSH-deficient cells. However, this is still not sufficient to completely compensate for methylation differences between the *Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ ESCs. This faster accumulation of 5mC can be explained by a more open chromatin state, which allows easier access to the *de novo* methyl transferase enzymes. It could potentially also be a consequence of an increased synthesis of methyl transferases, which would then operate faster on multiple loci in the genome.

Altogether these results show that the absence of LSH influences DNA methylation, as previously shown for later developmental stages. However, the system used here enable us to hypothesise that LSH is important earlier in development than previously reported. Furthermore, it seems clear that the lack of LSH affects both the dynamics of demethylation and the *de novo* methylation. However, it is still not evident how.

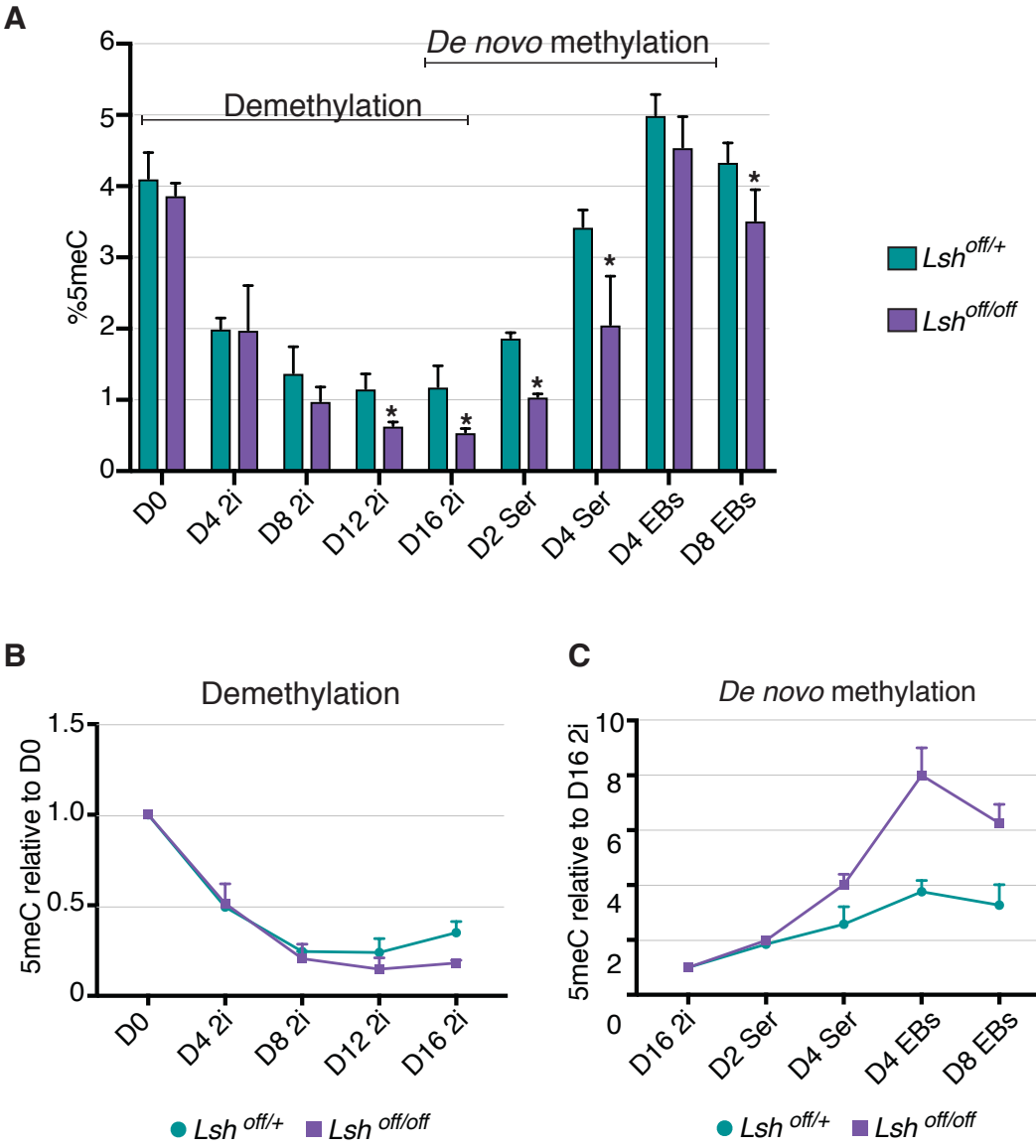


Figure 4.2.The absence of LSH affects the loss and gain of DNA methylation.
A. HPLC analysis of 5meC, as a percentage of the total C in the genome. Days and culturing conditions are annotated on the x axis (D0= cells in serum, D4/8/12/16 2i=4/8/12/16 days in culture with 2i medium, D2/4 Ser= 2/4 days in culture in serum-containing medium, D4/8 EBs=4/8 days into the formations of embryoid bodies). n=6 biological replicas; * indicates p-value < 0.05
B. Dynamic of DNA demethylation in *Lsh* $^{off/+}$ and *Lsh* $^{off/off}$ ESCs. The fold change in 5meC is shown relative to D0.
C. Dynamic of *de novo* DNA methylation in *Lsh* $^{off/+}$ and *Lsh* $^{off/off}$ ESCs. The fold change in 5meC is shown relative to D16 in 2i.
Error bars represent the standart deviation between biological replicas; p-values determined by Student's t-test. The percentage of methylated cytosine was calculated as percentage of the total cytosine present in the genome analysed.

4.3. The loss of 5meC in *Lsh*$^{off/off}$ ES cells correlates with accumulation of 5hmeC

The previous experiments showed that the absence of LSH leads to a bigger loss of 5meC in naïve pluripotent ESCs. A mathematical model was used to explain that the main mechanism involved in loss of DNA methylation in this culturing system is the impairment of maintenance (von Meyenn et al., 2016). However, in this system we observed a different response in cells lacking LSH to what expected. The deeper loss described could be explained by either an increase in active DNA demethylation or an impairment of methylation maintenance when the cells are cultured in 2i-containing medium.

The 5hmeC is formed by oxidation of the 5meC by the TET oxygenase enzymes (Figure 1.4). This is the first step of the demethylation process, which requires 2 more oxidative steps to be completed. The TET enzymes further convert the 5hmeC to 5-formylcytosin (5fC), and finally to 5-carboxylcytosine (5caC). The final step of removal of the modification from the bases is thought to involve the Thymine DNA Glycosylase (TDG) enzyme, which is part of the Base Excision Repair (BER) machinery (Kohli & Zhang, 2013). Therefore, the active removal of 5meC is typically accompanied by accumulation of 5hmeC. Importantly, it has been shown that PRDM14 upregulation leads to increased expression of the Tet enzymes in cells cultured in 2i-containing medium. Consequently, in these cells the 5hmeC levels are higher (Okashita et al., 2014).

On the other hand, when the maintenance of the DNA methylation is impaired, the DNA is not methylated on the newly synthesized strand at the replication fork. Thus, the 5meC is passively lost by dilution of the modification concomitantly with the cell division. When this happens, an increase of the amount of total 5hmeC is not detectable.

Therefore, in order to elucidate which mechanism is the main responsible for the increased demethylation in absence of LSH, the hydroxyl-methyl-cytosine (5hmeC) was measured by HPLC. Genomic DNA was extracted and analysed for both 5meC (Figure 4.3) and 5hmeC by HPLC. When 5hmeC was analysed as percentage of the total cytosine

(Figure 4.3A) at day 0, the amount of this modification in the two cell lines was very similar. Then, a 2.5 fold increase in 5hmeC was detected in the cells cultured in 2i-containing medium, particularly in the first passages in this medium, corresponding to the main demethylation events (Figure 4.3A). Although there is slightly more 5hmeC in the *Lsh*$^{off/off}$ ESCs at day 4 in 2i-containing medium compared to the *Lsh*$^{off/+}$ ESCs, this difference is not significant and later in the time course experiment it becomes negligible. Notably, there is a big variation between measurements of the 5hmeC. This is an issue when performing statistical analysis, as no significant difference can be established. However, this variation is due to technical limitations, as the amount of hydroxylated cytosine in the genome is very small and this affects the accuracy of quantification.

Expressing the amount of 5hmeC as percentage of 5meC (Figure 4.3 B) enabled us to determine how much of the measured methylated cytosine was actually being removed from the DNA. In this case as well, there was no significant difference between the two cell lines in the amount of 5hmeC at day 0. This was followed by a 3 to 4 fold increase in hydroxymethylation in 2i-medium. In fact, in both cell lines the 5hmeC levels in 2i stayed higher than at day 0 and until the cells were switched back to serum-containing medium. Moreover, 5hmeC decreased of about 40% in *Lsh*$^{off/+}$ ESCs after 4 days in 2i-containing medium, indicating that at this stage all the accessible methyl groups had been removed from the cytosine. Interestingly, this was not the case for *Lsh*$^{off/off}$ ESCs. More specifically, in the absence of LSH a higher portion of 5meC was oxidised when the cells were cultured in 2i-containing medium and these levels were maintained constantly around 10% of the 5meC until the cells were switched back to serum-containing medium. This suggests that some methyl groups were still accessible to the TET activities in the genome of *Lsh*$^{off/off}$ ESCs and could be removed, consistently with the more pronounced decrease of 5meC shown in the previous experiment (Figure 4.2).

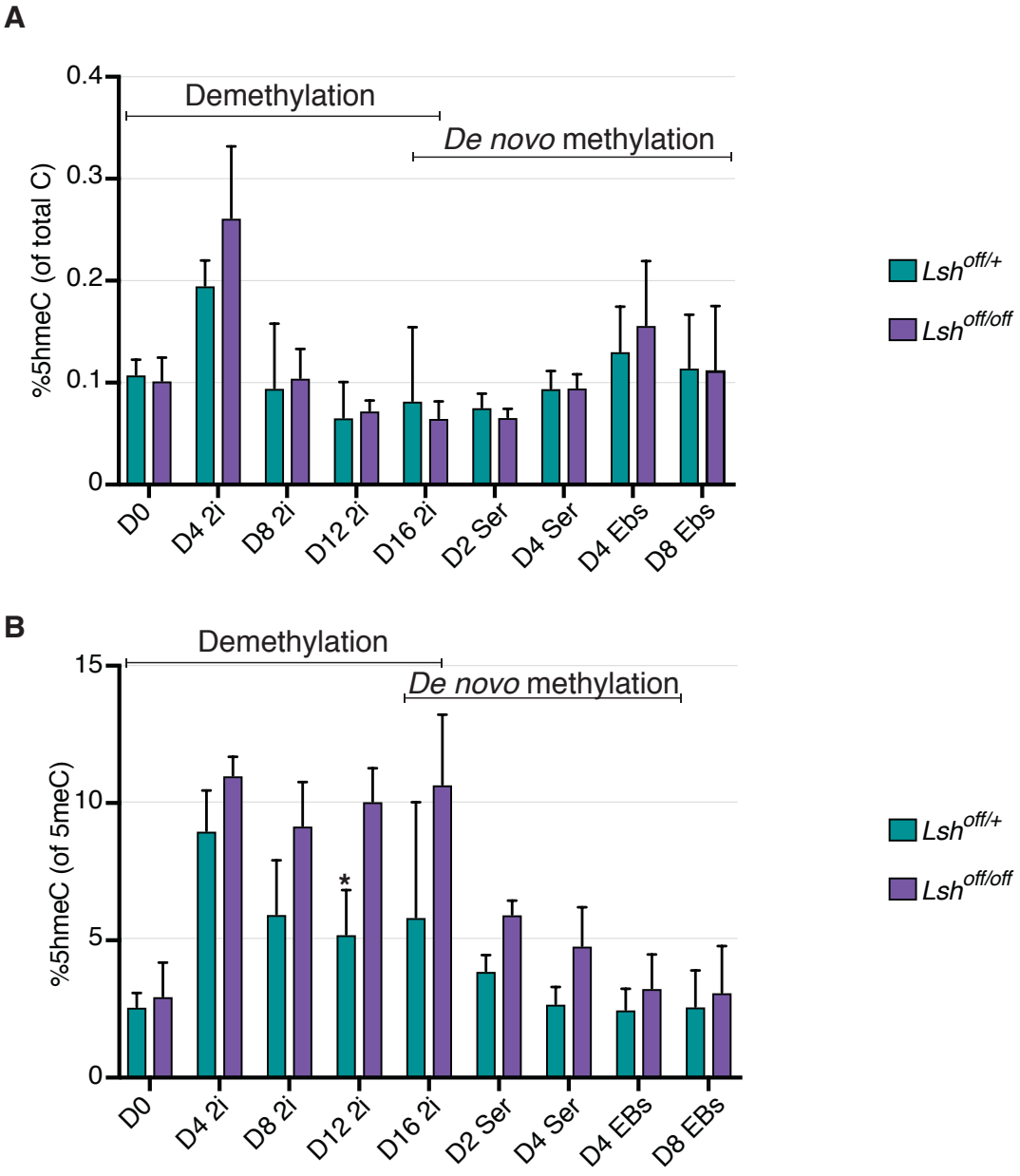


Figure 4.3. The loss of 5meC in *Lsh* $^{off/off}$ ES cells correlates with accumulation of 5hmeC.
A. HPLC analysis of 5hmeC, as a percentage of the total C in the genome. Days and culturing conditions are annotated on the x axis (D0= cells in serum, D4/8/12/16 2i=4/8/12/16 days in culture with 2i medium, D2/4 Ser= 2/4 days in culture in serum-medium, D4/8 EBs=4/8 days into the formations of Embryoid bodies). n=5 biological replicas; * indicates p-value < 0.05. The percentage of methylated cytosine was calculated as percentage of the total cytosine present in the genome analysed.
B. HPLC analysis of 5hmeC, as a percentage of 5meC in the genome. Days and culturing conditions are annotated on the x axis.n=5 biological replicas; * indicates p-value < 0.05
Error bars represent the standart deviation between biological replicas; p-values determined by Student's t-test.

These findings suggest that most of the loci are rapidly demethylated both in *Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ ESCs, when the cells are shifted to 2i-containing medium. However, in the absence of LSH the active demethylation can continue and occurs at more loci in the genome, causing a further accumulation of 5hmeC, until the cells are moved back to serum-containing medium. Therefore, some loci might be accessible to the TET enzymes only in cells lacking LSH, however, it is yet unclear which loci and why. It was previously shown *in vitro* that TET enzymes have a preference for open chromatin rather than for compacted chromatin. Consequently, it could be hypothesised that, when LSH is depleted, a more open chromatin environment is established and this might lead to a wider access to the genome of the TET enzymes, which can then oxidise more cytosine.

Altogether, these data seem to support the hypothesis that the lower level of 5meC occurring in naïve pluripotent ESCs lacking LSH is determined by an active demethylation, ruling out a possible role of LSH in the replicative maintenance of DNA methylation.

4.4. Removal of facultative heterochromatin has an additive effect on the loss and gain of 5meC and 5hmeC in *Lsh*$^{off/off}$ ES cells

Heterochromatic regions of the DNA are characterised by the presence of DNA methylation as well as specific histone marks. These are chemical modifications of the histone tails and can have a role in activation or repression of gene expression. The most important modifications for gene silencing are di- and tri-methylation of histone H3 at lysine 9 (H3K9me2/3) and tri-methylation of H3 at lysine 27 (H3K27me3). H3K9me2 is deposited mostly by the G9a/GLP methyl transferases complex, which is highly abundant in ESCs. H3K9me2 is also abundant in both pluripotent stem cells and differentiated cells (Lienert et al., 2011) covering about 50% of the genome in ESCs maintained in serum-containing medium and 53% of the genome in neuronal cells derived from ESCs. Interestingly, the domains of H3K9me2 are stable and, despite what was previously

thought, the difference between ESCs and differentiated cells is minimal and includes mostly newly established H3K9me2 domains at gene bodies in differentiated cells. However, it was subsequently shown that when ESCs are cultured in 2i-containing medium, H3K9me2 is depleted from most of the genome. Interestingly, this loss does not affect H3K9me3, which remains stable when the cells are driven to a naïve state of pluripotency (Walter et al., 2016). This suggests a specific role of the H3K9me2 in early stages of differentiation and the establishment of so called facultative heterochromatin. Furthermore, this modification is typically associated with lamina-associated domains (LAD) (Towbin et al., 2012; Wen et al., 2009) and has been shown to cooperate with H3K27me3 to relocate chromatin to the nuclear periphery (Harr et al., 2015).

LSH was shown to support DNA methylation at LADs (Yu et al., 2014), some of these chromatic regions are in fact demethylated in *Lsh*$^{-/-}$ MEFs. Furthermore, absence of LSH was shown to compromise recruitment of the G9a/GLP complex at a subset of loci, determining hypomethylation of the locus (Myant et al., 2011). However, no direct interaction between LSH and the histone methyl transferase complex was detected. Therefore, we hypothesise that LSH supports DNA methylation by remodelling heterochromatin at H3K9me2-marked LADs and thereby facilitating DNA methylation at these loci, concomitantly with *de novo* methylation. Thus, depletion of H3K9me2, and therefore heterochromatin, by inactivation of G9a, could promote accessibility of the chromatin for deposition of the methylation on the DNA, bypassing LSH activity.

To test this hypothesis, *Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ ESCs were treated with an inhibitor of the G9a/GLP methyltransferase complex, UNC0638. This small molecule inhibits the catalytic activity of the complex by acting as a strong competitor of H3K9me2 and occupying the substrate binding groove of G9a (Vedadi et al., 2011). The cells were treated with UNC0638 for 4 and 8 days and then tested by Western blotting for the depletion of H3K9me2. As expected, 4 days of treatment were sufficient to deplete most of H3K9me2 from chromatin of the inhibitor-treated cells (Figure 4.4B), with the cells still being able to divide. For the time course experiment to analyse the 5meC, the cells were treated for 6 days with UNC0638 prior to the switch to 2i-contaning medium (Figure 4.4A)

to allow sufficient time for dilution of H3K9me2 before starting the time course experiment. Cells of the same genotypes without treatment were used as a control. DNA methylation and hydroxymethylation were then analysed by HPLC.

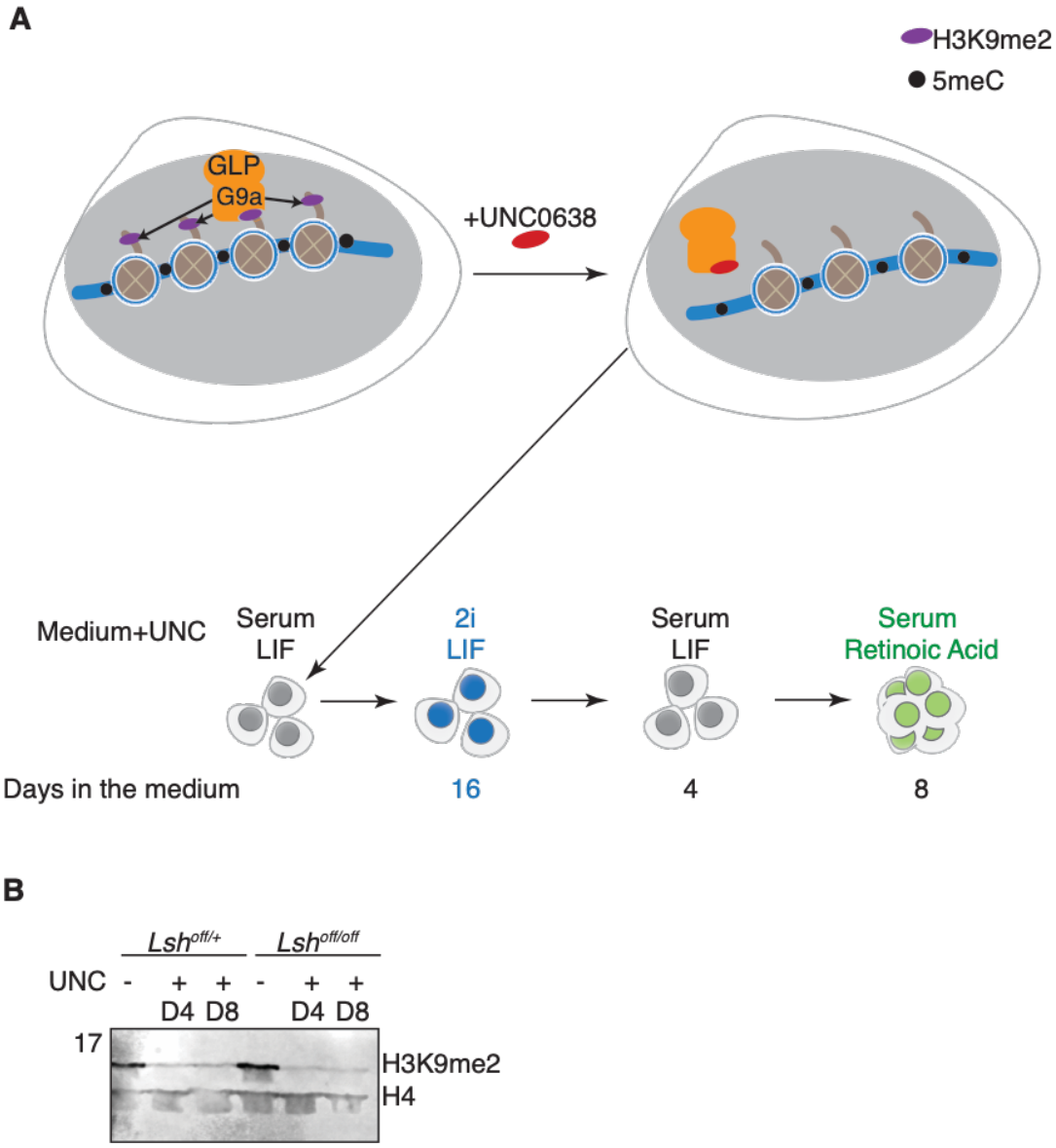


Figure 4.4. Removal of facultative heterochromatin has an additive effect on loss and gain of 5meC and 5hmeC in Lsh$^{off/off}$ ES cells.
A.Schematic representation of the experimental design. G9a/GLP is a methyl transferases complex which binds and methylate H3K9. When UNC0638 is added in culture, G9a binds the small molecule and the H3K9me2 is lost by dilution concomitantly with the replication.
B.Western Blot analysis of H3K9me2 and H4 in mESCs before and after treatment with the G9a/GLP inhibitor (UNC - and +). Days in culture in serum-medium are annotated on top.

Treating the cells with UNC0638 led to a decrease in the total cytosine methylation in serum, probably indicating impairment of the G9a-dependent DNA methylation maintenance (Shinkai & Tachibana, 2011; Tsumura et al., 2006) both in *Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ ESCs (Figure 4.5 A). Subsequently, the change of culturing conditions led to a quick decrease of DNA methylation from 3 to 1% in only 4 days in *Lsh*$^{off/+}$ ESCs. In *Lsh*$^{off/off}$ ESCs, this decrease was even faster, reaching 0.5% at 4 days in 2i-medium. In both cell lines, the loss of methylation was more substantial than in the untreated with the inhibitor control cell lines. While cells lacking LSH kept loosing DNA methylation further, reaching 0.3% in 16 days, *Lsh*$^{off/+}$ ESCs reached a steady state and maintained around 1% of the cytosine methylated, consistently with what occurred in the absence of UNC0638 treatment.

After activation of the *de novo* methylation machinery in serum, both cell lines were able to methylate the DNA in 4 days. However, only *Lsh*$^{off/+}$ ESCs showed a complete restoration of DNA methylation to initial values thus overcoming completely the maintenance defect caused by inhibition of G9a/GLP. On the other hand, *Lsh*$^{off/off}$ ESCs restored only 66% of the initial DNA methylation, recapitulating again the what occurred in the untreated control (Figure 4.5A).

Finally, when cells proceeded through the initial phases of differentiation and formed EBs, in both cell lines DNA methylation initially increased. Interestingly, at day 8 of differentiation some of the DNA methylation was lost, especially in *Lsh*$^{off/off}$ EBs. In fact, the 5meC decreased from the 3.8% to 2.8% in UNC0638 treated cells. The same trend was detected in the untreated control. At this stage, the difference in methylation detected between *Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ EBs is more substantial than the difference detected at the ESCs stage (Figure 4.5A).

In order to have a better understanding of the dynamics of these changes, the data were plotted as two separate graphs, demethylation and *de novo* methylation, as explained above (chapter 4.2). The 5meC was lost more quickly in the presence of G9a/GLP inhibitor in both *Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ ESCs (Figure 4.5 B). More specifically, after

4 days of culturing in 2i-medium, while the untreated control lost 50% of the initial DNA methylation, *Lsh*$^{off/+}$ ESCs treated with the inhibitor lost 68% of the initial 5meC and *Lsh*$^{off/off}$ ESCs treated with the inhibitor dropped to 0.12, corresponding to a loss of 88%. Hence, this effect could be attributed to the depletion of H3K9me2 histone modification. Interestingly, while *Lsh*$^{off/+}$ ESC grown with UNC0638 had similar levels of DNA methylation as those of the untreated control by day 16 in 2i, *Lsh*$^{off/off}$ ESCs treated with UNC0638 lost even more 5meC than their untreated counterparts, reaching 92.3% decrease compared to 88% of the control cell line (Figure 4.5 B). This suggests that the depletion of facultative heterochromatin and the absence of LSH have additive, yet independent, effect.

When exploring the *de novo* DNA methylation dynamics, surprisingly, the accumulation of methyl groups had similar trend in all the cell lines, but very different rates (Figure 4.5C). As seen above from these data analyses (Figure 4.2C), *Lsh*$^{off/off}$ ESCs in 4 days in serum medium accumulate more methyl groups than the *Lsh*$^{off/+}$ ESCs, while maintaining lower absolute levels of DNA methylation in the genome. The treatment of the cells with the UNC0638 accentuates this difference, by increasing the deposition (Figure 4.5C). A 4-fold difference in deposition rate was detected between in *Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ ESC grown with UNC0638, while the control cell lines differed by 2-fold. Importantly, this difference was dependent on an increased accumulation rate in *Lsh*$^{off/off}$ ESCs depleted of H3K9me2 compared to the untreated counterpart, as *Lsh*$^{off/+}$ ESC only increase by 1 fold in the rate of change compared to the cells not treated with the inhibitor (Figure 4.5C).

Altogether, these data suggest that the effect of the depletion of H3K9me2, and therefore increase in accessibility of the chromatin, has a similar effect on DNA methylation deposition as the absence of LSH. This supports the hypothesis that the absence of LSH leads to a more open state of chromatin which facilitates the action of DNA demethylase enzymes. Furthermore, this experiment highlights an interplay between LSH and G9a in the deposition of methylation in early differentiation. More specifically, in these experiments the inhibition of G9a resulted in an amplification of DNA methylation effects detected in the previous experiments. Therefore, if LSH is responsible for

facilitating DNA methylation of heterochromatic regions of the genome, and the permanent silencing of these loci, G9a might be essential for methylating the histone tails of the same loci. Alternatively, G9a might be required to a different subset of genomic loci to those that require LSH and the effect that we detected could be linked to the depletion of 5meC at the two separate subsets. However, *Lsh*$^{off/+}$ ESC treated with the UNC0638 showed very similar behaviour in terms of DNA methylation dynamics to the untreated cells. Therefore, the increase in the DNA methylation deposition rate is more plausible to be attributed to a combined effect, rather than to independent changes occurring are G9a-dependent regions.

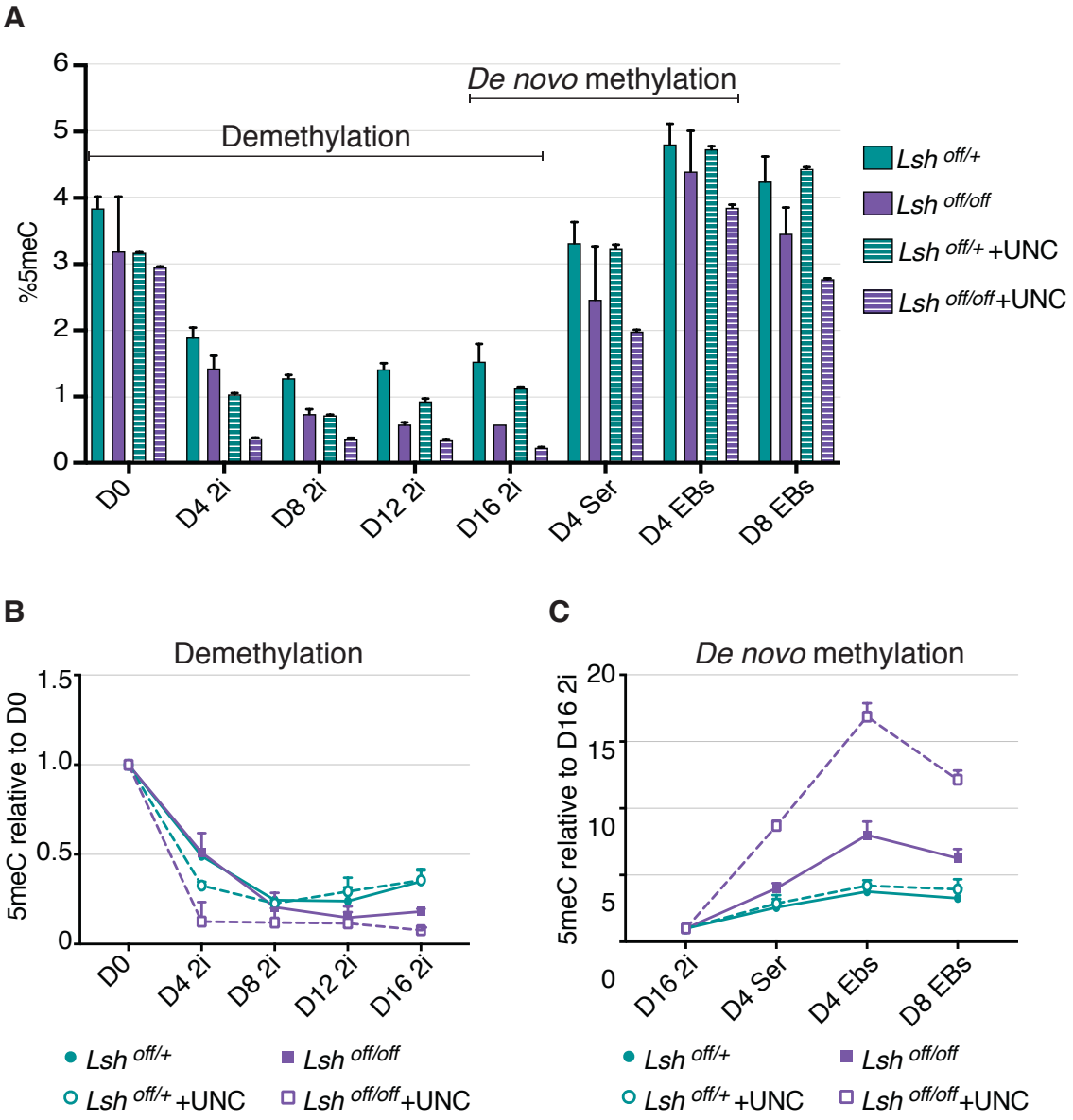


Figure 4.5. Removal of facultative heterochromatin has an additive effect on loss and gain of 5meC and 5hmeC in Lsh$^{off/off}$ ES cells.

A.HPLC analysis of 5meC, as a percentage of the total C in the genome. Days and culturing conditions are annotated on the *x* axis (D0= cells in serum, D4/8/12/16 2i=4/8/12/16 days in culture with 2i medium, D2/4 Ser= 2/4 days in culture in serum-medium, D4/8 EBs=4/8 days into the formations of Embryoid bodies).Treatment with the G9a/GLP inhibitor is indicated as +UNC. n=3 techinical replicas;

B.Dynamic of DNA demethylation in *Lsh* $^{off/+}$ and *Lsh* $^{off/off}$ ESCs treated with G9a/GLP inhibitor (+UNC). The fold change in 5meC is shown relative to D0.

C.Dynamic of the *de novo* methylation wave in *Lsh* $^{off/+}$ and *Lsh* $^{off/off}$ ESCs treated with G9a/GLP inhibitor (+UNC). The fold change in 5meC is shown relative to D16 in 2i.

Error bars represent the standart deviation between technical replicas. The percentage of methylated cytosine was calculated as percentage of the total cytosine present in the genome analysed.

4.5. The absence of LSH promotes chromatin accessibility

The experiments conducted so far demonstrated that cells lacking LSH are characterised by misregulation of the 5meC and the 5hmeC in the genome. To explain this, the facultative heterochromatin was depleted by inhibition of the G9a/GLP methyltransferases complex. The data collected suggests that the activity of LSH might be linked to the accessibility of the chromatin and therefore that the protein might be crucial for DNA methylation in regions marked by histone modifications, such as H3K9me2. Therefore, the absence of LSH might lead to a more open state of the chromatin. However, no evidence was yet produced to support this hypothesis.

In order to examine the state of the chromatin, Micrococcal nuclease (MNase) was used to digest the chromatin of *Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ ESCs. This endonuclease cuts exposed DNA between nucleosome and is used to infer information on nucleosome positioning and the overall chromatin accessibility. After digestion, the DNA associated to the nucleosomes is recovered and analysed. Digestion of nuclei with MNAse determines formation of mono-nucleosomes. However, depending on the accessibility of the chromatin, the formation of mono-nucleosomes requires long, if the chromatin is not accessible, or short time, if the chromatin is accessible (Figure 4.6).

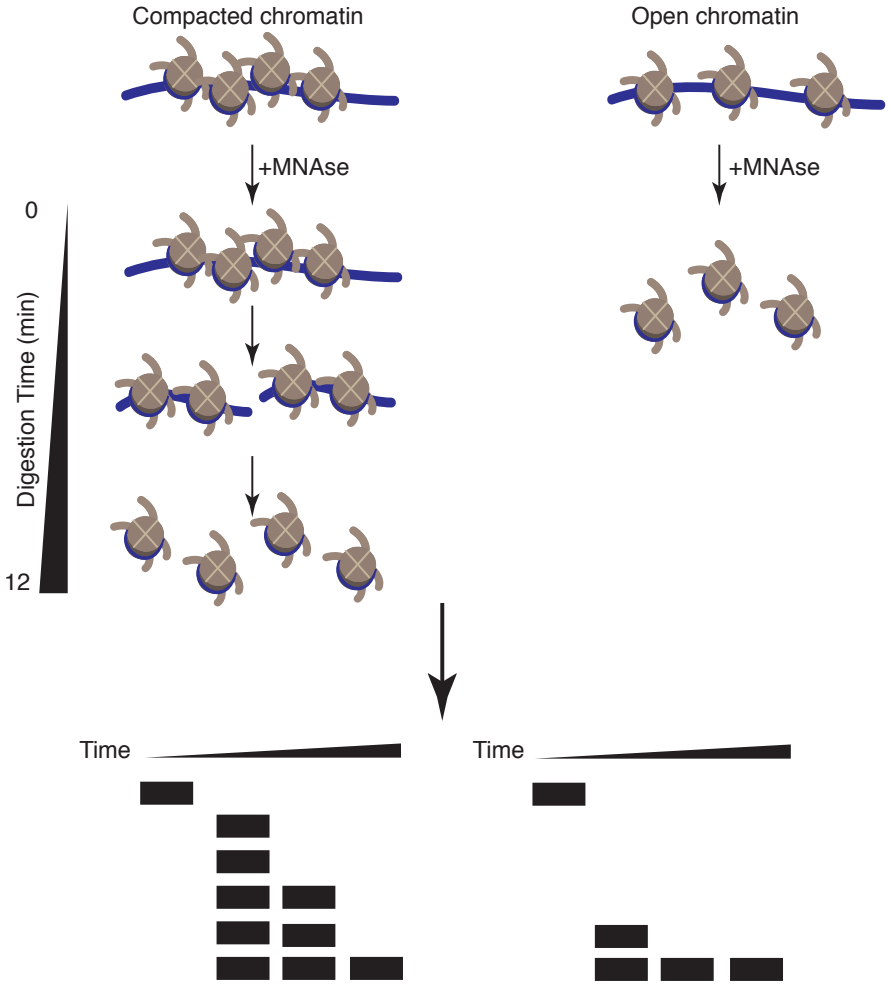


Figure 4.6.Chromatin accessibility assay
Schematic representation of the experimental design. Chromatin extracted from *Lsh* $^{off/+}$ and *Lsh* $^{off/off}$ ESCs was digested with the Micrococcal nuclease for 0 minutes (control) to 12 minutes and then DNA was extracted and analysed by gel electrophoresis. When the chromatin is compacted (left) and thereby difficult to access for the MNAse, longer time is necessary to produce di and mono-nucleosomes. On the other hand, accessible chromatin is easly digested by the MNAse, hence after very short incubation time mononucleosomes are produced.

To verify it absence of LSH is associated with change in chromatin accessibility, nucleosome positioning in *Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ ESCs at day 0, day 2 serum after 16 days in 2i medium and in *Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ EBs was analysed. After extraction of nuclei, these were treated with MNAse for 0 to 12 minutes. Subsequently to nucleosome-associated DNA extraction and purification, electrophoresis on agarose gel was used to determine the size of the digestion products. Although amount of chromatin was estimated by measuring the optical density ratio at 260/280 nm, the results were semi-quantitative and gave an indication of the state of chromatin in different samples. At day 0, when the difference in DNA methylation between the two genotypes is minimum, the digestion pattern of chromatin was similar, indicating that mono-nucleosomes are formed in similar times in *Lsh*$^{off/+}$ and *Lsh*$^{off/off}$ ESCs (Figure 4.7A). Similarly, at day 2 in serum-containing medium, following 16 days in 2i-containing medium, mono-nucleosomes in the two cell lines are formed in equal time (Figure 4.7A). However, the signal intensity profile analysis of samples digested for 3 minutes highlighted a difference (Figure 4.7B). In fact, the *Lsh*$^{off/+}$ ESCs sample was characterised by accumulation of high molecular weight DNA, represented by a very sharp peak at the top of the lane. This indicates that not all chromatin was digested at that time point of the experiment. However, the *Lsh*$^{off/off}$ ESCs sample showed a broader peak of high molecular weight and presence of low molecular weight at the tail of the curve, corresponding to di and mono-nucleosomes. This suggests that the digestion of the *Lsh*$^{off/off}$ ESCs chromatin had started already after 3 minutes of incubation with the enzyme (Figure 4.7B). Finally, a more visually obvious difference was observed at the EBs stage. In fact, the chromatin of the *Lsh*$^{off/off}$ EBs was largely digested after 3 minutes of incubation with the MNAse, while this was not the case for the *Lsh*$^{off/+}$ EBs (Figure 4.7).

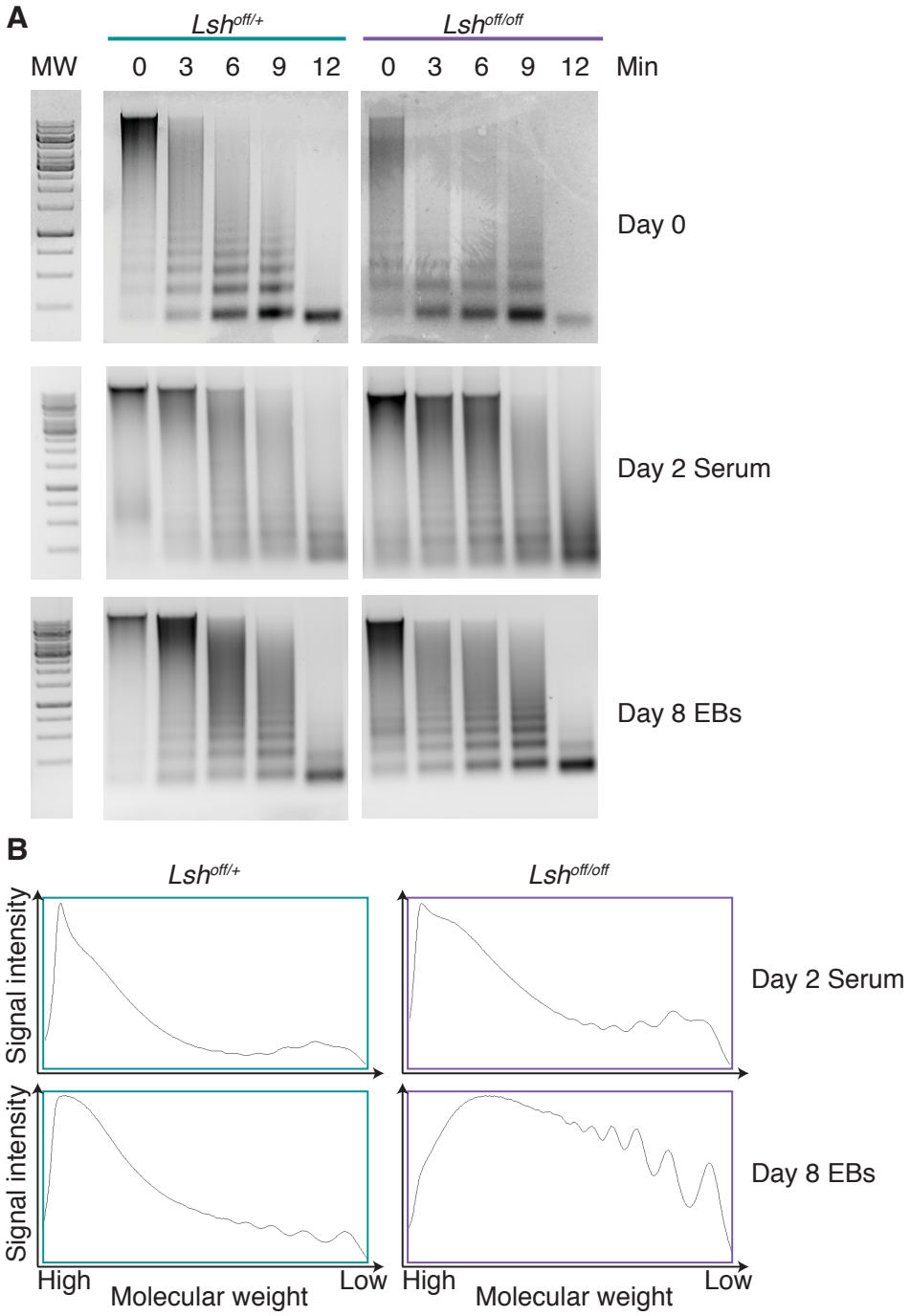


Figure 4.7.Absence of LSH determines increased chromatin accessibility
A.MNAse digestion assay in *Lsh* $^{off/+}$ and *Lsh* $^{off/off}$ ESCs at various time points (annotated on the right) in the time course experiment. Chromatin was analysed after 3, 6, 9 and 12 minutes of digestion with the enzyme. The 0 minutes lane correspond to the treated control, without incubation.
B.Plot of the lane profile of *Lsh* $^{off/+}$ and *Lsh* $^{off/off}$ ESCs Day 2 Serum and Day 8 EBs, after 3 minutes of incubation with the enzyme. Analysis was performed using ImageJ software.

These results support the hypothesis that the absence of LSH leads to a more accessible chromatin structure. Importantly, the biggest differences are evident after the demethylation/*de novo* methylation phases. Therefore, the absence of LSH alone is not enough to determine sever effects in ESC genome architecture. When cells go through methylation reprogramming and differentiate, therefore establishing a less dynamic and open chromatin compared to ESCs, consequences of absence of LSH are appreciable.

4.6. Conclusions

It is difficult to study the DNA methylation dynamics *in vivo*, at early stage of development, due to technical limitations and the difficulties of working with animals. However, using 2i-containing medium to culture ESCs recreates the naïve state of pluripotency that resembles day 3.5 of the embryonic development. At this stage, the *de novo* methyltransferases are not active either in the embryo or in the 2i-medium cultured cells. The subsequent switch in culturing medium enables reactivation of the *de novo* methyltransferases and recapitulates DNA methylation events taking place in the post implantation blastocyst. This system, although not perfectly representative of the real embryonic development, mimics these early phases and can be used to provide insights into the role of LSH at these stages.

The study of the DNA methylation dynamics in the absence of LSH (Figure 4.2A) showed that, when cells had never gone through DNA methylation reprogramming, there is no significant difference in the 5meC levels between $Lsh^{off/off}$ and $Lsh^{off/+}$ ESCs. This suggests that cells can maintain the 5meC without LSH, if they are not subject to any demethylation or *de novo* DNA methylation. At later time points of the experiment, ESCs lacking LSH lose more 5meC in 2i-containing medium than the heterozygous cells and accumulate 5hmeC (Figure 4.3 A). This suggests that although LSH is not involved in

maintenance of DNA methylation, it might contribute to the protection of some loci from the activity of the TET enzymes. Moreover, the *de novo* methylation can proceed in the absence of LSH (Figure 4.2 A). However, it is incomplete or slower compared to the heterozygous control cells. In fact, after 4 days in serum-containing medium the 5meC is not completely restored, even though the *de novo* methyltransferase act faster, possibly due to a more open state of the chromatin (Figure 4.2 B and C). Finally, $Lsh^{off/off}$ EBs at day 8 of differentiation show a lower level of DNA methylation compared to their heterozygous counterparts. This difference is more pronounced compared to the difference detected at day 0 of the time course between the two cell lines. Therefore, this indicated that LSH has a role concomitant with the *de novo* methylation and it might be important only for methylation of a subset of loci in the genome.

Overall, the dynamics of DNA methylation (Figure 4.2 B and C), are 2-fold faster in absence of LSH than in the control heterozygous for LSH cells. Also, when analysing the cells grown with the G9a/GLP inhibitor, the DNA methylation dynamic is more rapid in $Lsh^{off/off}$ ESCs compared to $Lsh^{off/+}$ ESCs (Figure 4.5 B and C). Therefore, when LSH is not present, the chromatin is more open and accessible (Figure 4.5 A and B), and this promotes the access of DNA for demethylases and methyltransferases. Thus, it can be hypothesized that this enhanced accessibility promotes faster DNA methylation dynamics. Furthermore, the additive effect of the inhibition of the G9a/GLP complex on top of LSH deficiency highlights a possible interplay between LSH and G9a in DNA and histone methylation early in embryonic development. More specifically, this might indicate the existence of a two step process in the deposition of the methylation: an early one, that is not dependent on LSH and G9a and a later one, leading to differentiation of the cells, that requires LSH and also G9a.

So far, this thesis work describes whole genome analyses which were performed to elucidate the role of LSH activity early in development. As mentioned above, only a subset of the genome seems to require LSH to be correctly methylated. Despite many studies have been published showing LSH activity on specific loci, using this culturing system can provide a further understanding of the importance of LSH linked to different stages of the

early development. In particular, experiments described in the next chapter will investigate which are the hypomethylated regions in ESCs that did not go through DNA methylation reprogramming (*Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ ESCs at day 0), ESCs that exit the naïve pluripotency and had started the *de novo* methylation wave (*Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ ESCs at day 4 in serum, after maintenance in 2i-containing medium) and in cells at early phases of differentiation (*Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ EBs).

5. DNA methylation at specific loci is LSH dependent

5.1. Introduction

Experiments performed so far have shown the importance of LSH for *de novo* DNA methylation, highlighting the role of the protein concomitantly with the reprogramming of the 5meC within a system mimicking early phases of the mammalian development. Furthermore, I showed that DNA methylation only at a subset of the genome is influenced by the presence of LSH. However, the use of HPLC to analyse the amount of DNA methylation only provided limited quantitative information and did not allow the identification of the genomic regions at which DNA methylation is misregulated in *Lsh*$^{off/off}$ cells concomitantly with DNA methylation reprogramming.

High-throughput sequencing analyses are a formidable tool to study in detail changes occurring at the genome level, by at the same time zooming in and looking at a whole genome level. In order to determine which loci require LSH to establish the correct DNA methylation patterns after depletion of the modification, we decided to use a high throughput sequencing technique called BioCAP-sequencing. By sequencing the unmethylated fraction of the genome, this approach will provide insights into which regions of the genome are depleted of 5meC and how these change when the genome goes through DNA methylation reprogramming. Altogether, this information will help elucidate the role of LSH in mammalian embryonic development.

5.2. BioCAP-sequencing is an efficient method to analyse the methylation state of the genome.

There are currently many next-generations sequencing technologies that enable analysis of the genome methylation state, such as sequencing of bisulfite converted DNA (whole genome bisulfite sequencing and reduced representation bisulfite sequencing) or sequencing of immunoprecipitated methylated DNA, such as the MeDIP-seq. A relatively less common technique is the Biotinylated CxxC affinity Purification-sequencing (BioCAP-seq). This technique is based on the CxxC Affinity Purification (CAP) approach that uses the high affinity of the ZINC finger CxxC protein domain for unmethylated cytosines to profile non-methylated DNA (Illingworth et al., 2008). This technique was further optimised in Rob Klose's lab (Blackledge et al., 2012), leading to the use of smaller amounts of input DNA and not requiring high-resolution chromatography systems. The BioCAP-sequencing uses a biotinylated CxxC peptide, derived from the lysine demethylase KDM2B, bound to neutravidin beads (Figure 5.1). These CxxC coated beads are then incubated with sonicated genomic DNA, which is then eluted in gradient salt wash. As a result, the eluted DNA is enriched in unmethylated cytosines and can be then used for massively parallel sequencing. The big advantage of using this technique derives from the fact that unmethylated cytosines are only 1-2% of total cytosines in a mammalian genome, primarily at CpG islands, especially in differentiated cells or primed pluripotent stem cells, such as the mESCs cultured in serum-containing medium. This specific enrichment allows deep sequencing with a smaller number of reads required for coverage compared to whole genome sequencing techniques. Therefore, this method, albeit indirectly, provides information about the methylation status of the DNA in a cost-efficient fashion.

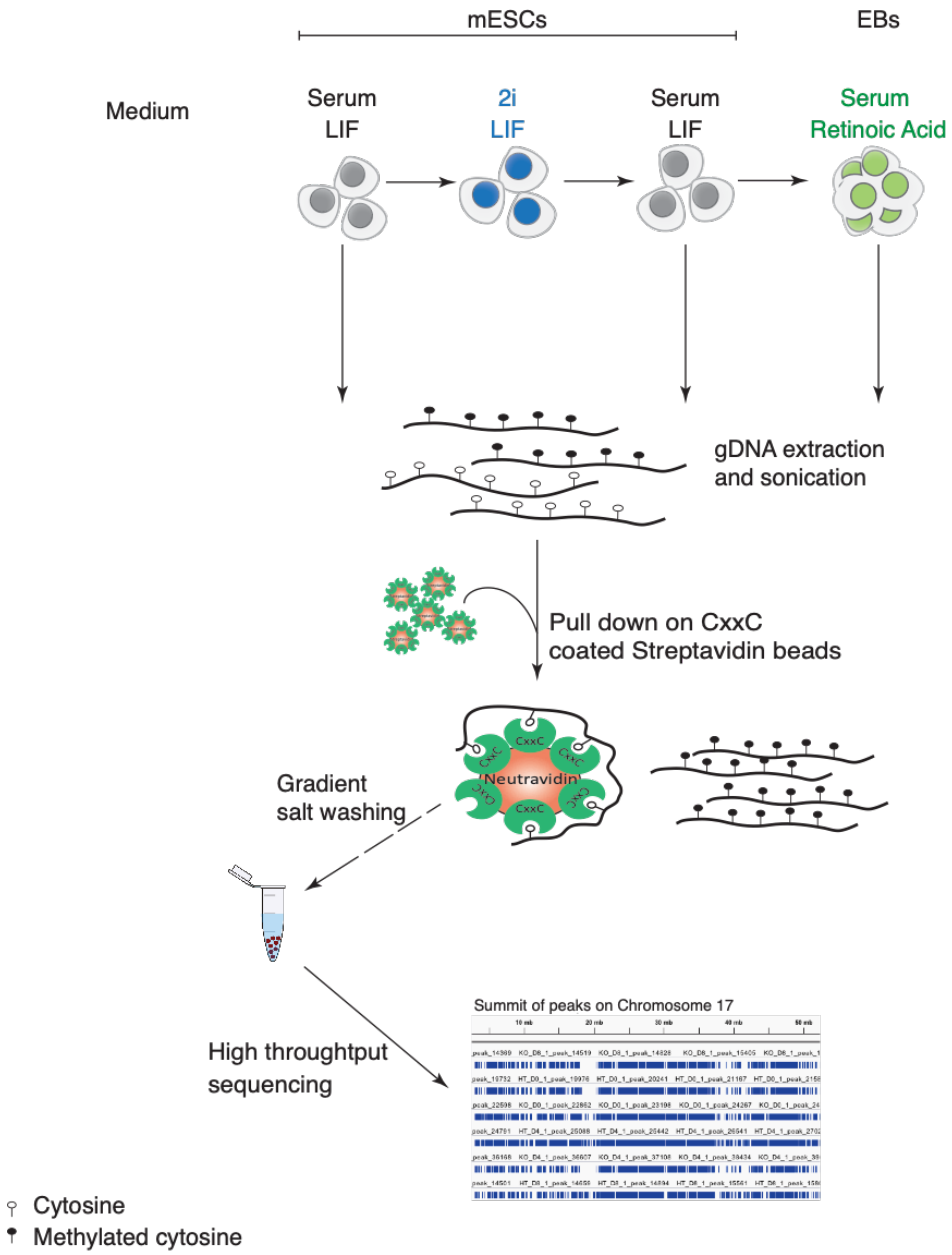


Figure 5.1 BioCAP-sequencing work flow.
Schematic overview of the work flow of the BioCAP-sequencing experiment. Cells were collected at 3 points of the time course experiment, day 0, day 4 in serum and day 8 EBs. After extraction and sonication of the gDNA, unmethylated cytosines (white lollipops) were pulled down by hybridisation with the CxxC domain-coated streptavidin beads. Unbound methylated (black lollipops) fractions of the gDNA were washed away and the highly unmethylated gDNA was then eluted and sequenced.

The crucial step in the BioCAP technique is the pull down of the unmethylated fraction of the genome, using CxxC coated neutravidin beads. To ensure sequencing of only the highly unmethylated fraction of the genome, after a first step of incubation of the coated beads with sonicated whole gDNA, these were washed in buffers with increasing salt concentration, from 100 to 1000 mM NaCl (Figure 5.1). To verify that the CxxC coated beads were highly specific for binding to unmethylated DNA, PCR analyses were performed after each washing step (Figure 5.2). Three genes were chosen for this analysis, 2 of which, Actin B and Moesin (Msn), were predicted to be unmethylated at the promoter and therefore expected to elute from the beads only in high salt buffers. The third locus chosen was the Fatty Acid Binding Protein 7, Fabp7, predicted to be methylated in mESCs and hence expected to elute from the beads in low salt buffer washes. The presence of Actin B and Msn DNA in fractions with 700mM and 1000mM NaCl confirmed that these regions bind to the CxxC coated bead with higher affinity than it could be expected from unmethylated regions and hence could be eluted from the beads only in presence of high concentration of salts (Figure 5.2, top panel). On the contrary, the presence of Fabp7 DNA at low NaCl concentration (Figure 5.2, bottom panel) showed that this sequence of DNA could be easily eluted from the beads, indicating low affinity binding presumably due to the high cytosine methylation level.

These preliminary results tested the reliability of the coated beads produced in the Stancheva lab by Dani Wickacsono and proved that they are suitable for further experiments. Therefore, we decided to proceed with the high throughput sequencing experiment, to investigate which regions of the genome require LSH for deposition of DNA methylation.

In order to address this question, the methylation status of *Lsh*$^{off/+}$ and *Lsh*$^{off/off}$ mESCs at day 0, cultured in serum-containing medium, were analysed by BioCAP-sequencing as control time point. *Lsh*$^{off/+}$ and *Lsh*$^{off/off}$ mESCs cultured in serum for 4 days after 16 days in 2i-containing medium were collected and DNA was analysed to investigate the changes occurring right after the demethylation, caused by treatment with 2i. Finally, DNA of *Lsh*$^{off/+}$ and *Lsh*$^{off/off}$ EBs at 8 days of growth was sequenced to consider how

absence of LSH could impact changes in 5meC of cells beginning the differentiation process. Sequencing of these samples will provide insights into which loci are hypomethylated in the absence of LSH. Furthermore, the comparison of samples at different time points of the time course experiment, that mimics the early stages of mouse embryonic development, will help understanding the timing of LSH activity. In fact, by analysing the difference in DNA methylation between day 0 and the subsequent time points, it should be possible to determine which regions become methylation-depleted as a consequence of the methylation reprogramming events, namely the demethylation wave, the activation of *de novo* DNA methyltransferases and finally a second *de novo* DNA methylation step to achieve cellular differentiation.

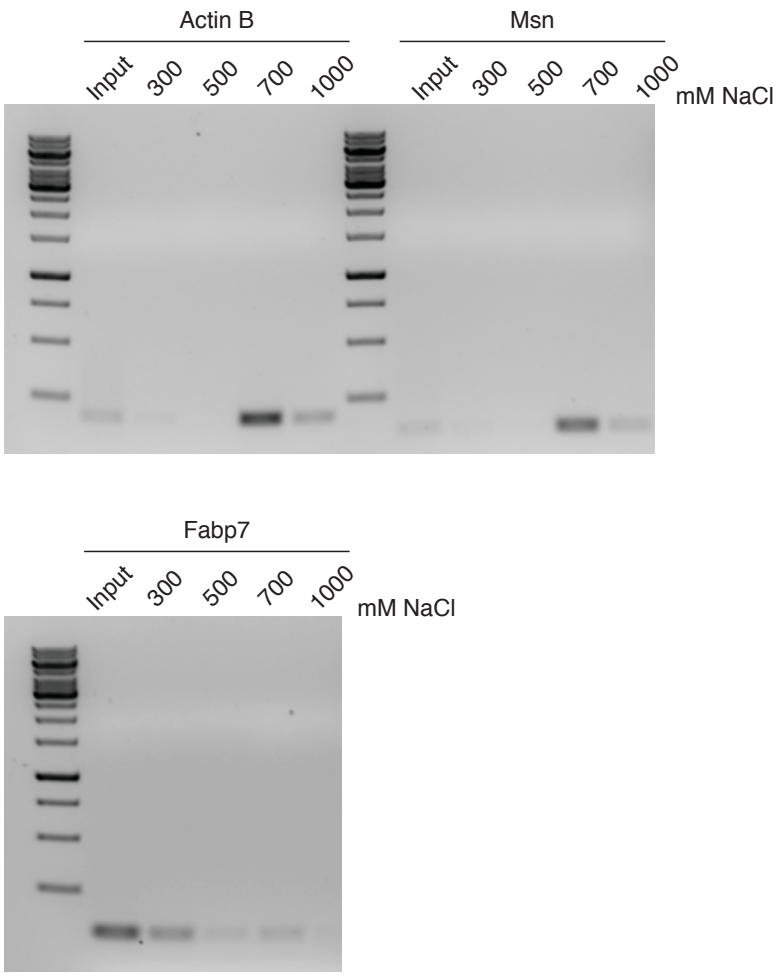


Figure 5.2 Optimisation of biotinylated CxxC affinity purification (BioCAP).
PCR analyses of the different fractions derived by the elution of gDNA from neutravidin-CxxC beads. ActinB, Msn and Fabp7 have CpG-rich promoter regions and were selected as unmethylated examples. Fabp7 promoter is known to be methylated in mESCs. As predicted, ActinB and Msn are enriched in fractions eluted with high salt (top panel). On the contrary, Fabp7 is enriched in fractions eluted with low salt, reflecting its predicted methylated status.

5.3. BioCAP-sequencing identifies loci characterised by LSH-dependent methylation.

Various studies have investigated the role of LSH at specific loci (Dennis et al., 2001; Fan, 2005; Myant et al., 2011; Xi et al., 2007). However, these studies provided insights into a static system, showing which specific genes are hypomethylated in the absence of LSH in a specific developmental stage or cell type. Combining BioCAP-sequencing and the culturing system set up in this study should provide not only information about the regions of the genome that are hypomethylated in absence of LSH at different stages, but also about the dynamics of these changes over time, depending on the regulatory processes within the cell.

High quality sequencing results (Appendix table 1 and figure 1) were analysed by mapping the reads to the mouse genome (Figure 5.1). Differentially methylated peaks in the *Lsh*$^{off/off}$ samples were identified by comparing reads found to the *Lsh*$^{off/+}$ controls to those found in samples lacking LSH, using the diffBind tool (Figure 5.3). At day 0, before any type of regulation of the DNA methylation is initiated by changes in culturing conditions, there was a small subset of peaks that corresponded to sequences hypomethylated in absence of LSH when compared to the control *Lsh*$^{off+f}$ ESCs (Figure 5.3, in red). This result is consistent with the previously described HPLC analyses (Figure 4.2). Interestingly, some hypermethylation was also detected at this time point (Figure 5.3, in blue), suggesting a mechanism of redistribution of the modification, including loss and acquisition of methylation at some loci in the LSH depleted cells. Subsequent culturing in 2i-containing medium leads to depletion of DNA methylation and inactivation of *de novo* methyltransferases. This step is then followed by the reactivation of the *de novo* methylation machinery caused by the culturing of the cells in serum. The sequencing at this point showed a larger number of peaks in *Lsh*$^{off/off}$ ESCs differentially methylated compared to the control cell line, almost 5000 peaks more than at day 0 (Figure 5.3). As seen in the HPLC analysis (Figure 4.2), after depletion of 5meC there is a recovery time

before acquisition of the 5meC to levels comparable to day 0 and this time is apparently longer for LSH depleted cells. Therefore, accumulation of hypomethylated regions in the genome in cells cultured for 2 days in serum was consistent with what expected. Interestingly, at the EBs stage, the last time point of the experimental system, there were still some differentially methylated regions in absence of LSH. Despite the absence of regions that accumulated 5meC, there were still hypomethylated regions in the *Lsh*$^{off/off}$ EBs compared to the control cell line (Figure 5.3, in red). The total number of peaks mapping at differentially methylated regions in *Lsh*$^{off/off}$ EBs decreases comparing to day 2 in serum, but there was still an enrichment compared to day 0. The overall dynamic of change in 5meC is similar to what was shown by HPLC in chapter 4. In fact, at the same time point a difference in the percentage of 5meC between the two genotypes was detected. As mentioned before, this indicates that *de novo* methylation can occur in absence of LSH, but it might be less efficient or slower. Furthermore, the decrease in the number of differentially methylated regions might indicate that a subset of loci can get methylated and some others cannot. However, it cannot be excluded that all the loci found hypomethylated in the absence of LSH at day 2 in serum-containing medium gain methylation and then, concomitantly with the beginning of differentiation, some loci cannot correctly retain methylation and therefore are found unmethylated in EBs, while in the control EBs the same loci are stably methylated.

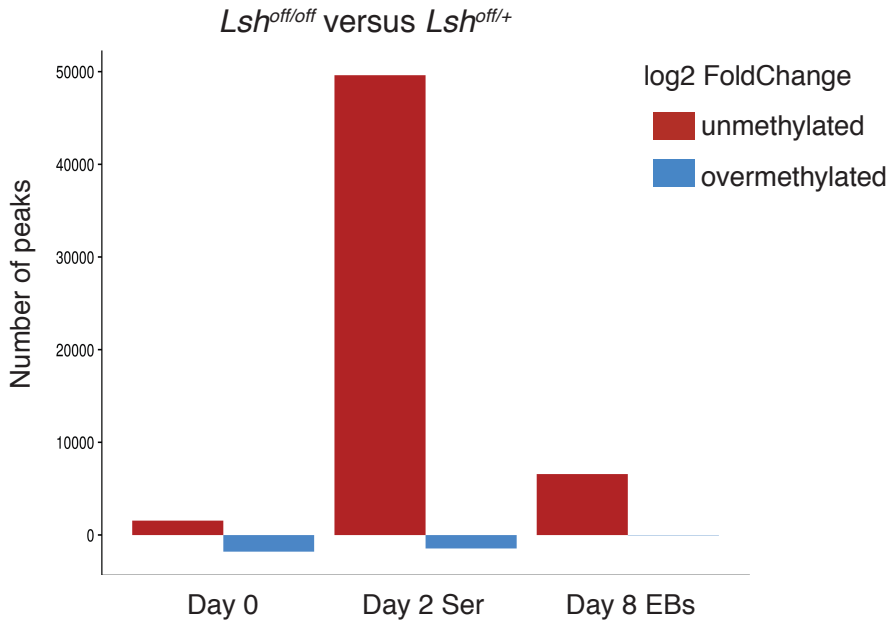


Figure 5.3 Differentially methylated peak analysis shows accumulation of methylation depleted genomic regions at day 2 serum and EBs.
The total number of peaks mapping to differentially methylated regions found in *Lsh*$^{off/off}$ ESCs or EBs after comparison to *Lsh*$^{off/+}$ ESCs or EBs at the 3 time points of the time course experiment analysed by BioCap-seq are shown on the x axis. Positive log2 fold change values indicate overrepresented DNA sequences, hence unmethylated regions and are represented in red. Negative log2FC values indicate under represented sequences and therefore hypermethilated genomic regions and are represented in blue. While there is not a big number of peaks mapping to hypomethylated loci at day 0, the enrichment becomes apparent when the ESCs are recovering from 2i-induced demethylation followed by differentiation. The enrichment in number of peaks decreases at the EBs stage, but remains higher than at day 0.
False Discovery Rate <0.05.

To distinguish between the hypothesis that a subset of loci requires LSH for being correctly methylated during the first wave of *de novo* methylation and remain hypomethylated until the EB stage or that loci hypomethylated in EBs lacking LSH differ from those found at day 2 in serum, the BioCAP-sequencing was a great tool.

The first step into analysing in details the regions of the genome hypomethylated in LSH depleted cells was determining the distribution of the differentially methylated peaks throughout the genome. Therefore, the genome was divided into features, as indicated in figure 5.4, and both the peaks found in hypomethylated and hypermethyalted regions (Figure 5.3) were assigned to a feature of the genome. While the majority of loci that acquire methylation represented promoter regions, interestingly most of the hypomethylated regions mapped to distal intergenic regions (Figure 5.4). More specifically, the percentage of peaks mapping to distal intergenic regions in cells lacking LSH increased from approximately 20% at day 0 to 70% day 8 of EBs formation. This suggests that the effect of the absence of LSH detected at the EBs stage might be mostly caused by hypomethylation of these distal intergenic regions, which include repetitive elements. Moreover, the percentage of hypomethylated peaks localising at promoter regions decreased from almost 40% at day 0 to about 10% day 2 in serum and remained stable at day 8 of EBs development. Bearing in mind the way smaller number of peaks that were differentially methylated between $Lsh^{off/off}$ and $Lsh^{off/+}$ ESCs at day 0, it is noticeable that this percentage decreased drastically, but then remained consistent between day 2 in serum and day 8 EBs. This might indicate that the promoters that do not gain methylation at day 2 in serum remain unmethylated at the stage of EBs and might be specific targets of LSH activity. However, it remains unclear at this point if these loci are the same at the two time points and therefore further analyses are necessary.

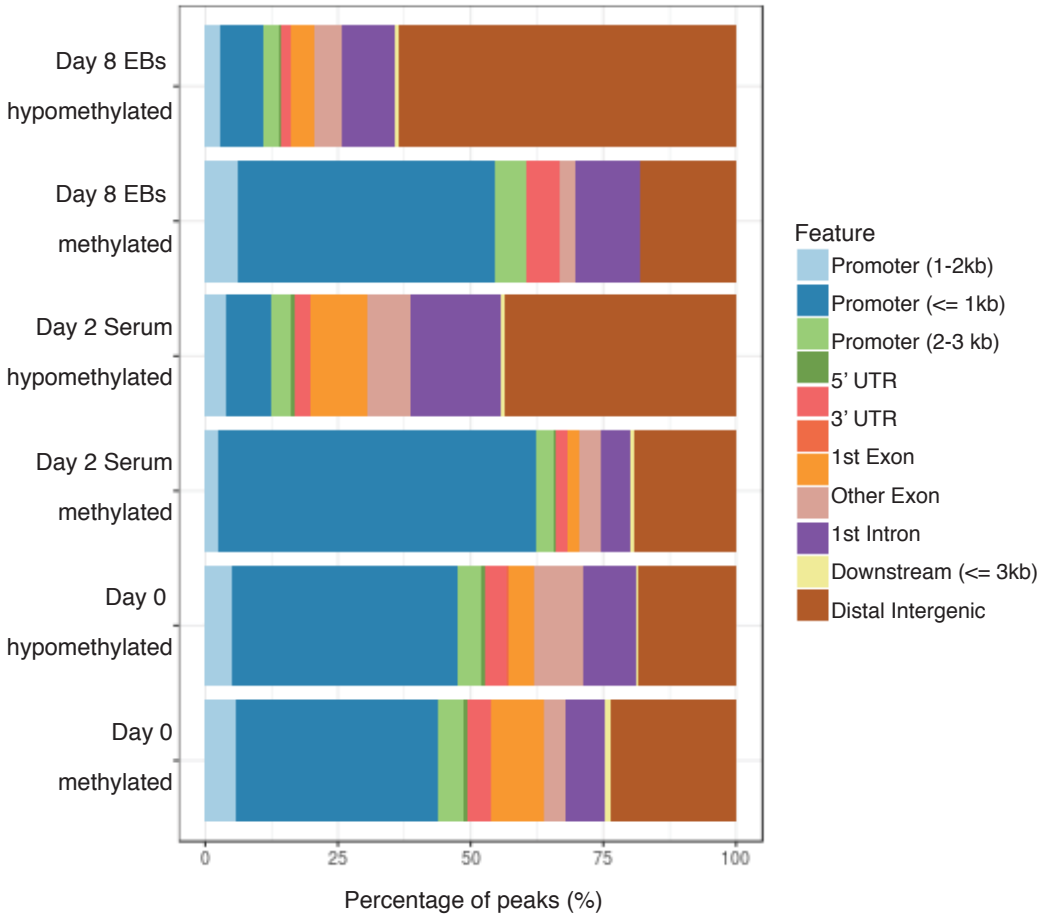


Figure 5.4 Feature distribution of differentially methylated regions found in *Lsh*$^{off/off}$ samples. Loci identified in the differential peaks analysis were categorised based on their location in the genome. These 11 categories, also referred to as features, are listed on the right, together with the colour assiged to each. Percentage of peaks falling into each category for every time point is shown, for loci depleted of methylation and enriched in methylation. The hypomethylated peaks fall mostly into the category of distal intergenic regions at day 2 in serum and day 8 EBs. On the other hand, the methylated peaks are mostly localised at promoter regions at day 2 in serum and day 8 of EBs.

These experiments provide first of all evidence for the reliability of the method used for sequencing due to the consistency with the analysis performed by HPLC and discussed in previous chapters. Furthermore, these experiments show that the hypomethylated regions of the genome in LSH-deficient cells represent mainly distal intergenic regions, but this approach did not provide detailed information on the specific regions involved. Interestingly, some promoter regions are found unmethylated in absence of LSH, supporting the hypothesis that LSH is important for methylation of the cytosine, not only at repetitive elements, but also at unique genomic regions, including promoters. Further analyses will address the question of which promoters are hypomethylated in the absence of LSH and, importantly, if these remain the same after the demethylation stage at day 2 in serum up to the EBs formation or not.

Finally, BioCAP-sequencing analysis result support HPLC data, highlighting a time frame of activity of LSH, as the hypomethylation appears massively after the demethylation wave, induced by the treatment with 2i, and remains evident until the EBs stage. It will be important to identify the specific genes and distal intergenic regions that are demethylated in the absence of LSH in order to understand if these genes are depleted of 5meC and maintained so through the time course experiment, suggesting that LSH might be necessary to methylate these regions *in vivo* during development after the first reprogramming of the 5meC. Or if, alternatively, the genomic regions involved change after the activation of the *de novo* methylation machinery and the subsequent differentiation, suggesting that 5meC loss in absence of LSH can be overcame first and that LSH therefore becomes important later in development, when the genome is more heterochromatic.

5.4. LSH is required for methylation at non-repetitive sequences

After looking at the distribution of differentially methylated regions in the genome of *Lsh*$^{off/off}$ cells, two main regions of interested emerged: distal intergenic regions, for their

abundance of differentially methylated peaks, and promoter regions, for the consistent percentage at day 2 serum and day 8 EBs. As previously shown by various studies (Dennis et al., 2001; Fan, 2005; Myant et al., 2011; Xi et al., 2007), LSH is required for methylation at specific unique loci, such as the *Rhox* genes and *Gm9* in MEFs. Therefore, further characterisation of genes with demethylated regions was performed.

We first looked at the Gene Ontology terms in which the regions of differential methylation could be categorised into (Figure 5.5). Comparison between $Lsh^{off/off}$ and $Lsh^{off/+}$ cells allowed to infer the most frequent terms describing the differential peaks. Interestingly, of the 3 categories, namely biological processes, cellular components and molecular function, at day 0 70% of the small number of differentially methylated genes identified belong to the biological process category, but this percentage decreased during the time course experiment (Figure 5.5). In fact, at day 2 in serum, 65% of the genes were categorised in the biological process category and at day 8 EBs this dropped to 50%. The decrease in genes related to this category was accompanied by an increase in the number of differentially methylated genes that were included in the cellular components or molecular function category from 10% and 15%, respectively, at day 0 to 27% and 22% at day 8 EBs (Figure 5.5). This change in the categorisation of the Gene Ontology terms suggests a shift in the regions of the genome hypomethylated in absence of LSH and might indicate a specificity of the protein towards some regions of the genome, arising during the stage of the time course experiment mimicking early phases of the embryonic development.

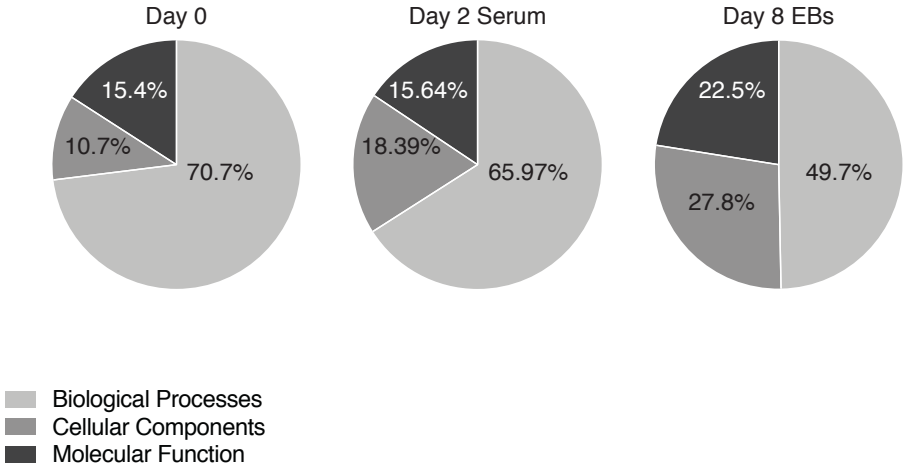


Figure 5.6 Gene Ontology analysis of hypomethylated genes shows differences in the distribution of GO terms over time.
Pie charts representing the percentage of hypomethylated genes described by each GO category at each time point of the experiment. The percentage of hypomethylated loci involved in formation of cellular components increases over the time course experiment. The time point in the experiment are shown on top of each the pie chart.

The GO categories falling into the 3 broader terms are listed in Figure 5.6. The most important feature to highlight is that there is a slow transition beginning at day 0 and moving towards day 8 EBs into terms more and more specific for neuronal specialisations. More specifically, while the categories coming up from the analysis at day 0 are more general, such as metabolism, cell and protein binding, or focused on the embryo development, such as embryonic morphogenesis or developmental process, and only some terms concern the neuronal lineage, such as synapsis, these change already when cells come out of the demethylation wave at day 2 in serum, and is even more pronounced at the EBs stage (Figure 5.6). Expression of these genes at day 8 EBs is expected, due to the addition after 4 days in culture of retinoic acid, which drives EBs towards neuronal lineage. Interestingly, however, almost all the terms describing hypomethylated regions

specifically in absence of LSH are connected to neuronal development or chemical activity of the nervous system. On one hand this result shows that the regions depleted of 5meC in absence of LSH partially differ depending on the developmental stage and, on the other hand, it suggests that LSH activity might be bound to the neuronal lineage. This feature becomes evident only later in the time point experiment because these are regions of the genome regulated only when cell differentiation is activated. If the methylation does not occur correctly at these regions, this might lead to impaired differentiation towards the various neural cell types.

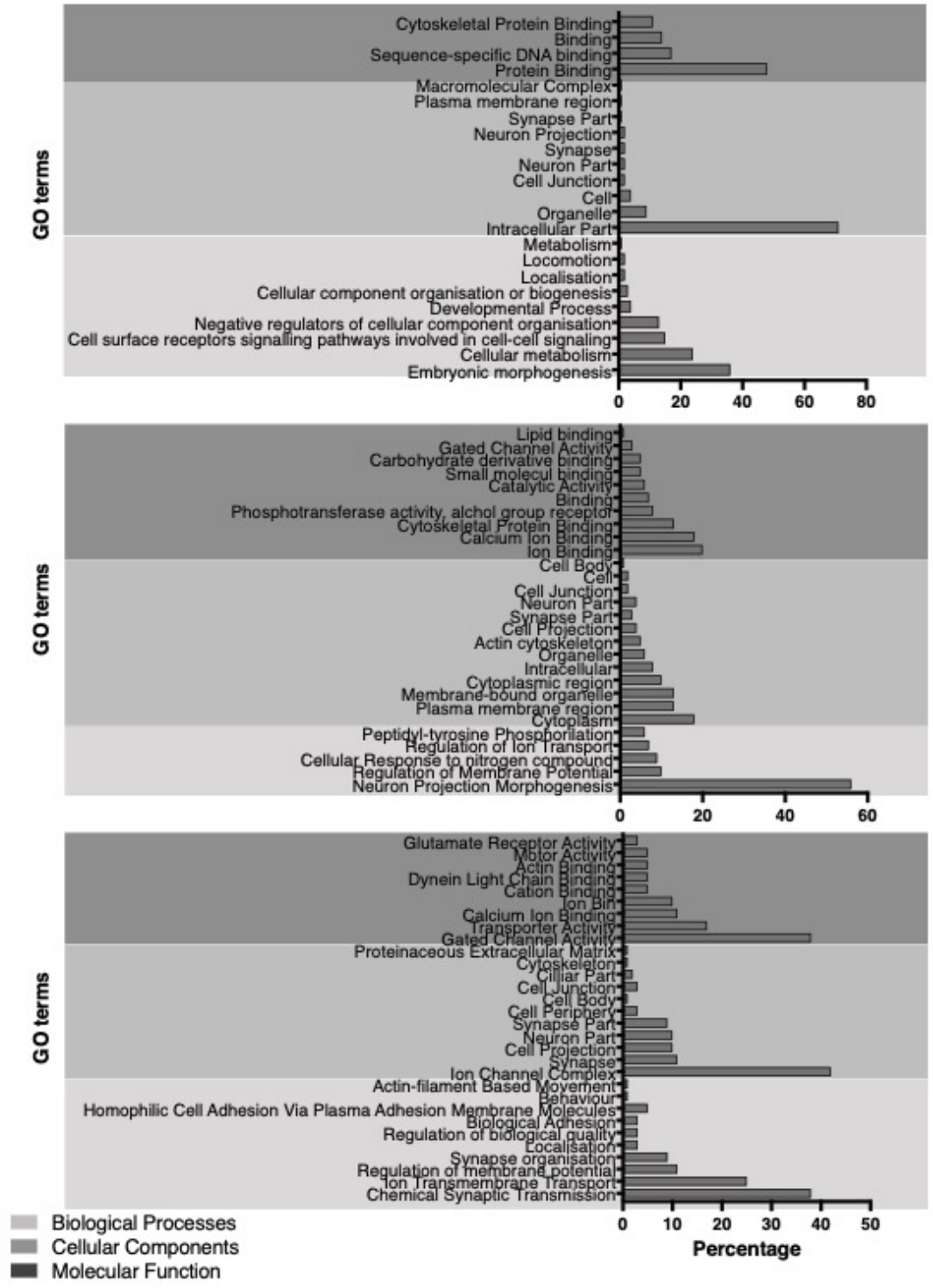


Figure 5.6 Gene Ontology terms analysis of hypomethylated loci.
gProfile analysis of genes identified as differentially methylated shows a progressive change in the components hypomethylated in absence of LSH. Boxes in different shades of greys are aimed to identify the 3 main GO terms, under which the specific terms are grouped.

To characterise further the peaks that map across promoter regions, we decided to ask whether the differentially methylated regions in *Lsh*$^{off/off}$ samples compared to *Lsh*$^{off/+}$ were the same at day 2 in serum and day 8 EBs. Over the total number of peaks mapping to these regions of the genome at the two time points, around 1000 are found in both samples. Of these genes, 77% are protein coding (Figure 5.7A). These genes were then analysed to explore the GO terms and it was found that around half of these genes belong to the biological process category (Figure 5.7B). This category distribution is very similar to what detected in day 8 EBs (Figure 5.5). When analysing the categorisations within the 3 aspects, biological processes, cellular components and molecular function (Figure 5.7 C), once again it seems that most of the genes fell into categories related to the nervous system, similarly to what was observed in the previous analysis (Figure 5.6). Altogether, these data show that there is a fraction of promoters that in absence of LSH are not methylated at day 2 in serum and they are kept in this state until the last time point of the experiment.

The characterisation of differentially methylated peaks identified by BioCAP-sequencing provided insights into the functions of the genes that are found hypomethylated in absence of LSH. What became evident from the changes in GO categories distribution following the regulations of 5meC during the time course experiments is that there is a gradual increase in the number of terms related to neuronal system. This was expected in the EBs samples, in both genotype, as the addition of retinoic acid is used to trigger differentiation into neuronal linage. However, comparing the differentially methylated peaks mapping at promoter regions at day 2 in serum and day 8 EBs, clarified that there is a subgroup of promoters that are already hypomethylated at day 2 in serum in absence of LSH and that these promoters do not acquire 5meC by the time the EBs are formed, at day 8. This, together with the presence of promoter regions that can gain methylation and others that lose it, suggests that there are some loci that might be the main targets for regulation by LSH as these promoters are unable to gain methylation in the absence of the protein, once it is lost. At the same time, there are other regions that can gain DNA methylation, but possibly require longer time compared to other

regions, despite the generally more accessible state of the chromatin in cells lacking LSH. Finally, when the cells go through differentiation, mimicking what happens in the embryo after E6.5, some genomic regions are not capable of acquiring 5meC, as suggest by the emerging at day 8 EBs of GO category terms related to the nervous system. Altogether, this suggests the potential existence of different sets of targets, the first requiring LSH for *de novo* methylation at the formation of the blastocyst and the second requiring LSH at the epiblast stage. However, it remains unclear what determines the need for LSH at those specific loci and what triggers its activity during the two main regulatory events of DNA methylation in the early stages of embryonic development.

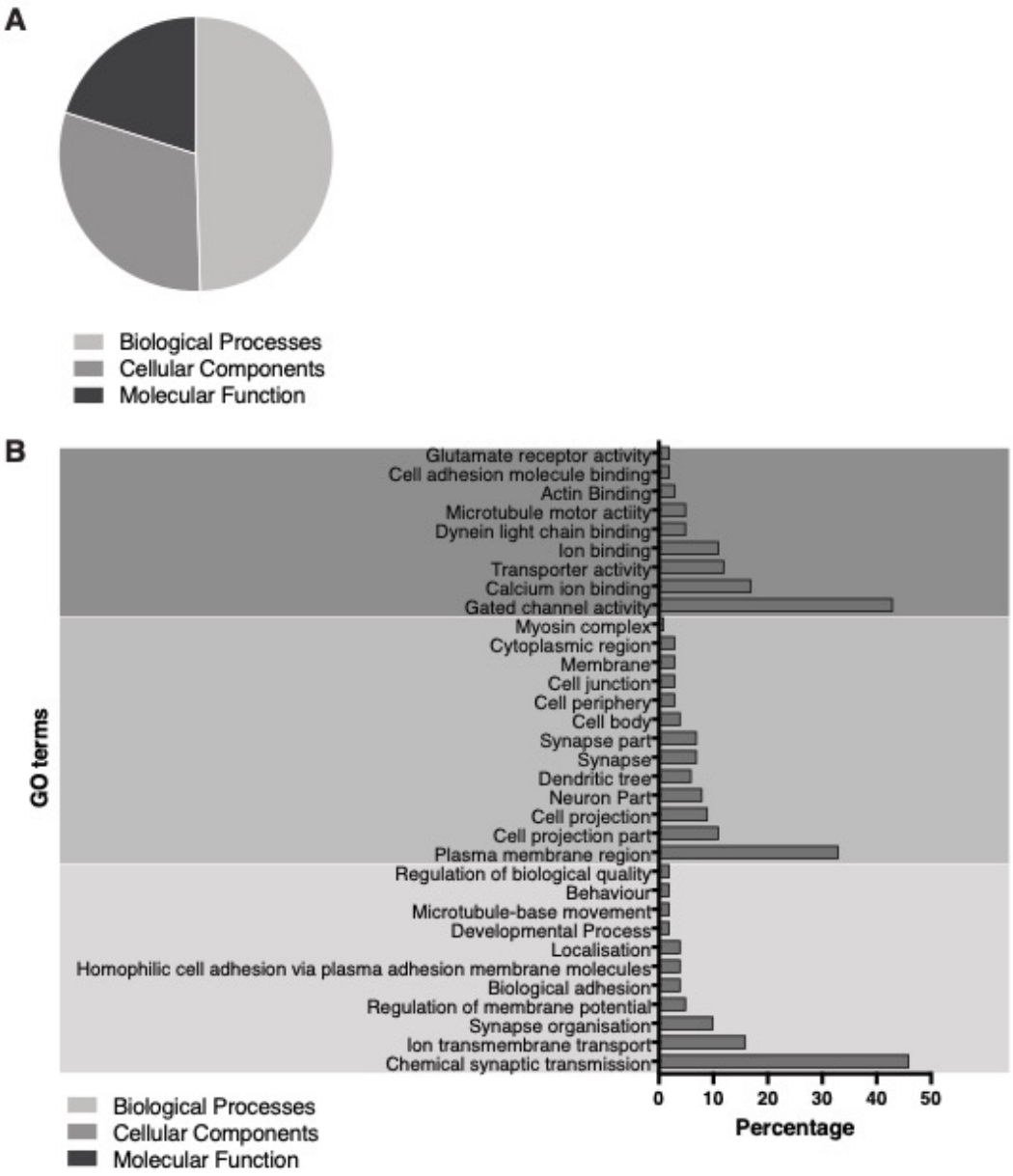


Figure 5.7 A subset of promoters are hypomethylated both at day 2 in serum and at EBs stage.
A.Gene Ontology categories associated with genes with hypomethylated promoters at day 2 serum and day 8 EBs.
B.Gene Ontology terms associated with genes with hypomethylated promoters at day 2 serum and day 8 EBS. The percentage of genes described by a term is shown on the x axis. The terms are grouped by aspect and represented by different background colours.

5.5. LSH is crucial for methylation of IAPs, a subfamily of repetitive elements, early in development.

From the feature distribution analysis, it emerged that the highest percentage of differentially methylated peaks map at distal intergenic regions. Within this feature, many genomic regions could be accounted, such as noncoding DNA, regulatory regions, enhancers, noncoding RNAs and repetitive elements. Since it was previously shown that LSH is necessary for the regulation of DNA methylation at repetitive elements (De La Fuente et al., 2006; Huang et al., 2004), and especially at endogenous mouse retrotransposons IAPs (intracisternal A-particle), we decided to focus our attention on these retroviral elements. Doing so, will provide, not only validation for the method used, but also further insights into which repetitive elements, other than IAPs, are hypomethylated in the absence of LSH, and what is the time frame for this regulation. Such information is crucial information to better understand the consequences of LSH KOs in mouse development.

Repetitive elements (RE) contained in heterochromatic regions are derived from transposable elements and make up around half of the mammalian genome (45% in human, (Lander et al., 2001). There are 2 main classes of repetitive elements, the first includes retrotransposons that produce reverse transcriptase, similarly to retroviruses, and is the most abundant. The second class include DNA transposons, are not active in the mouse and human genomes and are about 3% of the genome (Lander et al., 2001). These elements can be silenced by various post translational modifications and DNA methylation. It is crucial to understand the mechanism leading to repression of these very abundant elements in the genome, as it has been shown that reexpression of these is associated with diseases, such as cancer, and with a more sever outcome in those instances (Nguyen et al., 2018).

Retrotransposons fall into 2 categories, those characterised by long terminal repeats (LTR), such as the endogenous retroviral elements ERV, that account for about 10% of

the mouse genome (Lander et al., 2001). Belonging to the ERVK class are also the IAPs elements, retroviral elements containing 2 LTR at each end of the sequence, which can drive expression in cancer such as lymphoma (Lane et al., 2003; Qin et al., 2010). The second class of RE is characterised by the absence of LTRs and includes long interspersed elements (LINE) and short interspersed elements (SINE).

As mention in the introduction of this thesis, LSH activity has been linked to the expression of repetitive elements, such as LTRs in brain and liver tissue of KO mice for *Lsh* (Huang et al., 2004) and IAPs in oocytes and mouse embryonic fibroblasts (De La Fuente et al., 2006; Myant et al., 2011). Therefore, we analysed the differentially methylated regions in $Lsh^{off/off}$ cells compared to $Lsh^{off/+}$ cells at various time points in the time course experiment and interrogated more specifically the peaks mapping over repetitive sequences. More than 50 families of RE were analysed and the peak distribution was compared to a random distribution in order to establish a significance for the mapping of each of the families. More specifically, among the top 20 families with the highest number of mapped hypomethylated reads, only few showed significant enrichment in these reads compared to a random distribution (Figure 5.8). At day 0 there are some differentially methylated RE, namely Alu elements (member of the non-LTRs) and the simple repeats (Figure 5.8, bottom panel). After the cells go through DNA methylation reprogramming, there is an expected increase in the number of hypomethylated regions, including repetitive elements. In fact, at day 2 in serum, more than 25% of the peaks map at ERVK family, comprising the IAPS elements. The second most abundant RE are the L1, belonging to the LINE family, and the third are yet the simple repeats (Figure 5.8, middle panel). Finally, at day 8 of the EBs development, 46% of the total peaks map at ERVK elements and less than 40% to L1 (Figure 5.8, top panel). Altogether, this analysis shows that the type of repetitive element hypomethylated in $Lsh^{off/off}$ cells changes after the loss of 5meC in 2i-containing medium. Instead of acquiring DNA methylation due to the activation of the *de novo* methylation machinery in serum-containing medium, in absence of LSH ERVK and L1 elements specifically stay hypomethylated until the formation of the EBs, while the other elements acquire methylation or are not significantly

differentially methylated compared to the control. This suggests a LSH-dependent methylation mechanism that operates at these repetitive elements initiated after loss of 5meC. This methylation depletion is not overcome by the *de novo* methylation, possibly impaired by the absence of the protein.

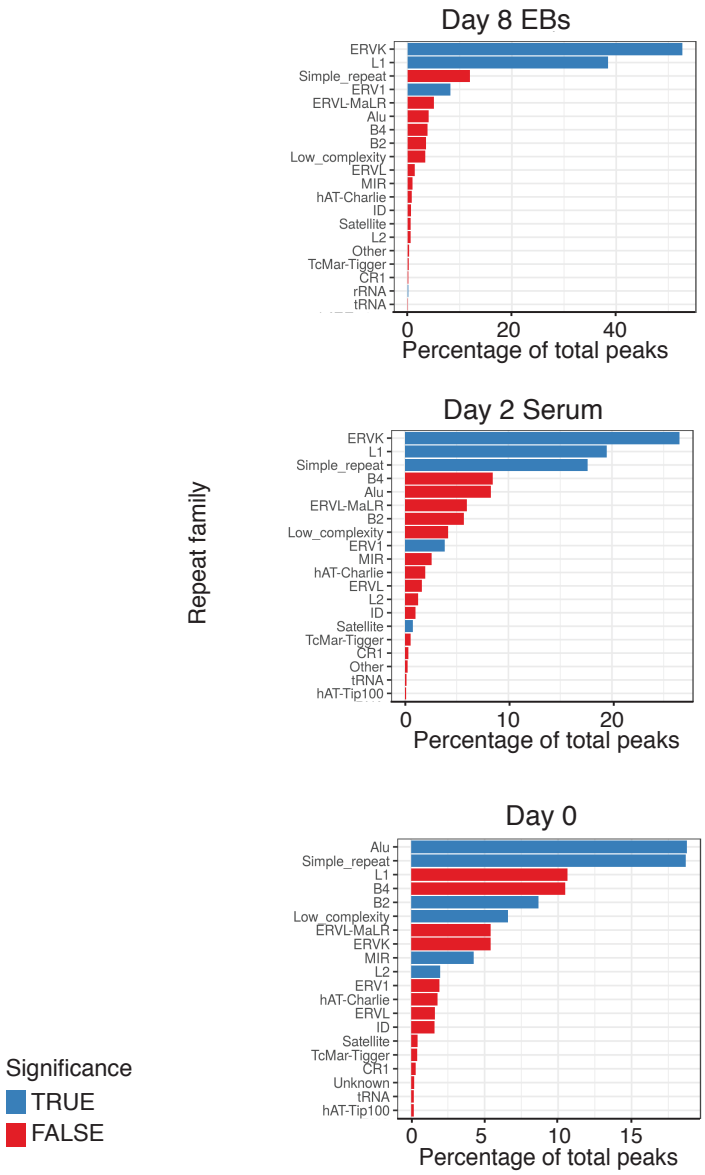


Figure 5.8 Families of repetitive elements differentially methylated in the absence of LSH change over time.

Top 20 repetitive element families that are differentially methylated in *Lsh*$^{off/off}$ compared to *Lsh*$^{off/+}$ cells are shown at day 0, day 2 serum and day 8 EBs. The percentage of peaks that overlap with each repetitive element family is shown. Only elements represented by the bars in blue are significantly enriched over a random distribution of the reads, with Permutation test p-value <=0.05. The most significantly differentially methylated repetitive element families change between day 0 and later time points of the time course experiment, to then stay durably unmethylated.

To better understand how IAPs methylation status changed through the time course experiment in presence and absence of LSH, reads of unmethylated CpG mapping at IAPs annotated sequences were analysed. While at day 0 there were virtually no reads mapping at IAPs in the genome of both cell lines analysed (Figure 5.9, left panel), at day 2 in serum there were peaks mapping at both LTRs, as might be expected due to the treatment of the mESCs with 2i-containing medium (Figure 5.9, central panel). There was however an increase in number of reads mapping at those regions in both biological replicas of *Lsh*$^{off/off}$ ESCs, compared to *Lsh*$^{off/+}$ ESCs (Figure 5.9, middle panel). At day 8 of EBs development, almost no reads mapped at IAPs in *Lsh*$^{off/+}$ ESCs, while there were still reads mapping at LTRs of the IAPs *Lsh*$^{off/off}$ EBs, despite a difference in the number of peaks between biological replicas (Figure 5.9, right panel). This difference could be explained by variability of EBs differentiation compared to mESCs in culture. The decrease in the number of reads mapping to IAPs at day 8 EBs compared to day 2 serum might suggest that some IAPs can always be methylated, both in *Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ EBs. However, there is yet a large proportion of these elements that in absence of LSH do not gain methylation. These results are consistent with LSH activity specifically targeting IAPs as was shown previously (De La Fuente et al., 2006; Huang et al., 2004). However, our experiments show that the hypomethylation of these regions only occurs after depletion of DNA methylation by treatment of ESCs with 2 inhibitors and cannot be recovered completely, at least in the time frame we investigated. We hypothesise that absence of LSH could lead to similar hypomethylation dynamics *in vivo*, impairing 5meC acquisition after the first DNA methylation reprogramming event.

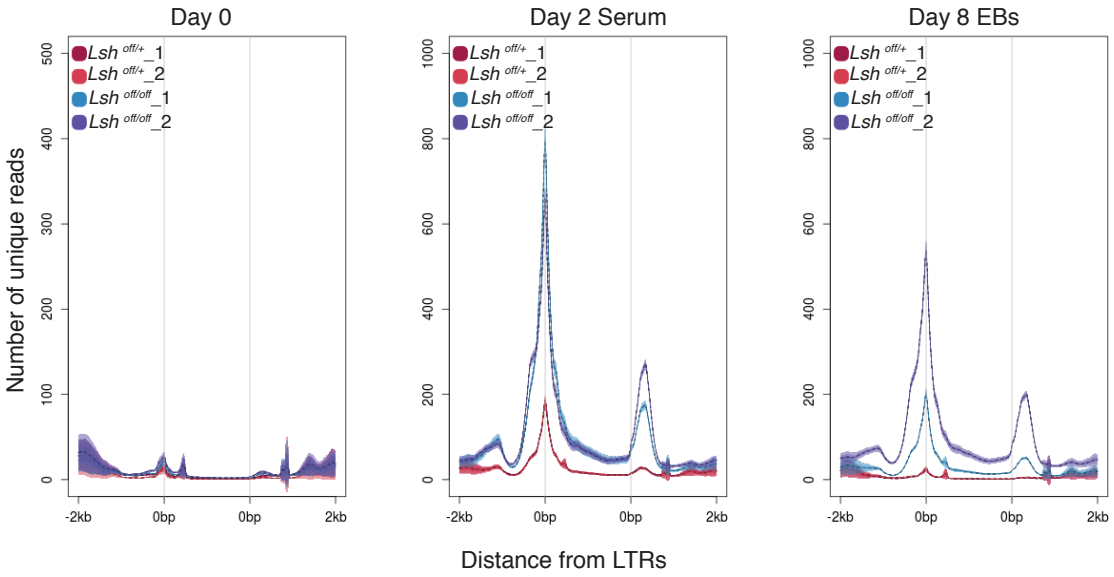
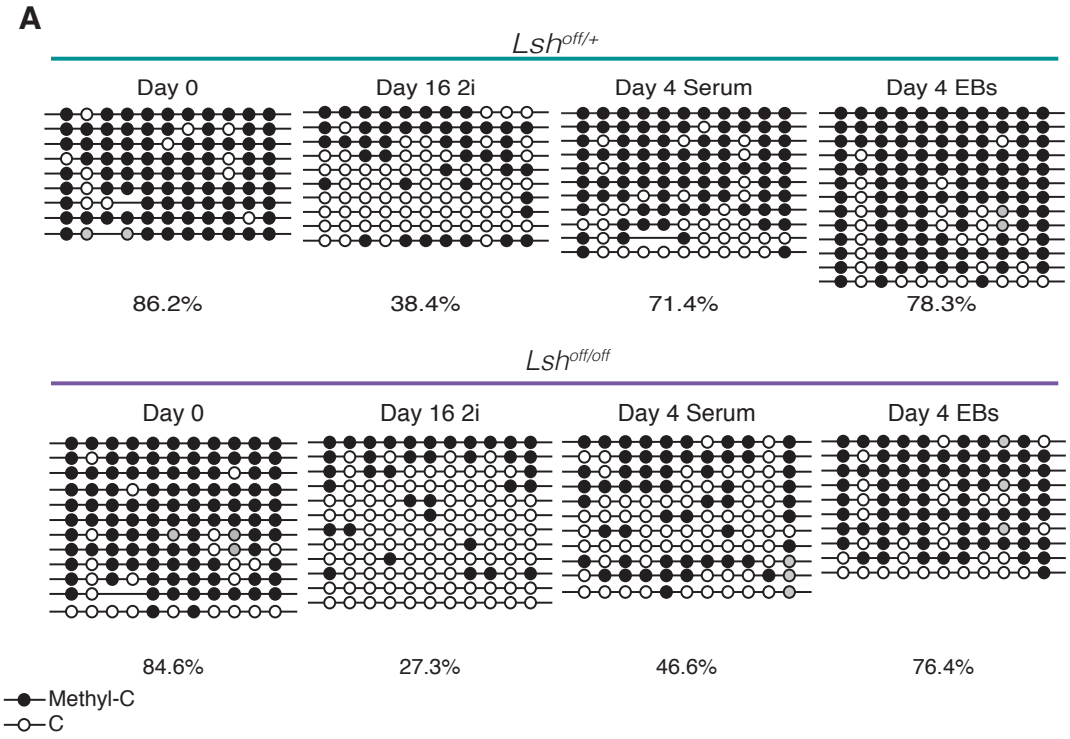


Figure 5.9 IAP elements are hypomethylated in the absence of LSH.
Plot of unique BioCAP reads across annotated IAPEz elements. At each time point, all the samples are shown in different colours. On the x axis the distance from the LTRs is annotated. On the y axis is annotated the number of reads mapping to that site. Almost no reads map to IAPEz at day 0. At day 2 in serum, these regions are enriched in reads, in all the samples analised. At day 8 of EBs development, while no reads map at these regions in *Lsh*$^{off/+}$ EBs, there are still reads mapping at these regions in *Lsh*$^{off/off}$ EBs, indicating hypomethylation of the elements when LSH is absent.

Misregulation of DNA methylation at IAPs was shown to activate the expression of these elements, for example an IAP element inserted into the *Agouti* locus can be activated by chemical treatment and other stimuli leading subsequently to activation of the *Agouti* gene in mice (Lane et al., 2003). Therefore, it was important to analyse the expression of these elements in the time course experiment. First, the hypomethylation of the elements was validated by bisulfite-sequencing. The analysis of the 5' UTR with this technique showed no difference in DNA methylation between *Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ ESCs at day 0, and a subsequent loss of 5meC at day 16 in 2i-containing medium (Figure 5.10A).

After 4 days in serum post 2i, the difference in 5meC between *Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ ESCs became more pronounced. However, this region of the IAPs gained methylation to comparable levels in the two cell lines at day 4 of the EBs development (Figure 5.10A). Unfortunately, the change in DNA methylation detected with this technique was not completely consistent with what was shown by the BioCAP-sequencing. However, it is important to highlight that Bisulfite-sequencing PCR only analysed the methylation of a specific and limited set of CpGs, compared to sequencing techniques that provide information on the whole sequence. Moreover, when analysing the expression of the IAPs by qRT-PCR (Figure 5.10B), an increase in the expression of IAPs was detected in both genotypes at day 2 in serum, as expected after treatment with the 2 inhibitors. Expression was even higher at day 8 of EBs, especially in cells lacking LSH. Furthermore, at both time points of the experiment, the expression of IAPs relative to day 0 was higher in *Lsh*$^{off/off}$ samples, consistent with BioCAP-sequencing results. The increase in the expression at day 8 in both *Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ EBs, despite a decrease in the DNA methylation might be dependent on an accumulation of IAP mRNA produced during previous stages of the experimental system and not completely degraded yet. Despite difference in expression of IAPs between *Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ samples not being statistically significant, it suggests that the hypomethylation of IAPs leads to expression of these elements, in both *Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ EBs, but to be a bigger extent in *Lsh*$^{off/off}$ EBs. It is not possible to conclude whether the difference in expression between cells with the two genotypes could lead to specific consequences for cell differentiation in the absence of LSH. It is however important to highlight that expression of these RE is initiated only after the depletion of DNA methylation and is more prominent in cells lacking LSH, highlighting a LSH dependent mechanism for the 5meC deposition at these sequences concomitant with acquisition of *de novo* methylation across the genome.



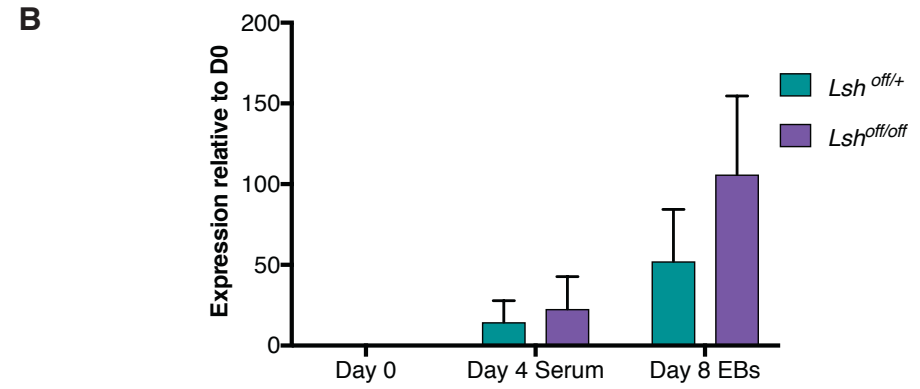


Figure 5.10 The IAP elements are hypomethylated in the absence of LSH and differentially expressed.
A.Bisulfite sequencing analysis of the 5'LTR of IAP elements shows loss of methylation at this end of the element and a slow recovery of DNA methylation in serum and EBs.
B.Quantitative RT-PCR detects expression of IAPS at day 4 in serum post 2i and in day 8 EBs, relative to day 0. GAPDH was used as internal control. Error bars represent standard deviation between mean values. The expression of IAPs increases at every step of the time course, consistantly with the depletion of 5meC. The increase in expression is higher in *Lsh*$^{off/off}$ EBs compared to the *Lsh*$^{off/+}$ control, despite the detected increase in the expression of these repetitive elements in cells of both genotypes.

5.6. Conclusions

Experiments described in previous chapters showed that LSH is required for *de novo* methylation throughout the genome, consistently with work previously published (Myant & Stancheva, 2008; Dunican et al., 2013; He et al., 2016; Jiang et al., 2017; Joshi et al., 2011; Myant et al., 2011), and suggest a time frame in which the role of LSH becomes relevant for the regulation of DNA methylation. In this last chapter, I used a high throughput sequencing technique to explore whether the absence of LSH influenced the methylation of specific genomic loci, i.e. genomic features. Identification of differentially methylated regions in $Lsh^{off/off}$ cells compared to $Lsh^{off/+}$ cells showed that, subsequently to the depletion of 5meC induced by culturing the cells in 2i medium, during the methylation recovery phase and the differentiation, the absence of LSH leads to demethylation of distal intergenic regions, promoter regions and gene bodies. When we interrogated genes and promoters depleted of 5meC, we found that only about 10% of promoter regions are hypomethylated at day 2 serum and day 8 EBs. As expected, most of the promoters are hypomethyalted only transiently at day 2 in serum, since at this time point of the experiment deposition of methyl groups has just been triggered by the change in culturing medium. This suggests the existence of two phases in the DNA methylation deposition in the absence of LSH. In the first phase, some genes acquire 5meC when *de novo* methylation machinery is activated, but require more time to be fully methylated compared to $Lsh^{off/+}$ cells, consistent with the deeper loss of methylation in 2i-containing medium occurring in LSH deficient cells possibly due to the increased accessibility of chromatin to demethylases enzymes. Later on in the time course experiment, when differentiation is triggered by addition of RA to the culture, a new phase of regulation of DNA methylation starts. At the EBs stage, in fact, most of the hypomethylated genes in $Lsh^{off/off}$ cells were categorised by GO terms related to the nervous system development or regulation. Since in the experimental design, the EBs were taken through the first steps of the differentiation into the neuronal lineage by addition of retinoic acid and withdrawal of LIF, this supports

previous findings in neural stem progenitor cells (Han et al., 2017) that LSH might be crucial for the correct regulation of DNA methylation and possibly the expression of those genes. A small number of promoter regions are hypomethylated immediately after the recovery from 2i-medium occurs, and these remain hypomethylated throughout EB formation and differentiation. These genes might be the main, direct or indirect, targets of LSH and are in fact associated to the neuronal lineage differentiation path. This might indicate that those genes are particularly sensitive to the absence of LSH and may not be correctly methylated early in the embryonic development, leading to misregulation of gene expression in the later differentiation stage. Further analysis of the expression of these genes in our cell culturing system and then *in vivo* would provide further insights into the consequences of the loss of DNA methylation in cells lacking LSH and whether this is relevant to embryonic development.

Consistently with what was previously shown (De La Fuente et al., 2006; Huang et al., 2004; Myant et al., 2011), distal intergenic regions and more specifically RE were found hypomethylated in LSH depleted cells. These regions, very much like the neuronal genes discussed above, are methylation depleted at day 2 in serum post 2i and remain hypomethylated in the EBs. ERVK and LINE1 elements were the main hypomethylated elements in cells lacking LSH after treatment with 2i, and IAPs in particular were further analysed and were not only hypomethylated, but also expressed immediately after recovery from 2i medium and in day 8 EBs lacking LSH. These results are consistent with what was reported in literature in different cell models, indicating that these regions require LSH to regulate the methylation of the DNA probably at the exit of naïve state of pluripotency of the cells.

Altogether, the data described in this chapter show that LSH is dispensable for methylation of the majority of the genome sequenced with the BioCAP-sequencing approach, but required for methylation of neuronal associated genes and repetitive elements. Importantly, this role emerges *in vitro* always after the demethylation caused by culturing the cells with 2i and seems crucial for the regulation of the DNA methylation in embryonic bodies. Therefore, it is possible to hypothesise that *in vivo* LSH might be

required at specific genomic regions after blastocyst formation at day E3.5 of the mouse development and during subsequent differentiation.

6. CONCLUSION AND FUTURE OUTLOOK

Despite the big efforts which led to the publications of many studies on LSH activity in mammals, there is yet a lot to uncover. It has been very recently proven LSH's activity in remodelling chromatin, as a member of a remodelling complex with CDCA7 (Jenness et al., 2018). A role for the protein was shown in *de novo* methylation of DNA but it is yet unclear whether it has a role in maintenance of DNA methylation, too. It was shown that absence of the protein determines defects in the DNA methylation in MEFs, oocytes and embryos after day 12.5 of the development. However, no information on earlier developmental stages nor on the molecular mechanism by which the protein modulates DNA methylation throughout development are available. This study was aimed at addressing whether LSH is involved in DNA methylation *de novo* deposition or in maintenance of this epigenetic modification. Furthermore, an *in vitro* culturing system was optimised to attempt understanding when during development the protein becomes pivotal for 5meC regulation and which loci in the genome require LSH for achieving correct DNA methylation state. This *in vitro* culturing system used mESCs to mimic early phases of the mouse embryonic development, between day E3.5 and E12, and hence facilitated the study of DNA methylation dynamics in absence of LSH activity.

Optimisation of the culturing system was the first step to investigate how 5meC levels are affected by absence of LSH in response to reprogramming events. The use of medium supplemented with inhibitors for GSK3beta and WNT allowed depletion of DNA methylation to a level similar to the blastocyst's, mimicking at the same time the loss of methylation occurring in the paternal and maternal genomes at the beginning of the embryonic development (Figure 1.6). *Lsh*$^{off/off}$ mESCs cultured in this condition show bigger loss of 5meC compared to the control cell line. The recovery of the 5meC determined by the initiation of *de novo* methylation in serum-containing medium also is affected by lack of LSH. In fact, *Lsh*$^{off/off}$ mESCs do not recover methylation to a level

comparable to the initial, despite accumulating methyl groups in a faster manner. LSH being capable of remodelling heterochromatic regions, it is hypothesised that its absence may influence the compaction of the chromatin, by making it more accessible. This hypothesis was supported by preliminary data obtained by chromatin digestion by MNAse, showing a more open chromatin state in cells lacking LSH at day 8 EBs. The accessibility of the chromatin should be investigated further, by using a high throughput approach such as MNAse-seq at different time points in the time course experiment. Despite not predicting sever effects on the chromatin structure of the whole genome, this experiment would provide more information into how the chromatin accessibility changes at target regions in response to the reprogramming events in absence of LSH and prove its importance in compaction of the chromatin.

The analysis of the intermediate product of oxidation of 5meC to cytosine, 5hmeC, provided some evidence to support a role of LSH in *de novo* methylation and not in maintenance of methylation. In fact, if the further loss of 5meC detected in $Lsh^{off/off}$ mESCs was accountable to deficiency in the maintenance of methylation, no accumulation of 5hmeC would be detected. Interestingly, this is not the case. Although this is not enough evidence to prove a unique role of LSH in *de novo* methylation, it is sufficient to say that in this instance, maintenance of DNA methylation is not imputable for the differences in quantity of methylation lost in 2i-medium culturing conditions. To further investigate how absence of LSH determines this increased loss of methylation, analyses of the expression of TET1/2 enzymes could be carried out. Some preliminary analysis shows that there is no significant difference in the expression of TET1 and TET2 between $Lsh^{off/+}$ and $Lsh^{off/off}$ mESCs (data not shown). To give a definite answer to whether LSH is linked to both *de novo* and maintenance of DNA methylation or exclusively to *de novo* methylation, $Lsh^{off/off}$ ESCs could be maintained in serum containing medium for longer passages and 5meC, 5hmeC and TET expression could be analysed at different times during the experiment and compared to $Lsh^{off/+}$ ESCs and ESCs depleted of *Dnmt1*. Direct comparison of the behavior of cells lacking the main protein responsible of *maintenance* of DNA methylation with cells lacking LSH could confirm or disproof the hypothesis of LSH being mainly

involved in *de novo* methylation.

HPLC analysis also showed that at the EBs stage, in absence of LSH 5meC is significantly lower compared to the control cell lines. This points toward two possible explanations: first, that cells need longer time to recover the 5meC in absence of LSH; second, that some loci cannot gain methylation in absence of LSH. Data derived from BioCAP-sequencing analysis were interrogated and a subset of genes hypomethylated in absence of LSH was identified. These genes lose methylation in 2i and do not gain the modification by the end of the experimental time window. A further step of completion of the differentiation process could be performed to determine whether the hypomethylation of these regions affects the differentiation process. Differentiating the EBs in different cell types might also provide information on the specificity of the target of LSH. In fact, Gene Ontology analysis showed that majority of the terms describing the genes differentially methylated in absence of LSH were related to neuronal system. These terms emerge from the analyses of the genes not methylated at day 2 in serum and day 8 EBs, hence indicating that it is not necessary any stimulation towards the neuronal lineage differentiation, such as retinoic acid addition, for these to be misregulated. This highlight a possible specific activity of LSH in this cell type and is consistent with the defects that the *Lsh*$^{off/off}$ mouse produced in the Stancheva lab shows (data not shown). However, this should be further investigated. ChIP-qPCR for LSH on a subset of candidate genes could be perform to determine the localization of the remodeling complex formed by LSH at those loci and therefore allowing to explain the depletion of DNA methylation in its absence.

As previously shown, LSH is important for methylation of DNA at repetitive elements, specifically LINE and IAPs. The culturing system used in this study could elucidate the time frame in which these repetitive sequences lose methylation in absence of LSH. As it happens for the unique loci, it emerges from BioCAP-sequencing analysis that the repetitive elements are differentially methylated at the initiation of *de novo* methylation and stay so until the EBs stage. It will be necessary to evaluate the expression of these elements and monitor this expression at subsequent differentiation

steps, by RNA-sequencing or simple qRT-PCR. Information about the expression status of these elements is necessary to fully understand the consequences of absence of LSH during embryonic development. In fact, it was shown in human and mouse how expression of repetitive elements affect the status not only of the single cell, but of a whole organism (Lane et al., 2003; Qin et al., 2010). RNA-sequencing could also provide information on the expression of unique loci, hence determining a clearer understanding of how lack of methylation is linked to gene expression in absence of LSH.

Altogether, these data suggest that some regions of the genome, both unique and repetitive, require LSH to be methylated at E3.5. Therefore, it can be hypothesised that LSH is driven to these loci and assure deposition of methylation by remodeling the heterochromatin and allowing access to DNMTs. These loci are therefore silenced. When LSH is absent, these genomic regions cannot be remodeler and DNMTs do not have access. Therefore, these regions cannot be silenced.

Importantly, it is yet unclear how LSH is driven to these repetitive or unique sequence. It will be therefore important to address this question, in order to fully understand the role of LSH in mammalian cells. Analysis of 5meC after depletion of H3K9me2 marked heterochromatin shows a possible additive effect on absence of LSH (Figure 4.5). It could be hypothesised that deposition of H3K9me2 by G9a could recruit LSH-CDCA7 to some loci, which could therefore be methylated and permanently silenced. To test this hypothesis, ChIP-qPCR analysis of H3K9me2 could be performed at LSH targets genes in *Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ cells, in presence or absence of G9a. If these genes showed enrichment in the modification in *Lsh*$^{off/off}$ and *Lsh*$^{off/+}$ cells and were not methylated in *Lsh*$^{off/+}$ cells in absence of G9a, it would show that H3K9me2 is required for recruiting LSH and methylating those specific regions.

In conclusion, this work has demonstrated that LSH is required for regulating DNA methylation *in vitro* concomitantly with the activation of the *de novo* methylation at IAPs and LINE-1 RE and at a subset of unique sequences, mostly related to features of the nervous system. Despite this being just a first step into a better understanding of how LSH

plays a role in embryonic development, these results could be launching pad for future investigations, providing a limited time window in the development to investigate and eventually determining a better understanding of the reprogramming processes of DNA methylation in the embryo.

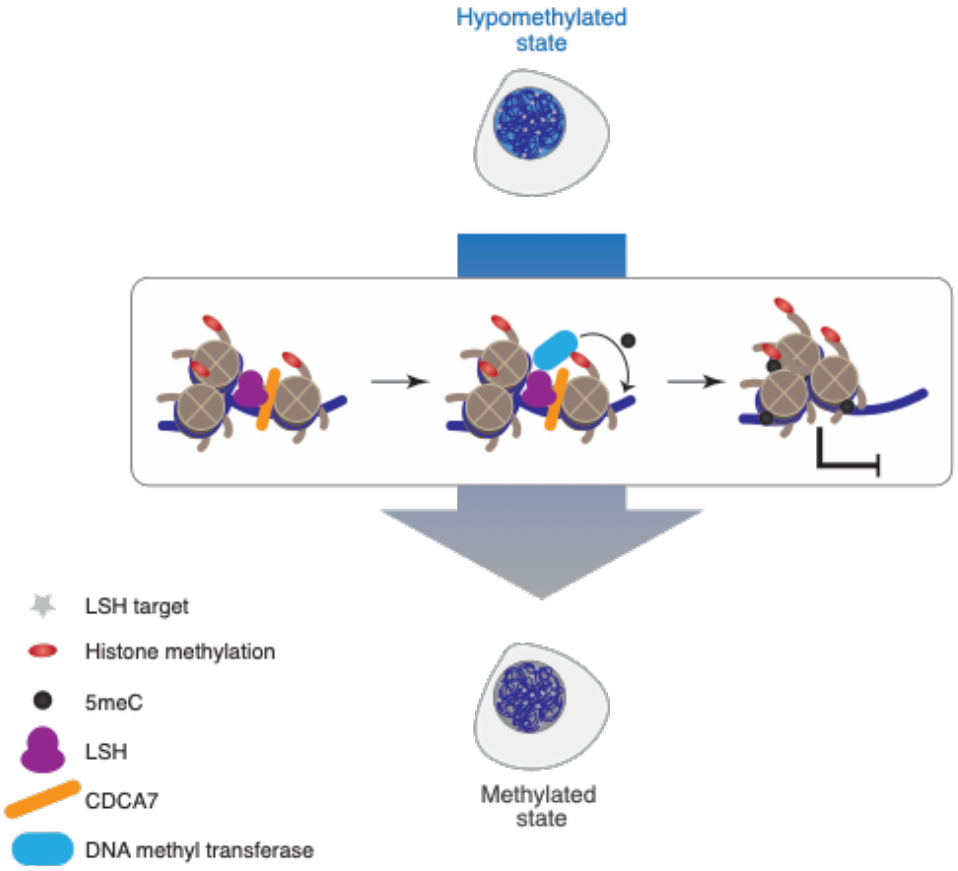


Figure 6.1 LSH-mediated DNA methylation.
Schematic representation of LSH role in DNA methylation. During the transition from hypomethylated to methylated status (in vitro by culturing medium change and in vivo by methylation reprogramming between E3.5 and E6.5 of embryonic development), a subset of loci in the genome requires LSH to be efficiently DNA methylated. Thanks to recent findings on LSH remodelling complex (Jennes at al., 2018), it can be hypothesised that CDCA7 associates to the chromatin at these repetitive or unique loci and recruits LSH. Chromatin is then remodelled, possibly by nucleosome sliding, allowing LSH to recruit DNA methyl transferase. The DNMT can then deposit a methyl group and therefore permanently transcriptionally silenced. In LSH-depleted cells, DNMTs do not gain access to the chromatin and these loci could eventually be reactivated and transcribed.

7. APPENDIX

BioCAP-seq quality control

Sample Name	% Dups	% GC	M Seqs
HT_D0_1.1	31.2%	62%	33.1
HT_D0_1.2	31.4%	62%	33.1
HT_D0_2.1	40.7%	64%	49.0
HT_D0_2.2	42.5%	64%	49.0
HT_D0_input.1	15.1%	41%	53.0
HT_D0_input.2	14.2%	41%	53.0
HT_Serum day 4_1.1	30.5%	60%	45.7
HT_Serum day 4_1.2	30.9%	60%	45.7
HT_Serum day 4_2.1	34.1%	60%	71.9
HT_Serum day 4_2.2	36.7%	60%	71.9
HT_Serum day 4_input.1	14.6%	41%	43.9
HT_Serum day 4_input.2	14.6%	41%	43.9
HT_EBs day 8_1.1	39.8%	65%	56.8
HT_EBs day 8_1.2	43.6%	65%	56.8
HT_EBs day 8_2.1	36.4%	63%	59.5
HT_EBs day 8_2.2	40.2%	63%	59.5
HT_EBs day 8_input.1	14.0%	41%	64.0
HT_EBs day 8_input.2	14.5%	41%	64.0
KO_D0_1.1	41.3%	63%	69.3
KO_D0_1.2	44.4%	63%	69.3
KO_D0_2.1	40.5%	62%	66.0
KO_D0_2.2	42.0%	62%	66.0
KO_D0_input.1	14.0%	41%	82.0
KO_D0_input.2	15.2%	41%	82.0
KO_Serum day 4_1.1	30.6%	56%	46.4
KO_Serum day 4_1.2	31.0%	57%	46.4
KO_Serum day 4_2.1	30.2%	51%	81.3
KO_Serum day 4_2.2	31.9%	51%	81.3
KO_Serum day 4_input.1	14.2%	41%	71.1

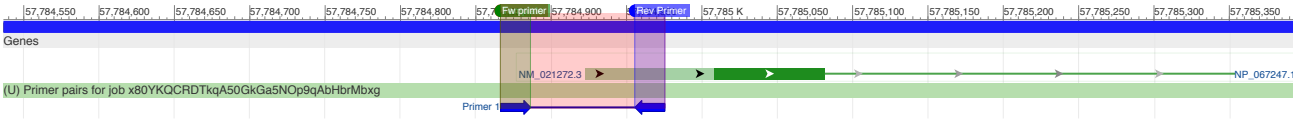
KO_Serum day 4_input.2	15.3%	41%	71.1
KO_EBs day 8_1.1	43.4%	60%	75.0
KO_EBs day 8_1.2	45.5%	60%	75.0
KO_EBs day 8_2.1	27.0%	54%	71.4
KO_EBs day 8_2.2	27.5%	54%	71.4
KO_EBs day 8_input.1	14.6%	41%	73.6
KO_EBs day 8_input.2	15.2%	41%	73.6

Appendix Table 1.BioCAP-sequencing quality control.
In the first column, samples sequenced are listed: HT = *Lsh*$^{off/+}$ KO = *Lsh*$^{off/off}$.
In the second column, the percentage of duplicates is indicated.
In the third column, the percentage of GC in each sample is indicated.
In the fourth column, the milion of reads sequenced per each sample is indicated.

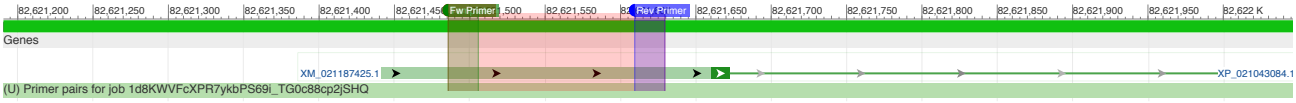
Appendix 1.BioCAP-sequencing quality control.
The mean quality value across each base position in the read, obtained with the FastQC quality control tool. All samples passed the quality control, reaching Phred values higher than 30.

BioCap-PCR PRIMER ALIGNMENT MAPS

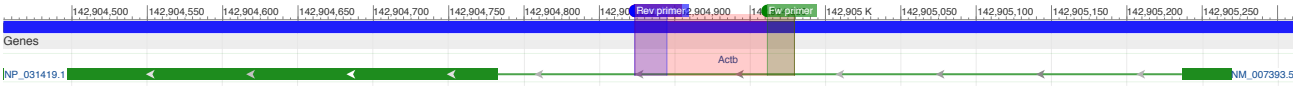
FABP7



MSN

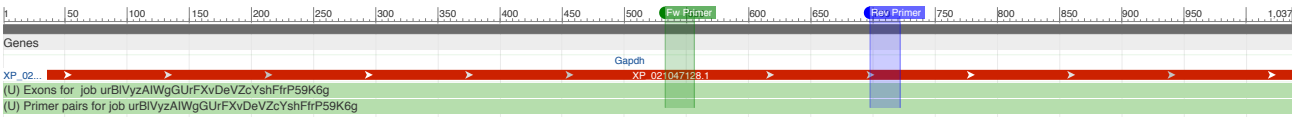


ACTIN B

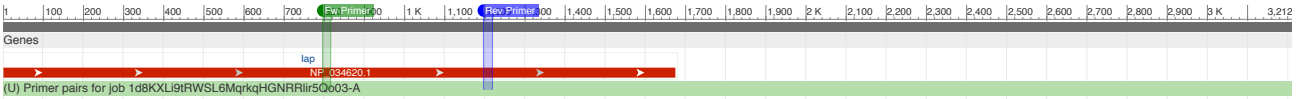


QRT PCR PRIMER ALIGNMENT MAPS

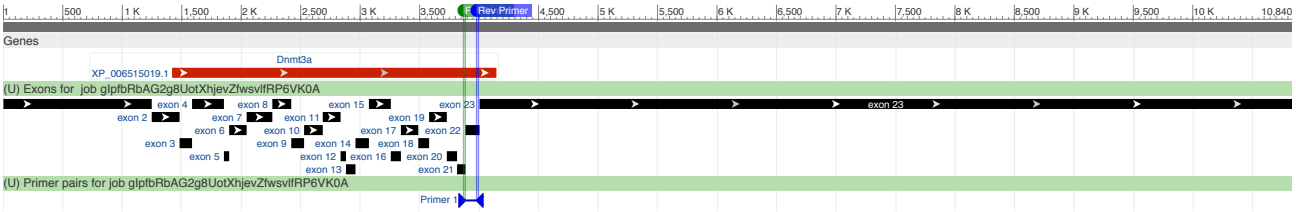
GAPDH



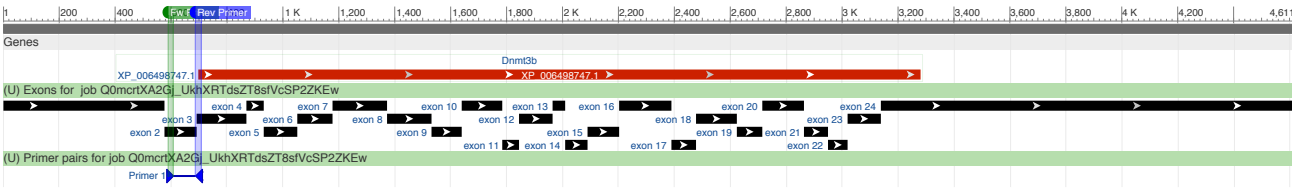
IAP



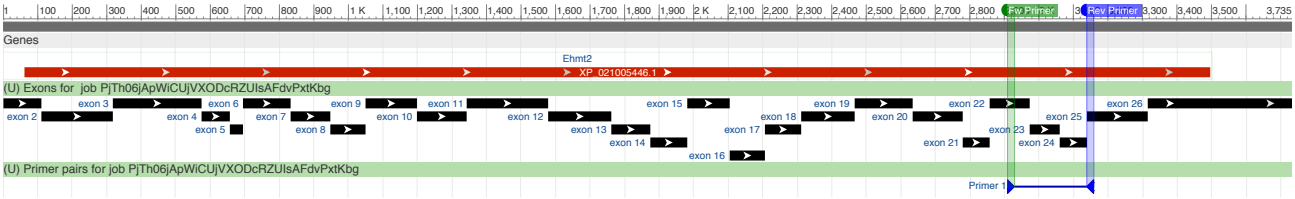
DNMT3A



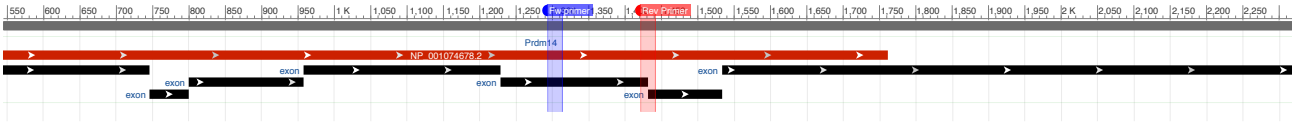
DNMT3B



G9A


Fw Primer
Rev Primer
Genes
Ehmt2
XP_021005446.1
(U) Exons for job PjTh06jApWiCUjVXODcRZUIsAFdvPxtKbg
(U) Primer pairs for job PjTh06jApWiCUjVXODcRZUIsAFdvPxtKbg
Primer 1

PRDM14


Fw primer
Rev Primer
Prdm14
NP_001074678.2

8. REFERENCES

Aagaard, L. (1999). Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31. *The EMBO Journal*, *18*(7), 1923–1938. https://doi.org/10.1093/emboj/18.7.1923

Aliaga, B., Bulla, I., Mouahid, G., Duval, D., & Grunau, C. (2019). Universality of the DNA methylation codes in Eucaryotes. *Science (New York, N.Y.)*, (October 2018), 1–11. https://doi.org/10.1038/s41598-018-37407-8

Allfrey, G., Faulkner, R., & Mirsky, A. E. (1964). ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. *Pnas*, *315*(1938), 786–794.

Alvaro, D., Lisby, M., & Rothstein, R. (2007). Genome-wide analysis of Rad52 foci reveals diverse mechanisms impacting recombination. *PLoS Genetics*, *3*(12), 2439–2449. https://doi.org/10.1371/journal.pgen.0030228

Ambrosi, C., Manzo, M., & Baubec, T. (2017). Dynamics and Context-Dependent Roles of DNA Methylation. *Journal of Molecular Biology*, *429*(10), 1459–1475. https://doi.org/10.1016/j.jmb.2017.02.008

Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P. W. S., D'Souza, Z., … Hajkova, P. (2016). De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nature Cell Biology*, *18*(November 2015). https://doi.org/10.1038/ncb3296

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, *21*(3), 381–395. https://doi.org/10.1038/cr.2011.22

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., … Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science*, *354*(6314), 909–912. https://doi.org/10.1126/science.aah5143

Barlow, D. P., & Bartolomei, M. S. (n.d.). Genomic Imprinting in Mammals.

Baubec, T., Colombo, D. F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A. R., … Schübeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature*. https://doi.org/10.1038/nature14176

Bednar, J., Horowitz, R. A., Grigoryev, S. A., Carruthers, L. M., Hansen, J. C., Koster, A. J.,

& Woodcock, C. L. (1998). Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proceedings of the National Academy of Sciences*, *95*(24), 14173–14178. https://doi.org/10.1073/pnas.95.24.14173

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., … Lander, E. S. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell*, *125*(2), 315–326. https://doi.org/10.1016/j.cell.2006.02.041

Bhutani, N., Burns, D. M., & Blau, H. M. (2011). DNA demethylation dynamics. *Cell*, *146*(6), 866–872. https://doi.org/10.1016/j.cell.2011.08.042

Bian, C., Chen, Q., & Yu, X. (2015). The zinc finger proteins ZNF644 and WIZ regulate the G9a/GLP complex for gene repression. *ELife*, *4*, 1–17. https://doi.org/10.7554/eLife.05606

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes & Development*, *16*(1), 6–21. https://doi.org/10.1101/gad.947102

Bird, A. (2007). Perceptions of epigenetics. *Nature*, *447*(7143), 396–398. https://doi.org/10.1038/nature05913

Blackledge, N. P., Long, H. K., Zhou, J. C., Kriaucionis, S., Patient, R., & Klose, R. J. (2012). Bio-CAP: A versatile and highly sensitive technique to purify and characterise regions of non-methylated DNA. *Nucleic Acids Research*, *40*(4). https://doi.org/10.1093/nar/gkr1207

Blossey, R., & Schiessel, H. (2018). The Latest Twists in Chromatin Remodeling. *Biophysical Journal*, 1–7. https://doi.org/10.1016/j.bpj.2017.12.008

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., … Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. *Nature Genetics*, *42*(12), 1093–1100. https://doi.org/10.1038/ng.708

Briones, V., & Muegge, K. (2012). The ghosts in the machine: DNA methylation and the mystery of differentiation. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1819*(7), 757–762. https://doi.org/10.1016/j.bbagrm.2012.02.013

Brzeski, J., & Jerzmanowski, A. (2003). Deficient in DNA methylation 1 (DDM1) defines a novel family of chromatin-remodeling factors. *Journal of Biological Chemistry*, *278*(2), 823–828. https://doi.org/10.1074/jbc.M209260200

Burrage, J., Termanis, A., Geissner, A., Myant, K., Gordon, K., & Stancheva, I. (2012). The SNF2 family ATPase LSH promotes phosphorylation of H2AX and efficient repair of

DNA double-strand breaks in mammalian cells. *Journal of Cell Science*, *125*(Pt 22), 5524–34. https://doi.org/10.1242/jcs.111252

Capuano, F., Kok, R., Blom, H. J., & Ralser, M. (2014). Cytosine DNA Methylation Is Found in Drosophila melanogaster but Absent in Saccharomyces cerevisiae, Schizosaccharomyces pombe , and Other Yeast Species. *Analytical Chemistry*, (86), 3697–3702.

Chen, T., Ueda, Y., Dodge, J. E., Wang, Z., & Li, E. (2003). Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Dnmt3b. *Molecular and Cellular Biology*, *23*(16), 5594–5605. https://doi.org/10.1128/MCB.23.16.5594

Clapier, C. R., & Cairns, B. R. (2009). The Biology of Chromatin Remodeling Complexes. *Annual Review of Biochemistry*, *78*(1), 273–304. https://doi.org/10.1146/annurev.biochem.77.062706.153223

Colot, V., & Rossignol, J. (1999). Eukaryotic DNA methylation as an evolutionary device. *Bioessays*, (21), 402–411.

Compere, S. J., & Palmiter, R. D. (1981). DNA methylation controls the inducibility of the mouse metallothionein-I gene in lymphoid cells. *Cell*, *25*(1), 233–240. https://doi.org/10.1016/0092-8674(81)90248-8

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., … Boone, C. (2010). SUPP - The genetic landscape of a cell. *Science*, *327*(5964), 425–31. https://doi.org/10.1126/science.1180823

De La Fuente, R., Baumann, C., Fan, T., Schmidtmann, A., Dobrinski, I., & Muegge, K. (2006). Lsh is required for meiotic chromosome synapsis and retrotransposon silencing in female germ cells. *Nature Cell Biology*, *8*(12), 1448–1454. https://doi.org/10.1038/ncb1513

De Sario, A. (2009). Clinical and molecular overview of inherited disorders resulting from epigenomic dysregulation. *European Journal of Medical Genetics*, *52*(6), 363–372. https://doi.org/10.1016/j.ejmg.2009.07.004

Dennis, K., Fan, T., Geiman, T., Yan, Q., & Muegge, K. (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes and Development*, *15*(22), 2940–2944. https://doi.org/10.1101/gad.929101

Dong, K. B., Maksakova, I. A., Mohn, F., Leung, D., Appanah, R., Lee, S., … Lorincz, M. C. (2008). DNA methylation in ES cells requires the lysine methyltransferase G9a but not

its catalytic activity, (April), 2691–2701. https://doi.org/10.1038/emboj.2008.193

Donohoe, M. E., Silva, S. S., Pinter, S. F., Xu, N., & Lee, J. T. (2009). The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature*, *460*(7251), 128–132. https://doi.org/10.1038/nature08098

Du, J., Johnson, L. M., Jacobsen, S. E., & Patel, D. J. (2015). DNA methylation pathways and their crosstalk with histone methylation. *Nature Reviews Molecular Cell Biology*, *16*(9), 519–532. https://doi.org/10.1038/nrm4043

Dunican, D. S., Cruickshanks, H. a, Suzuki, M., Semple, C. a, Davey, T., Arceci, R. J., … Meehan, R. R. (2013). Lsh regulates LTR retrotransposon repression independently of Dnmt3b function. *Genome Biology*, *14*(12), R146. https://doi.org/10.1186/gb-2013-14-12-r146

Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., … Bergman, Y. (2008). De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nature Structural and Molecular Biology*, *15*(11), 1176–1183. https://doi.org/10.1038/nsmb.1476

Erdel, F., Müller-Ott, K., Baum, M., Wachsmuth, M., & Rippe, K. (2011). Dissecting chromatin interactions in living cells from protein mobility maps. *Chromosome Research*, *19*(1), 99–115. https://doi.org/10.1007/s10577-010-9155-6

Erdel, F., & Rippe, K. (2011). Binding kinetics of human ISWI chromatin-remodelers to DNA repair sites elucidate their target location mechanism. *Nucleus*, *2*(2), 105–112. https://doi.org/10.4161/nucl.2.2.15209

Espada, J., Ballestar, E., Fraga, M. F., Villar-Garea, A., Juarranz, A., Stockert, J. C., … Esteller, M. (2004). Human DNA methyltransferase 1 is required for maintenance of the histone H3 modification pattern. *Journal of Biological Chemistry*, *279*(35), 37175–37184. https://doi.org/10.1074/jbc.M404842200

Fan, T. (2005). Lsh controls silencing of the imprinted Cdkn1c gene. *Development*, *132*(4), 635–644. https://doi.org/10.1242/dev.01612

Fan, T., Yan, Q., Huang, J., Austin, S., Cho, E., Ferris, D., & Muegge, K. (2003). Lsh-deficient murine embryonal fibroblasts show reduced proliferation with signs of abnormal mitosis. *Cancer Research*, *63*(15), 4677–4683.

Felle, M., Hoffmeister, H., Rothammer, J., Fuchs, A., Exler, J. H., & Längst, G. (2011). Nucleosomes protect DNA from DNA methylation in vivo and in vitro. *Nucleic Acids Research*, *39*(16), 6956–6969. https://doi.org/10.1093/nar/gkr263

Ferguson-Smith, A. C. (2011). Genomic imprinting: the emergence of an epigenetic

paradigm. *Nature Reviews Genetics*, *12*, 565. Retrieved from https://doi.org/10.1038/nrg3032

Ferry, L., Fournier, A., Tsusaka, T., Adelmant, G., Shimazu, T., Matano, S., … Defossez, P. A. (2017). Methylation of DNA Ligase 1 by G9a/GLP Recruits UHRF1 to Replicating DNA and Regulates DNA Methylation. *Molecular Cell*, *67*(4), 550–565.e5. https://doi.org/10.1016/j.molcel.2017.07.012

Ficz, G., Branco, M. R., Seisenberger, S., Santos, F., Krueger, F., Hore, T. a, … Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature*, *473*(7347), 398–402. https://doi.org/10.1038/nature10008

Ficz, G., Hore, T. a, Santos, F., Lee, H. J., Dean, W., Arand, J., … Reik, W. (2013). FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell*, *13*(3), 351–9. https://doi.org/10.1016/j.stem.2013.06.004

Fujita, N., Watanabe, S., Ichimura, T., Tsuruzoe, S., Shinkai, Y., Tachibana, M., … Nakao, M. (2003). Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression. *Journal of Biological Chemistry*, *278*(26), 24132–24138. https://doi.org/10.1074/jbc.M302283200

Geiman, T. M., Tessarollo, L., Anver, M. R., Kopp, J. B., Ward, J. M., & Muegge, K. (2001a). Lsh, a SNF2 family member, is required for normal murine development. *Biochimica et Biophysica Acta - General Subjects*, *1526*(2), 211–220. https://doi.org/10.1016/S0304-4165(01)00129-5

Geiman, T. M., Tessarollo, L., Anver, M. R., Kopp, J. B., Ward, J. M., & Muegge, K. (2001b). Lsh, a SNF2 family member, is required for normal murine development. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1526*(2), 211–220. https://doi.org/10.1016/S0304-4165(01)00129-5

Goll, M. G., & Bestor, T. H. (2005). Eukaryotic Cytosine Methyltransferases. *Annual Review of Biochemistry*, *74*(1), 481–514. https://doi.org/10.1146/annurev.biochem.74.010904.153721

Greer, E. L., Blanco, M. A., Gu, L., Sendinc, E., Liu, J., Aristizábal-corrales, D., … He, C. (2015). DNA methylation on N6-adenine in C. elegans. *Cell*, *161*(4), 868–878. https://doi.org/10.1016/j.cell.2015.04.005.DNA

Han, Y., Ren, J., Lee, E., Xu, X., Yu, W., & Muegge, K. (2017). Lsh/HELLS regulates self-renewal/ proliferation of neural stem/ progenitor cells. *Scientific Reports*, *7*(1), 1136. https://doi.org/10.1038/s41598-017-00804-6

Harikumar, A., & Meshorer, E. (2015). Chromatin remodeling and bivalent histone modifications in embryonic stem cells. *EMBO Reports*, *16*(12), 1609–1619. https://doi.org/10.15252/embr.201541011

Harr, J. C., Luperchio, T. R., Wong, X., Cohen, E., Wheelan, S. J., & Reddy, K. L. (2015). Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins. *Journal of Cell Biology*, *208*(1), 33–52. https://doi.org/10.1083/jcb.201405110

He, X.-J., Chen, T., & Zhu, J.-K. (2011). Regulation and function of DNA methylation in plants and animals. *Cell Research*, *21*(3), 442–465. https://doi.org/10.1038/cr.2011.23

He, X., Yan, B., Liu, S. S. S., Jia, J., Lai, W., Xin, X., … Tao, Y. (2016). Chromatin remodeling factor LSH drives cancer progression by suppressing the activity of fumarate hydratase. *Cancer Research*, *76*(19), 5743–5755. https://doi.org/10.1158/0008-5472.CAN-16-0268

Holliday, R. (1994). Epigenetics : An Overview. *Developmental Genetics*, *15*, 453–457.

Holliday, R., & Pugh, J. E. (1975). DNA modification mechanisms and gene activity during development. *Science (New York, N.Y.)*, *187*(4173), 226–232. https://doi.org/10.1126/science.1111098

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

Howlett, S. K., & Reik, W. (1991). Methylation levels of maternal and paternal genomes during preimplantation development. *Development*, *113*(1), 119–127.

Huang, J., Fan, T., Yan, Q., Zhu, H., Fox, S., Issaq, H. J., … Muegge, K. (2004). Lsh, an epigenetic guardian of repetitive elements. *Nucleic Acids Research*, *32*(17), 5019–5028. https://doi.org/10.1093/nar/gkh821

Hulten, M. (1978). Selective Somatic Pairing and Fragility at 1q12 in a Boy with Common Variable Immuno Deficiency. *Clinical Genetics*, *14*(5). https://doi.org/doi:10.1111/j.1399-0004.1978.tb02170.x

Illingworth, R., Kerr, A., DeSousa, D., Jørgensen, H., Ellis, P., Stalker, J., … Bird, A. (2008). A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biology*, *6*(1), 0037–0051. https://doi.org/10.1371/journal.pbio.0060022

Iqbal, K., Jin, S.-G., Pfeifer, G. P., & Szabo, P. E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proceedings of the National Academy of Sciences*, *108*(9), 3642–3647.

https://doi.org/10.1073/pnas.1014033108

Jaenisch, R., & Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics*, *33 Suppl*(march), 245–54. https://doi.org/10.1038/ng1089

Jarvis, C. D., Geiman, T., Vila-Storm, M. P., Osipovich, O., Akella, U., Candeias, S., … Muegge, K. (1996). A novel putative helicase produced in early murine lymphocytes. *Gene*, *169*(2), 203–207. https://doi.org/10.1016/0378-1119(95)00843-8

Jeddeloh, J. A., Stokes, T. L., & Richards, E. J. (1999). Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nature Genetics*, *22*(1), 94–97. https://doi.org/10.1038/8803

Jenness, C., Giunta, S., Müller, M. M., Kimura, H., Muir, T. W., & Funabiki, H. (2018). HELLS and CDCA7 comprise a bipartite nucleosome remodeling complex defective in ICF syndrome. *Proceedings of the National Academy of Sciences*, 201717509. https://doi.org/10.1073/pnas.1717509115

Jiang, Y., Mao, C., Yang, R., Yan, B., Shi, Y., Liu, X., … Tao, Y. (2017). EGLN1/c-Myc induced lymphoid-specific helicase inhibits ferroptosis through lipid metabolic gene expression changes. *Theranostics*, *7*(13), 3293–3305. https://doi.org/10.7150/thno.19988

Jones, P. A. (2012). Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, *13*(7), 484–492. https://doi.org/10.1038/nrg3230

Joshi, A. A., Chaudhari, A. J., Li, C., Dutta, J., Cherry, S. R., Shattuck, D. W., … Leahy, R. M. (2011). DNA hypomethylation caused by Lsh deletion promotes erythroleukemia development, *55*(20), 6197–6214. https://doi.org/10.1088/0031-9155/55/20/011.DigiWarp

Jurkowska, R. Z., Jurkowski, T. P., & Jeltsch, A. (2011). Structure and Function of Mammalian DNA Methyltransferases. *ChemBioChem*, *12*(2), 206–222. https://doi.org/10.1002/cbic.201000195

Kakutani, T., Jeddeloh, J. A., & Richards, E. J. (1995). Characterization of an Arabidopsis thaliana DNA hypomethylation mutant. *Nucleic Acids Research*, *23*(1), 130–137. https://doi.org/10.1093/nar/23.1.130

Keshet, I., Lieman-Hurwitz, J., & Cedar, H. (1986). DNA methylation affects the formation of active chromatin. *Cell*, *44*(4), 535–543. https://doi.org/10.1016/0092-8674(86)90263-1

Keyes, W. M., Pecoraro, M., Aranda, V., Vernersson-Lindahl, E., Li, W., Vogel, H., … Mills,

A. a. (2011). ΔNp63α is an oncogene that targets chromatin remodeler Lsh to drive skin stem cell proliferation and tumorigenesis. *Cell Stem Cell*, *8*(2), 164–76. https://doi.org/10.1016/j.stem.2010.12.009

Kohli, R. M., & Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*, *502*(7472), 472–479. https://doi.org/10.1038/nature12750

Kornberg, R. (1974). Chromatin Structure : A Repeating Unit of Histones and DNA Chromatin structure is based on a repeating unit of eight. *Science*, *184*, 868–871.

Kouzarides, T. (2007). Chromatin Modifications and Their Function. *Cell*, *128*(4), 693–705. https://doi.org/10.1016/j.cell.2007.02.005

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., … International Human Genome Sequencing, C. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. https://doi.org/10.1038/35057062

Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., & Reik, W. (2003). Resistance of IAPs to Methylation Reprogramming May Provide a Mechanism for Epigenetic Inheritance in the Mouse. *Genesis*, *35*, 88–93. https://doi.org/10.1002/gene.10168

Längst, G., & Manelyte, L. (2015). Chromatin Remodelers: From Function to Dysfunction. *Genes*, *6*(2), 299–324. https://doi.org/10.3390/genes6020299

Law, J. A., & Jacobsen, S. E. (2011). Establising, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet.*, *11*(3), 204–220. https://doi.org/10.1038/nrg2719.Establishing

Lee, D. W., Zhang, K., Ning, Z. Q., Raabe, E. H., Tintner, S., Wieland, R., … Arceci, R. J. (2000). Proliferation-associated SNF2-like gene (PASG): A SNF2 family member altered in leukemia. *Cancer Research*, *60*(13), 3612–3622. https://doi.org/10.1002/(sici)1521-1878(199808)20:8<634::aid-bies6>3.0.co

Leitch, H. G., McEwen, K. R., Turp, A., Encheva, V., Carroll, T., Grabole, N., … Hajkova, P. (2013). Naive pluripotency is associated with global DNA hypomethylation. *Nature Structural & Molecular Biology*, *20*(3), 311–6. https://doi.org/10.1038/nsmb.2510

Leitch, H. G., Surani, M. A., & Hajkova, P. (2016). DNA (De)Methylation: The Passive Route to Naïvety? *Trends in Genetics*, *32*(10), 592–595. https://doi.org/10.1016/j.tig.2016.08.005

Li, E. (2002). Chromatin modification and epigenetic reprogramming in mammalian development. *Nature Reviews Genetics*, *3*(9), 662–673. https://doi.org/10.1038/nrg887

Li, E., Bestor, T. H., & Jaenisch, R. (1992). Targeted mutation of the DNA

methyltransferase gene results in embryonic lethality. *Cell*, *69*(6), 915–926. https://doi.org/10.1016/0092-8674(92)90611-F

Li, Z., Chen, L., Kabra, N., Wang, C., Fang, J., & Chen, J. (2009). Inhibition of SUV39H1 methyltransferase activity by DBC1. *Journal of Biological Chemistry*, *284*(16), 10361–10366. https://doi.org/10.1074/jbc.M900956200

Liang, G., Chan, M. F., Tomigahara, Y., Tsai, Y. C., Gonzales, F. a, Li, E., … Jones, P. a. (2002). Cooperativity between DNA Methyltransferases in the Maintenance Methylation of Repetitive Elements Cooperativity between DNA Methyltransferases in the Maintenance Methylation of Repetitive Elements. *Molecular and Cellular Biology*, *22*(2), 480. https://doi.org/10.1128/MCB.22.2.480

Liao, J., Karnik, R., Gu, H., Ziller, M. J., Clement, K., Tsankov, A. M., … Meissner, A. (2015). Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. *Nature Genetics*, *47*(5), 469–478. https://doi.org/10.1038/ng.3258

Lienert, F., Mohn, F., Tiwari, V. K., Baubec, T., Roloff, T. C., Gaidatzis, D., … Schübeler, D. (2011). Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells. *PLoS Genetics*, *7*(6). https://doi.org/10.1371/journal.pgen.1002090

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., … Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, *462*(7271), 315–322. https://doi.org/10.1038/nature08514

Liu, P., Dou, X., Peng, G., Han, J.-D. J., & Jing, N. (2015). Genome-wide analysis of histone acetylation dynamics during mouse embryonic stem cell neural differentiation. *Genomics Data*, *5*, 15–6. https://doi.org/10.1016/j.gdata.2015.05.001

Lopez, D., Hamaji, T., Kropat, J., Hoff, P. De, Morselli, M., Rubbi, L., … Pellegrini, M. (2015). Dynamic Changes in the Transcriptome and Methylome of Chlamydomonas reinhardtii throughout Its Life Cycle 1. *Plant Physiology*, *169*(December), 2730–2743. https://doi.org/10.1104/pp.15.00861

Lungu, C., Muegge, K., Jeltsch, A., & Jurkowska, R. Z. (2015). An ATPase-deficient variant of the SNF2 family member HELLS shows altered dynamics at pericentromeric heterochromatin. *Journal of Molecular Biology*, *427*(10), 1903–1915. https://doi.org/10.1016/j.jmb.2015.03.014

Maier, V. K., Chioda, M., & Becker, P. B. (2008). ATP-dependent chromatosome remodeling. *Biological Chemistry*, *389*(4), 345–352. https://doi.org/10.1515/BC.2008.040

Martin, C., & Zhang, Y. (2005). The diverse functions of histone lysine methylation. *Nature Reviews Molecular Cell Biology*, *6*(11), 838–849. https://doi.org/10.1038/nrm1761

Maunakea, A. K., Chepelev, I., Cui, K., & Zhao, K. (2013). Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Research*, *23*(11), 1256–1269. https://doi.org/10.1038/cr.2013.110

Meng, H., Meng, H., Cao, Y., Qin, J., Song, X., Zhang, Q., … Cao, L. (2015). DNA methylation, its mediators and genome integrity. *International Journal of Biological Sciences*, *11*(5), 604–617. https://doi.org/10.7150/ijbs.11218

Mohn, F., Weber, M., Rebhan, M., Roloff, T. C., Richter, J., Stadler, M. B., … Schübeler, D. (2008). Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors. *Molecular Cell*, *30*(6), 755–766. https://doi.org/10.1016/j.molcel.2008.05.007

Moore, L. D., Le, T., & Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology*, *38*(1), 23–38. https://doi.org/10.1038/npp.2012.112

Morrison, A. J., & Shen, X. (2009). Chromatin remodelling beyond transcription: The INO80 and SWR1 complexes. *Nature Reviews Molecular Cell Biology*, *10*(6), 373–384. https://doi.org/10.1038/nrm2693

Moshkin, Y. M., Chalkley, G. E., Kan, T. W., Reddy, B. A., Ozgur, Z., van Ijcken, W. F. J., … Verrijzer, C. P. (2012). Remodelers Organize Cellular Chromatin by Counteracting Intrinsic Histone-DNA Sequence Preferences in a Class-Specific Manner. *Molecular and Cellular Biology*, *32*(3), 675–688. https://doi.org/10.1128/MCB.06365-11

Myant, K., & Stancheva, I. (2008). LSH cooperates with DNA methyltransferases to repress transcription. *Molecular and Cellular Biology*, *28*(1), 215–26. https://doi.org/10.1128/MCB.01073-07

Myant, K., Termanis, A., Sundaram, A. Y. M., Boe, T., Li, C., Merusi, C., … Stancheva, I. (2011). LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. *Genome Research*, *21*(1), 83–94. https://doi.org/10.1101/gr.108498.110

Narlikar, G. J., Sundaramoorthy, R., & Owen-Hughes, T. (2013). Mechanisms and Functions of ATP-Dependent Chromatin-Remodeling Enzymes. *Cell*, *154*(3), 490–503. https://doi.org/10.1016/j.cell.2013.07.011

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., … Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. *Nature*, *543*(7643), 72–77. https://doi.org/10.1038/nature21373

Nguyen, T. H. M., Carreira, P. E., Sanchez-Luque, F. J., Schauer, S. N., Fagg, A. C., Richardson, S. R., … Faulkner, G. J. (2018). L1 Retrotransposon Heterogeneity in Ovarian Tumor Cell Evolution. *Cell Reports*, *23*(13), 3730–3740. https://doi.org/10.1016/j.celrep.2018.05.090

Niwa, H., Burdon, T., Chambers, I., & Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes & Development*, *12*(13), 2048–60. https://doi.org/10.1101/gad.12.13.2048

Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell*, *99*, 247–257. https://doi.org/https://doi.org/10.1016/S0092-8674(00)81656-6

Okashita, N., Kumaki, Y., Ebi, K., Nishi, M., Okamoto, Y., Nakayama, M., … Seki, Y. (2014). PRDM14 promotes active DNA demethylation through the Ten-eleven translocation (TET)-mediated base excision repair pathway in embryonic stem cells. *Development*, *141*(2), 269–280. https://doi.org/10.1242/dev.099622

Okashita, N., Sakashita, N., Ito, K., Mitsuya, A., Suwa, Y., & Seki, Y. (2015). PRDM14 maintains pluripotency of embryonic stem cells through TET-mediated active DNA demethylation. *Biochemical and Biophysical Research Communications*, *466*(1), 138–145. https://doi.org/10.1016/j.bbrc.2015.08.122

Qin, C., Wang, Z., Shang, J., Bekkari, K., Liu, R., Pacchione, S., … Storer, R. D. (2010). Intracisternal a particle genes: Distribution in the mouse genome, active subtypes, and potential roles as species-specific mediators of susceptibility to cancer. *Molecular Carcinogenesis*, *49*(1), 54–67. https://doi.org/10.1002/mc.20576

Raabe, E. H., Abdurrahman, L., Behbehani, G., & Arceci, R. J. (2001). An SNF2 factor involved in mammalian development and cellular proliferation. *Developmental Dynamics*, *221*(1), 92–105. https://doi.org/10.1002/dvdy.1128

Ramsahoye, B. . (2002). Measurement of genome wide DNA methylation by reversed-phase high-performance liquid chromatography. *Methods*, *27*(2), 156–161. https://doi.org/10.1016/S1046-2023(02)00069-5

Ramsahoye, B. H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, a P., & Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(10), 5237–5242. https://doi.org/10.1073/pnas.97.10.5237

Rasmussen, K. D., & Helin, K. (2016). Role of TET enzymes in DNA methylation, development, and cancer. *Genes and Development*, *30*(7), 733–750. https://doi.org/10.1101/gad.276568.115

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*, *447*(7143), 425–432. https://doi.org/10.1038/nature05918

Reik, W., Dean, W., & Walter, J. (2001). Epigenetic Reprogramming in Mammalian Development. *Science*, *293*(5532), 1089–1093. https://doi.org/10.1126/science.1063443

Ren, J., Briones, V., Barbour, S., Yu, W., Han, Y., Terashima, M., & Muegge, K. (2015). The ATP binding site of the chromatin remodeling homolog Lsh is required for nucleosome density and de novo DNA methylation at repeat sequences. *Nucleic Acids Research*, *43*(3), 1444–55. https://doi.org/10.1093/nar/gku1371

Ren, J., Hathaway, N. A., Crabtree, G. R., & Muegge, K. (2017). Tethering of Lsh at the Oct4 locus promotes gene repression associated with epigenetic changes. *Epigenetics*, *2294*(June), 00–00. https://doi.org/10.1080/15592294.2017.1338234

Rippe, K., Schrader, A., Riede, P., Strohner, R., Lehmann, E., & Langst, G. (2007). DNA sequence- and conformation-directed positioning of nucleosomes by chromatin-remodeling complexes. *Proceedings of the National Academy of Sciences*, *104*(40), 15635–15640. https://doi.org/10.1073/pnas.0702430104

Robertson, A. K., Geiman, T. M., Sankpal, U. T., Hager, G. L., & Robertson, K. D. (2004). Effects of chromatin structure on the enzymatic and DNA binding functions of DNA methyltransferases DNMT1 and Dnmt3a in vitro. *Biochemical and Biophysical Research Communications*, *322*(1), 110–118. https://doi.org/10.1016/j.bbrc.2004.07.083

Rodríguez-Paredes, M., & Esteller, M. (2011). Cancer epigenetics reaches mainstream oncology. *Nature Medicine*, *17*(3), 330–339. https://doi.org/10.1038/nm.2305

Rondelet, G., Dal Maso, T., Willems, L., & Wouters, J. (2016). Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. *Journal of Structural Biology*, *194*(3), 357–367. https://doi.org/10.1016/j.jsb.2016.03.013

Rose, N. R., & Klose, R. J. (2014). Understanding the relationship between DNA methylation and histone lysine methylation. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1839*(12), 1362–1372. https://doi.org/10.1016/j.bbagrm.2014.02.007

Samuelsson, J., Dumbovic, G., Polo, C., Moreta, C., Alibés, A., Ruiz-Larroya, T., … Perucho, M. (2016). Helicase Lymphoid-Specific Enzyme Contributes to the Maintenance of Methylation of SST1 Pericentromeric Repeats That Are Frequently Demethylated in Colon Cancer and Associate with Genomic Damage. *Epigenomes*,

1(1), 2. https://doi.org/10.3390/epigenomes1010002

Sarraf, S. A., & Stancheva, I. (2004). Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Molecular Cell*, *15*(4), 595–605. https://doi.org/10.1016/j.molcel.2004.06.043

Shaked, H., Avivi-Ragolsky, N., & Levy, A. A. (2006). Involvement of the arabidopsis SWI2/SNF2 chromatin remodeling gene family in DNA damage response and recombination. *Genetics*, *173*(2), 985–994. https://doi.org/10.1534/genetics.105.051664

Shinkai, Y., & Tachibana, M. (2011). H3K9 methyltransferase G9a and the related molecule GLP. *Genes and Development*, *25*(8), 781–788. https://doi.org/10.1101/gad.2027411

Silva, J., Nichols, J., Theunissen, T. W., Guo, G., van Oosten, A. L., Barrandon, O., … Smith, A. (2009). Nanog Is the Gateway to the Pluripotent Ground State. *Cell*, *138*(4), 722–737. https://doi.org/10.1016/j.cell.2009.07.039

Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M., & Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature*, *336*(6200), 688–690. https://doi.org/10.1038/336688a0

Smith, Z. D., Chan, M. M., Humm, K. C., Karnik, R., Mekhoubad, S., Regev, A., … Meissner, A. (2014). DNA methylation dynamics of the human preimplantation embryo. *Nature*, *511*(7511), 611–615. https://doi.org/10.1038/nature13581

Smith, Z. D., Chan, M. M., Mikkelsen, T. S., Gu, H., Gnirke, A., Regev, A., & Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature*, *484*(7394), 339–44. https://doi.org/10.1038/nature10960

Smith, Z. D., & Meissner, A. (2013a). DNA methylation: roles in mammalian development. *Nature Reviews. Genetics*, *14*(3), 204–20. https://doi.org/10.1038/nrg3354

Smith, Z. D., & Meissner, A. (2013b). DNA methylation: Roles in mammalian development. *Nature Reviews Genetics*, *14*(3), 204–220. https://doi.org/10.1038/nrg3354

Struhl, K., & Segal, E. (2013). Determinants of nucleosome positioning. *Nature Structural and Molecular Biology*, *20*(3), 267–273. https://doi.org/10.1038/nsmb.2506

Sun, L. Q., Lee, D. W., Zhang, Q., Xiao, W., Raabe, E. H., Meeker, A., … Arceci, R. J. (2004). Growth retardation and premature aging phenotypes in mice with disruption of the SNF2-like gene, PASG. *Genes and Development*, *18*(9), 1035–1046. https://doi.org/10.1101/gad.1176104

Suzuki, T., Farrar, J. E., Yegnasubramanian, S., Zahed, M., Suzuki, N., & Arceci, R. J. (2008). Stable knockdown of PASG enhances DNA demethylation but does not accelerate cellular senescence in TIG-7 human fibroblasts. *Epigenetics : Official Journal of the DNA Methylation Society*, *3*(5), 281–291. https://doi.org/10.4161/epi.3.5.6914

Tahiliani, M., Kian, P. K., Yinghua, S., A., P. W., Bandukwala, H., Brudno, Y., … Anjana, R. (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1, (May), 930–936.

Tan, A. Y., & Manley, J. L. (2009). The TET family of proteins: Functions and roles in disease. *Journal of Molecular Cell Biology*, *1*(2), 82–92. https://doi.org/10.1093/jmcb/mjp025

Tao, Y., Xi, S., Shan, J., Maunakea, A., Che, A., Briones, V., … Muegge, K. (2011). Lsh, chromatin remodeling family member, modulates genome-wide cytosine methylation patterns at nonrepeat sequences. *Proceedings of the National Academy of Sciences*, *108*(14), 5626–5631. https://doi.org/10.1073/pnas.1017000108

Termanis, A., Torrea, N., Culley, J., Kerr, A., Ramsahoye, B., & Stancheva, I. (2016). The SNF2 family ATPase LSH promotes cell-autonomous de novo DNA methylation in somatic cells. *Nucleic Acids Research*, *44*(16), 7592–7604. https://doi.org/10.1093/nar/gkw424

Thijssen, P. E., Ito, Y., Grillo, G., Wang, J., Velasco, G., Nitta, H., … Sasaki, H. (2015). Mutations in CDCA7 and HELLS cause immunodeficiency–centromeric instability–facial anomalies syndrome. *Nature Communications*, *6*, 7870. https://doi.org/10.1038/ncomms8870

Towbin, B. D., González-Aguilera, C., Sack, R., Gaidatzis, D., Kalck, V., Meister, P., … Gasser, S. M. (2012). Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell*, *150*(5), 934–947. https://doi.org/10.1016/j.cell.2012.06.051

Trojer, P., & Reinberg, D. (2007). Facultative Heterochromatin: Is There a Distinctive Molecular Signature? *Molecular Cell*, *28*(1), 1–13. https://doi.org/10.1016/j.molcel.2007.09.011

Tsumura, A., Hayakawa, T., Kumaki, Y., & Takebayashi, S. (2006). Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes to Cells*, *11*, 805–814. https://doi.org/10.1111/j.1365-2443.2006.00984.x

Vedadi, M., Barsyte-Lovejoy, D., Liu, F., Rival-Gervier, S., Allali-Hassani, A., Labrie, V., …

Jin, J. (2011). A chemical probe selectively inhibits G9a and GLP methyltransferase activity in cells. *Nature Chemical Biology*, *7*(8), 566–574. https://doi.org/10.1038/nchembio.599

Velasco, G., Grillo, G., Touleimat, N., Ferry, L., & Ivkovic, I. (2018). Comparative methylome analysis of ICF patients identifies heterochromatin loci that require ZBTB24 , CDCA7 and HELLS for their methylated state ., *0*(April), 1–16. https://doi.org/10.1093/hmg/ddy130/4969376

Voigt, P., Tee, W. W., & Reinberg, D. (2013). A double take on bivalent promoters. *Genes and Development*, *27*(12), 1318–1338. https://doi.org/10.1101/gad.219626.113

Von Eyss, B., Maaskola, J., Memczak, S., Möllmann, K., Schuetz, A., Loddenkemper, C., … Ziebold, U. (2012). The SNF2-like helicase HELLS mediates E2F3-dependent transcription and cellular transformation. *EMBO Journal*, *31*(4), 972–985. https://doi.org/10.1038/emboj.2011.451

von Meyenn, F., Iurlaro, M., Habibi, E., Liu, N. Q., Salehzadeh-Yazdi, A., Santos, F., … Stunnenberg, H. G. (2016). Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells. *Molecular Cell*, *62*(6), 848–861. https://doi.org/10.1016/j.molcel.2016.04.025

Vongs, a, Kakutani, T., Martienssen, R. a, & Richards, E. J. (1993). Arabidopsis thaliana DNA methylation mutants. *Science (New York, N.Y.)*, *260*(5116), 1926–1928. https://doi.org/10.1126/science.8316832

Waddington, C. H. (2012). The epigenotype. 1942. *International Journal of Epidemiology*, *41*(1), 10–13. https://doi.org/10.1093/ije/dyr184

Walter, M., Teissandier, A., Pérez-Palacios, R., & Bourc'His, D. (2016). An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *ELife*, *5*(JANUARY2016), 1–30. https://doi.org/10.7554/eLife.11418.001

Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., … Liu, J. (2014). Programming and inheritance of parental DNA methylomes in mammals. *Cell*, *157*(4), 979–991. https://doi.org/10.1016/j.cell.2014.04.017

Wen, B., Wu, H., Shinkai, Y., Irizarry, R. A., & Feinberg, A. P. (2009). Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nature Genetics*, *41*(2), 246–250. https://doi.org/10.1038/ng.297

Williams, R. L., Hilton, D. J., Pease, S., Willson, T. a, Stewart, C. L., Gearing, D. P., … Gough, N. M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental

potential of embryonic stem cells. *Nature*, *336*(6200), 684–687. https://doi.org/10.1038/336684a0

Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C. J., Zakhartchenko, V., Boiani, M., … Walter, J. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nature Communications*, *2*(1). https://doi.org/10.1038/ncomms1240

Xi, S., Zhu, H., Xu, H., Schmidtmann, A., Geiman, T. M., & Muegge, K. (2007). Lsh controls Hox gene silencing during development. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(36), 14366–71. https://doi.org/10.1073/pnas.0703669104

Xiao, D., Huang, J., Pan, Y., Li, H., Fu, C., Mao, C., … Tao, Y. (2017). Chromatin remodeling factor LSH is upregulated by the LRP6-GSK3β-E2F1 axis linking reversely with survival in gliomas. *Theranostics*, *7*(1), 132–143. https://doi.org/10.7150/thno.17032

Yan, Q., Huang, J., Fan, T., Zhu, H., & Muegge, K. (2003). Lsh, a modulator of CpG methylation, is crucial for normal histone methylation. *EMBO Journal*, *22*(19), 5154–5162. https://doi.org/10.1093/emboj/cdg493

Yearim, A., Gelfman, S., Shayevitch, R., Melcer, S., Glaich, O., Mallm, J. P., … Ast, G. (2015). HP1 Is Involved in Regulating the Global Impact of DNA Methylation on Alternative Splicing. *Cell Reports*, *10*(7), 1122–1134. https://doi.org/10.1016/j.celrep.2015.01.038

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., … Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature*, *453*(7194), 519–23. https://doi.org/10.1038/nature06968

Yu, W., McIntosh, C., Lister, R., Zhu, I., Han, Y., Ren, J., … Muegge, K. (2014a). Genome-wide DNA methylation patterns in LSH mutant reveals de-repression of repeat elements and redundant epigenetic silencing pathways. *Genome Research*, *24*(10), 1613–23. https://doi.org/10.1101/gr.172015.114

Yu, W., McIntosh, C., Lister, R., Zhu, I., Han, Y., Ren, J., … Muegge, K. (2014b). Genome-wide DNA methylation patterns in LSH mutant reveals de-repression of repeat elements and redundant epigenetic silencing pathways. *Genome Research*, *24*(10), 1613–23. https://doi.org/10.1101/gr.172015.114

Yun, M., Wu, J., Workman, J. L., & Li, B. (2011). Readers of histone modifications. *Cell Research*, *21*(4), 564–578. https://doi.org/10.1038/cr.2011.42

Zhang, G., Huang, H., Wang, H., Chen, D., Zhang, G., Huang, H., … Zhang, D. (2015). N 6-Methyladenine DNA Modification in Drosophila. *Cell*, *161*(4), 893–906. https://doi.org/10.1016/j.cell.2015.04.018

Zhang, T., Termanis, A., Özkan, B., Bao, X. X., Culley, J., de Lima Alves, F., … Stancheva, I. (2016). G9a/GLP Complex Maintains Imprinted DNA Methylation in Embryonic Stem Cells. *Cell Reports*, *15*(1), 77–85. https://doi.org/10.1016/j.celrep.2016.03.007

Zhang, W., & Xu, J. (2017). DNA methyltransferases and their roles in tumorigenesis. *Biomarker Research*, 1–8. https://doi.org/10.1186/s40364-017-0081-z

Zhu, H., Geiman, T. M., Xi, S., Jiang, Q., Schmidtmann, A., Chen, T., … Muegge, K. (2006). Lsh is involved in de novo methylation of DNA. *The EMBO Journal*, *25*(2), 335–345. https://doi.org/10.1038/sj.emboj.7600925

Zylicz, J. J., Dietmann, S., Günesdogan, U., Hackett, J. A., Cougot, D., Lee, C., & Surani, M. A. (2015). Chromatin dynamics and the role of G9a in gene regulation and enhancer silencing during early mouse development. *ELife*, *4*(NOVEMBER2015), 1–25. https://doi.org/10.7554/eLife.09571