THE ROLE OF SMYD2 DURING HUMAN EMBRYONIC STEM CELLS DIFFERENTIATION

Borja Sesé Ballesteros

TESI DOCTORAL UPF / 2013

Thesis supervisor:

Dra. María José Barrero Núñez,
Center of Regenerative Medicine in Barcelona

UNIVERSITAT POMPEU FABRA

Por su apoyo incondicional y saber que siempre van a estar ahí,

a mi familia

ACKNOWLEDGMENTS

Me gustaría aprovechar este pequeño apartado para recordar y agradecer a todas esas personas que han sido fundamentales para la realización de la presente tesis doctoral.

En primer lugar, a María José y Juan Carlos por haberme dado la oportunidad de formar parte del CMR[B] para realizar mis estudios de doctorado. A María José no le puedo estar más agradecido por la confianza puesta en mi desde el primer día. Ella ha sido una excelente profesora y le debo todo cuanto he aprendido a lo largo de estos años de trabajo.

A Antonio por haber sido un gran compañero además de un gran amigo en todo momento. Con Antonio compartí la mesa de laboratorio durante los dos primeros años, hasta que se marchó y me quedé con la mesa entera para mí. La gente suele alegrarse cuando le dan una mesa de laboratorio entera para él solo. Yo no.

A Julio por haber estado ahí siempre que lo he necesitado. Cuando a uno le surgen dudas que desafían los conceptos de la Biología Molecular puramente académica, él es el hombre a quien acudir. Para mí ha sido un apoyo importante tanto dentro como fuera del laboratorio. Me siento tremendamente afortunado de haber formado parte del mismo grupo de trabajo junto a él, Antonio y María José.

A Carme y Veronika por todo lo aprendido acerca de la experimentación con embriones de pez cebra en esta última etapa. A Carme por su total entrega y dedicación al proyecto e irradiarnos cada día, sin excepción, con su simpatía y buen humor. To Veronika for her enthusiasm and optimism about the project and offering me all the necessary assistance every time I needed.

A toda la gente con la que he tenido el placer de trabajar durante estos cuatro años, puesto que todos y cada uno de ellos han aportado, de una forma u otra, su granito de arena en esta tesis. La lista es muy extensa pero hay un colectivo que merece una mención especial, los estudiantes de doctorado. A mis compañeros de fatiga, Ignasi, Edu, Cristina, Alex, Adriana, Erika y Lorena por haber compartido tan buenos momentos y hacer del doctorado un camino más llevadero.

To Dr. Elodie Drapeau and Dr. Joseph D. Buxbaum for hosting me as a visiting student in the Department of Psychiatry at Mount Sinai School of Medicine. Thanks for their kindness and generosity.

No menos importante, a Anna por aceptar ser mi tutora de tesis.

A todos vosotros, muchas gracias.

ABSTRACT

Embryonic stem (ES) cells are able to differentiate into any cell type, a property called pluripotency, and have unlimited potential for self-renewal. Although the molecular mechanisms responsible for maintaining self-renewal and pluripotency in ES cells are not well known, recent studies have demonstrated the importance of epigenetic mechanisms in maintaining these processes. Histone modifying enzymes play decisive roles in differentiation and development. This study describes that SMYD2 (SET and MYND domain containing protein 2), a histone lysine methyltransferase, is induced during human ES cells differentiation and it is preferentially expressed in somatic cells versus pluripotent cells. Gain and loss-of-function experiments have shown that knockdown of SMYD2 in human ES cells promotes the induction of endodermal markers during differentiation, while overexpression has opposite effects. *In vivo* experiments in zebrafish revealed that knockdown of *smyd2a* (a homologue gene of human SMYD2) causes developmental delays and aberrant tail formation. The phenotype of *smyd2a*-morphant embryos correlates with a low expression of *ntl* and over induction Nodal-related genes during gastrulation. Finally, SMYD2 is shown to stimulate the activation of BMP signaling pathway and promotes the induction of BMP2-target genes in human ES cells. Overall, these findings suggest that SMYD2 plays a critical role at early stages during development and in human ES cells differentiation.

RESUM

Les cèl·lules mare embrionàries (ES) són capaces de diferenciar-se a qualsevol tipus cel·lular, una propietat coneguda amb el nom de pluripotència, i presenten un potencial il·limitat d'auto-renovació. Tot i que els mecanismes moleculars responsables per al manteniment de la pluripotència i l'auto-renovació en ES encara no es coneixen bé, estudis recents han demostrat la importància dels mecanismes epigenètics en mantenir aquests processos. Els enzims modificadors d´histones juguen un paper decisiu durant la diferenciació i el desenvolupament. Aquest estudi descriu que SMYD2 (SET and MYND domain containing protein 2), una metiltransferasa d'histones, s'indueix durant la diferenciació de cèl·lules ES i s'expressa preferentment en cèl·lules somàtiques envers cèl·lules pluripotents. Experiments de guany i pèrdua de funció mostren que el noqueig de SMYD2 en cèl·lules ES humanes promou la inducció de marcadors d'endoderm durant la diferenciació, mentre que la sobre-expressió té efectes oposats. Experiments *in vivo* en el peix zebra van revelar que el noqueig de *smyd2a* (un gen homòleg de SMYD2 humà) causa retard en el desenvolupament i formació aberrant de la cua. El fenotip dels embrions absents de *smyd2a* es correlaciona amb una baixa expressió de *ntl* i una sobre-inducció dels gens relatius a Nodal durant la gastrulació. Finalment, SMYD2 estimula l'activació de la via de senyalització BMP i promou la inducció dels gens diana de BMP2 en cèl·lules ES humanes. En general, aquests descobriments suggereixen que SMYD2 juga un paper important durant els estadis primerencs en el desenvolupament embrionari i durant la diferenciació de cèl·lules ES humanes.

PREFACE

Over the last three decades, human embryonic stem (ES) cells have been proposed as a promising tool for many therapeutic applications focused on personalized and regenerative medicine. ES cells are able to differentiate into any cell type, a property called pluripotency, and have unlimited potential for self-renewal. For many years it was thought that the differentiation of ES cells into specialized cell types was an irreversible process. However, recently it has been described that it is possible to reprogram adult somatic cells into pluripotent cells by overexpressing a combination of a subset of transcription factors. These so called induced pluripotent stem (iPS) emerged as an alternative to human ES cells, given their ability to self-renew and to differentiate into a large cohort of cell types. Although the molecular mechanisms responsible for maintaining self-renewal and pluripotency in ES are not well-known, recent studies have demonstrated the importance of epigenetic mechanisms in maintaining these processes. The chromatin state in ES cells is generally less compact and more permissive to the transcriptional machinery, compared with differentiated cells. Upon differentiation, chromatin becomes organized into a more repressive state. Histone modifications play a key role in chromatin compaction and organization, and many of these modifications can be correlated with gene expression activation or repression. In ES cells, a range of key developmental genes present a very particular histone modification pattern on their regulatory regions consisting on the simultaneous presence of marks associated with gene expression (H3K4 methylation) and repression (H3K27 methylation). This so-called bivalent domain has been suggested to maintain developmental genes silenced but poised and ready to be activated upon differentiation to specific linages. Differentiation causes the resolution of these domains into H3K4 methylation only in genes that become active and H3K27 methylation only in genes that become permanently repressed. We anticipated that histone marks present at the bivalent domains are controlled by a complex interplay of histone modifying enzymes in ES cells. These enzymes would contribute to the establishment and maintenance of the bivalent domains in ES cells and to their resolution during differentiation. The main aim of our study is to identify critical histone lysine methyltransferases (HKMTs) and demethylases (HKDMs) that participate in the homeostasis of these domains in human

ES cells and understand the molecular mechanisms of action of these enzymes. Preliminary results identified the histone lysine methyltransferase SMYD2 as a gene that is highly induced during the differentiation of human ES cells and likely to play a role in this process. SMYD2 is a member of a protein family containing a SET domain that is split into two segments by a MYND domain, and has been described to have methyltransferase activity specific for histone H3K36 and H3K4, both marks frequently associated with active gene expression. In this study we used several biological approaches based on *in vitro* cell culture techniques and *in vivo* zebrafish model system as well, in order to address the role of SMYD2 during human ES cell differentiation.

CONTENTS

LIST OF FIGURES

1. INTRODUCTION

1.1. CHROMATIN STRUCTURE

Genomic DNA in eukaryotic organisms is highly compacted but well organized into the cell nucleus. The genetic material is packaged into a complex structure named chromatin, where DNA is assembled together with proteins. The compaction of DNA into a higher-order structure inhibits gene transcription by limiting the access of the transcriptional machinery to DNA [1]. However, chromatin presents a dynamic architecture which offers a precise control on gene expression by modulating its conformation for either facilitate or impair the accessibility of transcription factors to DNA [2]. Depending on the level of condensation, in non-dividing cells, chromatin presents two different conformation states. Recent findings have shown that genome can be classified into five different chromatin types, based on their unique combinations of proteins [3]. Transcriptionally active euchromatin can be classified in two different subgroups: one containing genes with a broad expression pattern, such as DNA repair, ribosome and nucleic acid metabolic process; and another one containing genes linked to a more specific processes like receptor binding, transcription factor activity and signal transduction. Inactive heterochromatin can also be subdivided in two subgroups: one corresponds to classic heterochromatin marked with SU(VAR)3-9, HP1 and HP1-interacting proteins; and another one corresponds to Polycomb binding chromatin. The fifth chromatin type, which covers about a half of the genome, corresponds to the prevalent type of repressive chromatin and lacks the classic heterochromatin markers. This prevalent repressive chromatin presents no signals for transcriptional activity, and harbour most of the silent genes in the genome.

Histones are the main protein component of chromatin and they are composed of a globular domain and flexible unstructured tails. The biological function of histones was first described to merely organize DNA within the nucleus, but in recent years mounting evidence suggest that histones also play an important role in regulating gene expression [4]. There are five primary histone isotypes designated as core histones H2A, H2B, H3, H4 and linker histone H1. In addition, with exception of H4, all different histone

isotypes present several histone replacement variants that are recruited to specific regions of the genome to carry out specific functions [5].

The nucleosome is the basic structural element of chromatin organisation [6]. The structure of the nucleosome consists of two copies of each core histone proteins, H2A, H2B, H3 and H4, with 145-147 base pairs of DNA wrapped 1.65 times around the histone octamer [7,8] (Fig. 1). During nucleosome assembly, a first formation of H3-H4 heterodimers give rise to a $(H3\text{-}H4)_2$ tetramer as a stable complex. At the same time, H2A heterodimerizes with H2B, but in this case two separate H2A-H2B dimers binds one at each side of the $(H3\text{-}H4)_2$ [9,10]. Besides the histone octamer fold core, each histone presents N-terminal tails that protrude from the nucleosome, exposed as ideal surfaces for post-translational covalent modifications [11]. Moreover, these modifications can modulate the chromatin structure and gene expression by affecting the accessibility of transcription factors to DNA [12,13].

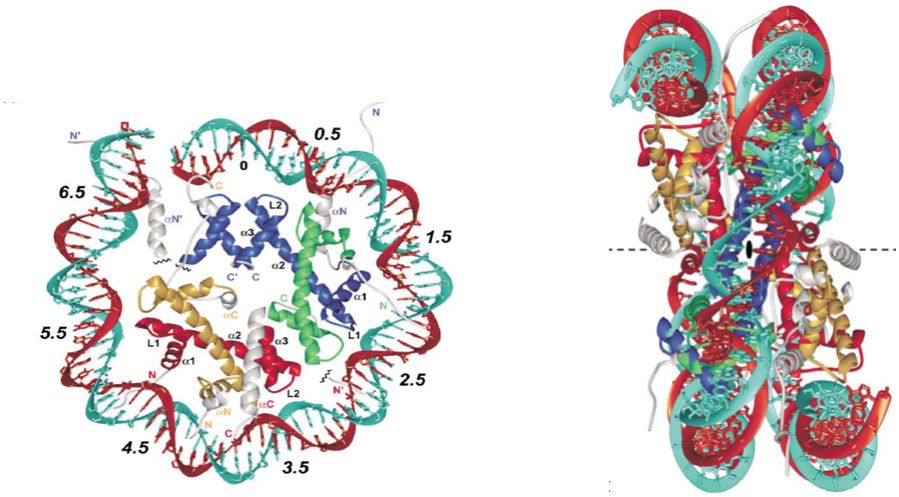


Figure 1. Structure of the nucleosome core particle at 1.9 Å resolution. It shows the double helix of DNA wrapped around the histone octamer. The histone-fold domains of the histone core proteins are blue for H3, green for H4, yellow for H2A and red for H2B. The histone-fold extensions and N-terminal tail regions shown are white [8].

Recently, a new model for chromatin packaging has been proposed, similar to the one used for protein folding [14] (Fig. 2). The primary structure consists of a 10-nm chromatin fiber organized in linear arrays of nucleosomes connected by short segment of DNA. The secondary structure is formed by the folded nucleosomal arrays into a compact fiber with a diameter of ~30nm, which is stabilized by the binding of linker

histone H1. The tertiary structure is associated with interactions between secondary structures and the establishment of long-distance chromatin fiber interactions.

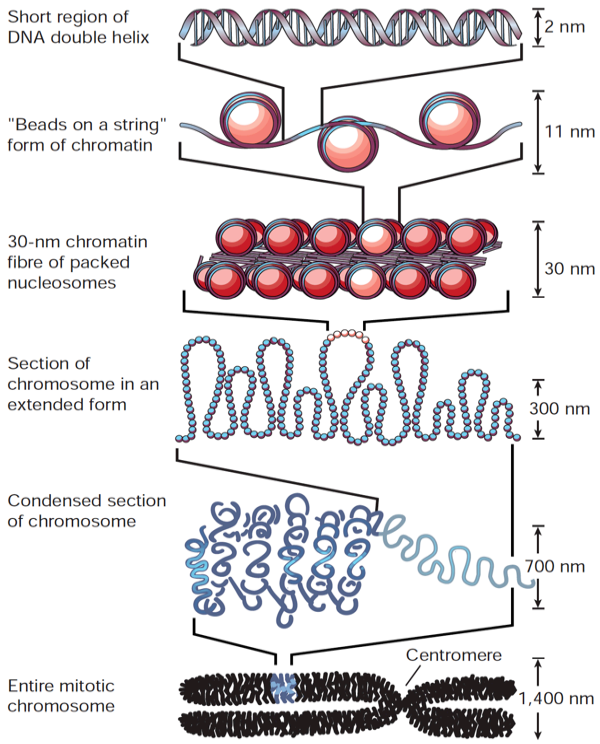


Figure 2. DNA organization within the chromatin structure. Two superhelical turns of DNA are wound around the histone octamer forming the nucleosome. Strings of nucleosomes are folded into a 30 nm fibers, which are further folded into high-order structures [2].

1.2. EPIGENETIC MODIFICATIONS

1.2.1. DNA methylation

DNA methylation is a covalent modification catalyzed by DNA methyltransferases (DNMTs) that occurs in the C-5 cytosine position at the CpG stie-regions of DNA [15]. In the genome, CpG-rich regions are known as CpG islands, that is, unmethylated GC-rich regions that possess high relative densities of CpG and are commonly found at the 5' regulatory region of many genes [16]. In mammals, the DNMTs family is composed of three active enzymes: Dnmt1, which is the responsible for the maintenance of DNA methylation, and Dnmt3a and Dnmtb are the responsibles for the *de novo* methylation of DNA [17]. DNMTs catalyzes the transfer of methyl group at the 5 position of cytosine ring, using S-adenosyl-L-methionine as a methyl-donor, and produces the 5-methylcytosine (5mC) and S-adenosyl-L-homocysteine [17].

While 5mC has been extensively studied, recent studies have demonstrated that Tet family proteins are capable of the generate 5-hydroximethylcitosine (5hmC) from oxidation of 5mC in mammalian cells [18,19]. Expression and functional analysis have shown that Tet1, a member of the Tet family, is highly expressed in mouse ES cells, and necessary for its maintenance and self-renewal [20]. Moreover, Tet1 occupies regulatory regions of both pluripotent genes and Polycomb-repressed developmental genes in mouse ES cells [21]. Conversion of 5mC to 5hmC is believed to be the first step in the DNA demethylation process, suggesting that generation of 5hmC can lead to passive demethylation in a replication-dependent manner [22].

The methylation of DNA is associated with gene silencing and plays a critical role in many cellular processes such as X chromosome inactivation, genomic stability, and embryonic development [23,24]. A mode of repression consists of direct interference of the methyl group with the binding ability of transcription factors to their cognate DNA sequence. The conserved family of proteins containing methyl-CpG binding domain (MBD) which binds selectively to methylated CpG and contribute to the repressive properties of DNA methylation [16]. Mammalian Mbd1, Mbd2 and Mbd4 binds to 5mC, in contrast to Mbd3, which binds to 5hmC regions of DNA and regulate genes whose regulatory sequences are enriched for this modification [18,25]. There is a clear evidence of interaction between of MBD-containing proteins and histone deacetylases and remodelling complexes, providing a mechanistic relationship of DNA hypermethylation and histone deacetylation to promote transcriptional repression [26].

In contrast to normal cells, cancer cells often show genome-wide DNA hypomethylation and hypermethylation of CpG islands located in the promoters of tumour suppressors genes such as *CDKN2A*, *BRCA1* and *VHL*, leading to the silencing of tumor suppressors and genomic instability [27].The mechanisms responsible for aberrant DNA methylation have not yet been concluded. One hypothesis is that aberrant DNA methylation occurs randomly throughout the genome, but hypermethylation at genes that limit cell proliferation (e.g., tumor suppressor genes) provides a selective advantage [28]. Another hypothesis is that aberrant DNA methylation is the result of aberrant targeting of DNMTs to certain regions, or that certain regions possess intrinsic features that make them better substrates for *de novo* DNA methylation [28].

1.2.2. Histone modifications

Post-translational covalent modifications on histone tails present a wide range of variation in epigenetic regulation with more than 60 different sites of modification. In the last few years, the use of chromatin immunoprecipitataion (ChIP) techniques has improved the analysis of histone modifications present at specific genomic sites, allowing a better understanding of such modifications and its enrolment in regulation of gene expression. Histones are subject to several distinct types of modifications which include acetylation, methylation (lysine/arginine), phosphorylation, ubiquitylation, sumoylation, ADP ribosylation, deamination and proline isomerization [29] (Fig. 3). Histone modifications have distinct impact on chromatin structure: acetylation causes a decondensation of the chromatin fiber since it neutralizes the basic charge of the lysine, whereas methylation can act as a mark for the recruitment of specific regulatory proteins in order to modulate DNA accessibility and therefore, gene expression. The coexistence of different histone modifications on the same histone tail suggests the existence of a "histone code" that can be read by external proteins to trigger a specific transcriptional program [30,31]. The recognition of histone marks is performed by proteins containing specific domains, which can recognize specific residues e.g. acetylation (bromodomain) or methylation (chromodomain).

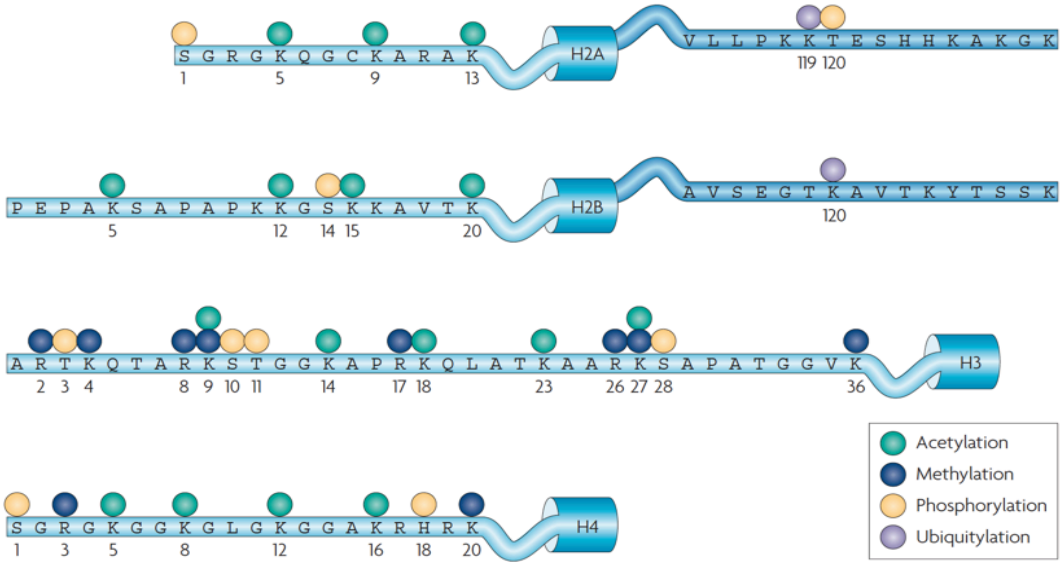


Figure 3. Histone code hypothesis. Amino-acid residues of histone tails are subject to various post-translational modifications, such as methylation, acetylation, phosphorylation, ubiquitination, sumoylation, citrullination and ADP ribosylation. Histone modifications have been associated with either active or inactive chromatin states, as well as with particular cellular processes. Based on these observations, it has been proposed that patterns of post-translational modification form a combinatorial "histone code" [151].

1.2.2.1. Histone acetyltransferases

The acetylation of histone residues is associated with transcriptional activation [32]. Hyper-acetylation of histone lysine residues is thought regulate transcriptional activity beacause acetylation brings a negative charge that neutralizes the positive charge of histone tails and decreases the electrostatic interaction between N-termini of histone with negatively charged phosphate groups of DNA [33-35]. As a consequence, the condensed chromatin adopts a more relaxed state to allow the access of transcription factors to gene promoters. However, this process can be reverted by histone deacetylation to restore the initial positive charge [36]. In addition to conformational changes, histone acetylation has been described to act as a binding platform to recruit different transcriptional regulators to promote gene activation [35]. In fact, acetylation of histone tails facilitates the binding of subunits of the TFIID transcription initiation complex on the gene promoters [37]. The transcriptional coactivator complex GCN5 is recruited to target promoters and acetylates histone H4K8, and histone H3K9 and H3K14. Subsequently, H4K8 acetylation is required for recruitment of BRG1, the ATPase subunit of the SWI/SNF complex that mediate nucleosome remodeling. As postulated by the histone code hypothesis, these findings demonstrate a functional interaction between chromatin remodeling complexes, such as SWI/SNF, and histone acetylase complexes, such as GCN5 [37]. Since the first histone acetyltransferase (HAT) was discovered in 1995, several proteins, previously described as transcriptional coactivators, have been identified to present HAT activity [32]. According to sequence similarities, HAT proteins can be classified in different groups such as Gcn5/PCAF, p300/CBP, and MYST family [38]. Moreover, protein acetylation can induce a crosstalk between other post-translational modifications such phosphorylation or methylation, among others, to control complex regulatory programs [39].

1.2.2.2. Histone deacetylases

Conversely to HAT, histone deacetylases (HDAC) are traditionally associated with transcriptional repression and gene silencing [40]. The first mammalian HDAC was related to the yeast transcriptional regulator Rpd3 [41]. Since then, several HDAC have been identified and classified in three different subgroups, based on its sequence similarities with their yeast homologs [36]. (1) Human HDAC class I include HDACs

1,2,3, and 8, and present a high degree of homology with the yeast Rpd3. Members of HDAC class I present a specific sensitivity for small molecule compounds that can act as inhibitors to block their catalytic activity [42]. Also class I members have been closely associated with other protein subunits that participate in transcriptional repression, such as Sin3 and N-CoR [43]. (2) HDAC class II include HDACs 4,5,6,7,9,and 10, and are similarly related with yeast Hda1. Members of class II display tissue-specific expression and can shuttle in and out between the nucleus and the cytoplasm. As a special mention to HDAC 11, although being mostly related with class I HDAC, it can't be classified in any of the three HDAC classes given that the overall sequence of similarity is too low. (3) HDAC class III, also named sirtuins, consists of Sir1 to Sir7 members, which are homologous to the yeast Sir2. Class III members exhibit significant structural and functional differences from class I and II groups, and present a NAD-dependent deacetylase activity [44]. Given the divergence with class I and II, HDAC class III members are not sensitive to traditional HDAC inhibitors. Unexpectedly, paradoxical findings have shown that histone deacetylation may also be required for transcriptional activation, suggesting that recruitment of HDACs complexes into a cytokine-inducible promoters is necessary for its activation [45]. Disrupting the balance between histone acetylation and deacetylation may have critical consequences in the regulation of gene expression, and it has been associated with tumor development and progression. Several studies have shown aberrant expression levels of individual HDACs in tumor samples. Therefore, the use of HDACs inhibitors emerged as a promising therapeutic treatment in cancer progression [46].

1.2.2.3 Histone methyltransferases

The methylation of histone tails does not affect directly the conformational state of chromatin. Instead, methyl marks on histone tails provide a binding site for the recruitment of transcriptional regulatory complexes. Histone methylation has been described as a critical player in the regulation of many biological processes such as transcriptional regulation, genome stability, and nuclear architecture [47-49]. The methylation of histone tails can be associated either to activation or repression, in a residue-depending manner [29]. Histone methyltransferases (HMT) methylates histone tails by transferring a methyl group from S-adenosyl-L-methionine into a lysine or arginine residue from the amino group of histone tails, leaving the cofactor byproduct

S-adenosyl-L-homocysteine [47]. According to this we can distinguish between histone lysine methyltransferases (HKMT) and histone arginine methyltransferases (PRMT).

1.2.2.3.1 Histone arginine methyltransferases

Protein arginine methyltransferases (PRMT) are able to catalyze mono- and di-methylation of arginine residues. There are three main forms of methylated arginine in mammals: monomethylarginines (MMA), asymmetric dimethylarginines (ADMA), and symmetric dimethylarginines (SDMA) [50]. PRMTs are evolutionary conserved from yeast to humans. The mammalian family of PRMT family consist of nine members, and can be classified as either type I, II or III. All members of PRMT type I and II can catalyse the formation of an MMA intermediate, whereas type I (PRMT1, 2, 3, 4, 6 and 8) can perform the formation of ADAMA, and type II (PRMT5 and 7) produce the SDMA. PRMT7 can only monomethylate certain residues, which is referred as type III [50]. Several members of this protein family have been described to participate in transcriptional regulation, e.g., PRMT6 mediates methylation of H3R2 and antagonizes to H4K4 trimethylation [51].

1.2.2.3.2 Histone lysine methyltransferases

Histone lysine methyltransferases (HKMT) are characterized by the presence of the catalytic SET domain, a module encoded within many proteins and strongly conserved among evolution. There is one exception, Dot1L, which is a HKMT member but does not have a SET domain [52]. The function of the SET domain is to catalyze the methylation of lysine residues using the AdoMet as a methyl-donor (Fig. 4). The SET domain was first identified as a conserved sequence of approximately 130 amino acids containing three *Drosophila* genes involved in epigenetic processes: the suppressor of variegation 3-9 (Su(var)3-9) [53], the Polycomb-group (PcG) chromatin regulator Enhancer of zeste (Ez), and the Trithorax-group chromatin regulator trithorax (Trx) [54]. The SET domains-containing proteins have been classified in different groups depending on the grade of sequence conservation in regions flanking the SET domain, where each group seems to present a specific affinity corresponding with a lysine residue [29,49]. Most of the targeted lysine methylation that has been reported so far occurs mainly on histone H3 tail, including H3K4, K9, K27, K36 and K79, and histone

H4 tail on H4K20. The biological significance of these modifications is dependent on a residue-specific manner and correlates with distinct states of gene expression [55]

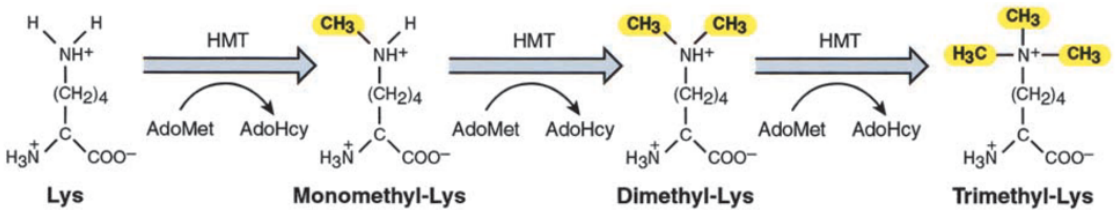


Figure 4. Chemistry of lysine methylation. Molecular structure of lysine and mono-, di- and tri-methylated lysisne [47].

The first HMT to be characterized was the mammalian homolog of *Drosophilia* Su(var)3-9, which specifically methylates H3K9 to mediate gene silencing [56]. In humans, SUV39H1 has been characterized as a nuclear protein that acts on the centromeres during mitosis methylating H3K9. This mark facilitates the recruitment of the repressive protein HP1, which is involved in higher-order heterochromatin organization together with HDACs enzymes [57].

Methylation of H3K27 has been associated with gene silencing and has been found on euchromatic gene loci, at pericentromeric heterochromatin, and at X inactivated chromosome in mammals [58,59]. The methylation of H3K27 is catalyzed by the Polycomb-repressive complex (PRC), which plays a central role in the epigenetic regulation of chromatin structure and gene expression. Recently, Polycomb has also been associated with the recruitment of several DNA methyltransferases, involved in the control and regulation of developmental programs such as stem cell self renewal and different cell linage commitment [60]. At least there are two different PRC well described, PRC1 and PRC2. PRC1 is a multiprotein complex that includes at least one paralog of the Pcgf, Ring1, Phc and Cbx [61]; whereas PCR2 consists of the three core components Ezh2, Edd and Suz12 among others [47]. Methylation of H3K27 is catalysed by PRC2 through its catalytic subunit Ezh2. This modification can be recognized by the Cbx subunits of PRC1, mediating its recruitment and subsequent ubiquitylation of H2AK119 by the Ring1 component. However, several reports have shown that global levels of H2AK119ub1 are preserved in PRC2-deficient cells,

suggesting the existence of H3K27me3-independent mechanisms of PRC1 recruitment [62,63]. Accordingly, recent findings show that Kdm2b participates in the recruitment of PRC1 complexes devoid of Cbx subunits to unmethylated CpG islands [64-66]

The methylation of H3K4 has been generally associated with transcriptional activation. In particular, H3K4 trimethylation has been found in the promoters of active genes and participates in the recruitment of the TFIID transcriptional initiation complex [67]. In humans, several H3K4 methyltransferases have been identified such as mixed-linage leukemia (MLL) family proteins, which play a critical role in gene expression. Among all H3K4 methyltransferases, MLL1 emerged as a potential candidate to become the yeast Set1 analog, given its homology on the SET domain and its interaction with basal transcription factors, in a similar manner as the yeast Set1 complex [68]. MLL1 is a mammalian member of the Drosophila Trithorax-group proteins. These proteins has been described to regulate and maintain the transcriptional states of Hox genes during development [69].

Trimethylation of H3K36 has been identified on the coding region of active genes, which correlates with active transcription. H3K36 has been associated with transcriptional elongation [59]. In yeast, RNA pol II is activated for transcriptional initiation after Ser5 phosphorylation of its carboxy-terminal domain (CTD), which permits the recruitment of Set1 to methylate H3K4. During transcriptional elongation, CTD-Ser2 is phosphorylated and recruits Set2, which subsequently methylates H3K36. In addition, H3K36me act as a binding site to recruit HDACs complexes, possibly to provide a transcriptional memory of the passing polymerase and ensure the fidelity of transcriptional initiation [70]. On the contrary, H3K36me can also recruit HAT through its PHD finger domain [71]. H3K36me mark has been observed on the exons of active genes relative to the promoter occupancy, while it remained enriched through the entire exon region [72]. Although it still remains unclear whether H3K36me is implicated in any developmental process [73], it has been related to participate in transcriptional activation and repression, dosage compensation, splicing, DNA replication, recombination and repair [74].

Methylation of H4K20 was one of the first histone modifications to be discovered, but its biological significance was not well understood until recently. This mark has been

related with transcriptional repression, and H4K20me3 appears to be functionally associated with H3K9me3 mark in the establishment and maintenance of pericentric heterochromatin [75]. The histone methyltransferse PR-Set7/Set8 specifically monomethylates H4K20 to regulate cell cycle progression, and to maintain silent chromatin for mitotic chromosome separation [76]. MMSET, another methyltransferase, dimethylates H4K20 and participate in DNA damage response [77]. Recently, SUV420H2-mediated trimethylation of H4K20 has been described to mediate gene silencing, through pausing Pol II, by acetylation of H4K16 [78].

1.2.2.4 Histone demethylases

Together with the discovery of histone methyltransferases, there was a permanent debate about the existence of unidentified histone demethylases (HDM) capable of catalyze the removal of methyl marks from histone tails lysines. The first HDM described was the lysine specific demethylase 1 (LSD1), which consists of an amine oxidase subunit, frequently associated within various HDAC corepressor complexes, and presents a specific demethylase activity for H3K4 mono- and di-methylated [79]. LSD1 belongs to the flavine adenine dinucleotide (FAD)-dependent amine oxidase family, where the removal of methyl groups consists of a substrate oxidation by the presence of FAD as a cofactor [80].

Subsequently, a new family of HDMs was identified containing a jumonji catalytic (JmjC) domain that has been previously implicated in chromatin dependent functions. Therefore, most of the JmjC family members possess a lysine demethylase activity with distinct lysine methylation sites, including the JHDM1A (H3K36me1/2) [81], JHDM2A (H3K9me1/2) [82], JMJD2 (H3K9me2/3 and H3K36me2/3) [83], JARID1 (H3K4me2/3) [84], and UTX/JMJD3 (H3K27me2/3) [85], among others. In the last years, JMJD3 has gained special attention because it was shown to play a decisive role in stem cells differentiation, by removing the repressive H3K27me3 mark of some lineage-specific genes during differentiation [86]. In addition to histone H3 tail demethylation, a PHD and JmjC containing-domain protein, PHF8, has been described as a H4K20me1 demethylase and plays a role in cell cycle progression [87]. Recently, another member of the JmjC family, JMJD6, was identified as the first arginine demethylase by removing methyl-groups from H3R2me and H4R3me [88].

Unlike LSD1, the catalytic mechanism of JmjC family members is capable to demethylate tri-methylated, as well as mono- and di-methylated lysines, via Fe(II)-dependent hydroxylation reaction. Importantly, genetic studies have identified mutations or aberrant expression levels of several demethylases in human diseases such as neurological disorders and cancer [89].

1.3. CHROMATIN REMODELING

The regulation of chromatin architecture is necessary to facilitate the control of gene expression. Transcription is regulated by the integration of many *cis*-regulatory elements such as core promoters and promoter-proximal elements, in addition to some others *cis*-regulatory modules which are located at relatively large distances from the transcriptional star sites, like enhancers, silencers, insulators, and tethering elements [90]. During transcriptional activation, transcription factors need to bind to specific DNA sequences of a gene promoter to facilitate transcription, but most of these DNA-binding sequences are wrapped inside the nucleosome, remaining inaccessible [90]. In order to gain access to these specific binding sites, several chromatin-remodeling complexes participate in nucleosome displacement to modulate promoter accessibility and generating a suitable environment for transcriptional initiation [91]. Nucleosomes are dynamic structures, and the histone octamer can either be displaced to a neighbouring DNA sequence, called sliding, or exchanged in and out of DNA, known as replacement [92].

The process of nucleosome mobilization is carried out by a family of enzymes, the ATP-dependent chromatin remodelling complex, in a non-covalent manner. Due to its implication in regulating gene expression, these complexes are fundamental elements that present an important role during the regulation of cell cycle, cell differentiation, and development [93]. Depending of their ATPase catalytic subunits and domain structure, chromatin remodeling complexes can be classified in three distinct groups: the SWI/SNF-type complex [94], the ISWI/SNF2L-containing machines [95], and the CHD-containing complexes [96]. The ATPase subunits Swi2/Snf2, ISWI, and Mi-2 define the three different groups, respectively. Remodeling complexes can be associated with other regulatory proteins, such as histone modifying enzymes, to form multiprotein complexes [97] (Fig. 5).

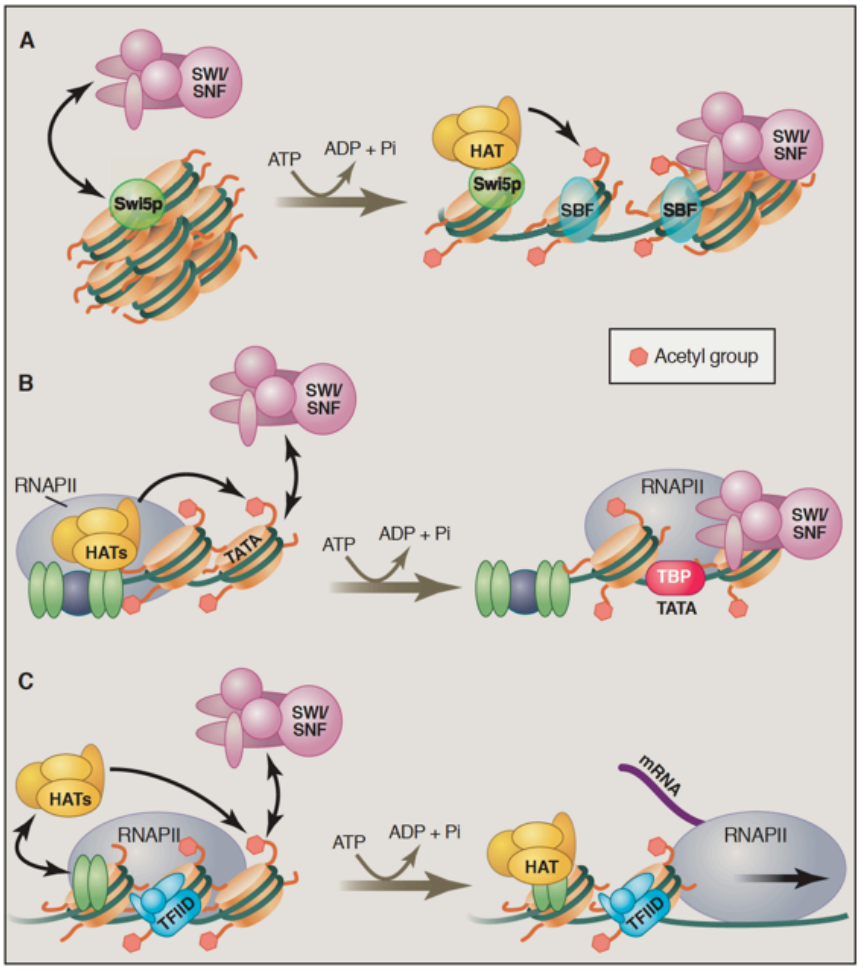


Figure 5. Chromatin remodeling. Chromatin-remodeling is necessary for activation of gene expression. (A) Yeast HO promoter. Swi5p activator recruits the SWI/SNF and Gcn5p HAT very early in gene activation. SWI/SNF and HAT complex cooperate to facilitate binding of a second activator, SBF. (B) Human *IFN-β* gene promoter. Upstream activators (green and purple) recruit multiple HAT proteins during the assembly of the preinitiation complex (PIC). Histone acetylation promotes recruitment of the SWI/SNF, which facilitates the TBP binding to the TATA element, completing the PIC assembly. (C) Human α_1-*AT* gene promoter. Multiple HAT complexes, CBP and P/CAF, and the SWI/SNF complex are recruited after PIC assembly [97].

The binding of chromatin remodeling complexes to nucleosomes is achieved through the presence of certain protein domains that are able to recognize various chromatin structures. Among different domains, two well-characterized domains are the bromodomain, which can recognize the acetylated tails of histones, and the chromodomain, which can recognize the methylated tails of histones. The bromodomain was first identified in the *Drosophila* chromatin remodeling protein Brahma and consist on a left-handed four-helix bundle that recognizes acetylated N-terminal histone tails [98,99]. Bromodomains have been found in three major chromatin-associated protein families: ATP-dependent chromatin remodelling factors, histone acetyltransferases, and

BET transcriptional regulators [99]. Chromodomains consist on a monomeric three anti-parallel beta sheet flanked by a single C-terminal alpha helix [100]. This domain was first identified as a 37 amino acid homologue region with *Drosophila* Polycomb protein and heterochromatin protein 1 (HP1) [101]. Chromodomains define CMi-2HD family remodelers, which typically contain two tandem chromodomains in their N-terminal region [102]. Certain tandem CHD domains function as a unit to bind one methylated lysine. Unlike HP1 and Polycomb proteins that use single chromodomains to bind their respective methylated histone H3 tails, the two chromodomains of human CHD1 cooperate to interact with one methylated H3 tail [103].

1.4. SMYD2

1.4.1. SMYD family of histone lysine methyltransferases

The family of **S**ET and **MYN**D containing proteins (SMYD), contain a SET domain that is split into two segments by an MYND domain/zinc finger motif, followed by a cysteine-rich post-SET domain. The SET domain provides its methyltransferase activity by adding methyl groups to lysine residues using AdoMet as a donor substrate. The MYND domain encompasses a putative zinc-finger motif that facilitates protein-protein interactions through the conserved PXLXP motif-containing proteins. This domain is present in several other transcriptional regulators where it is known to contribute in developmental processes [104,105].

In human, there are five members in the SMYD protein family (SMYD1-5) and have been shown to participate regulating gene transcription and cell proliferation (Fig 6). SMYD1 is a heart and muscle specific histone methyltransferase involved in cardiomyocyte and myogenic differentiation [106,107]. The lack of Smyd1 in mice development results in embryonic death due to cardiac defects [106]. Knockdown of smyd1a/b in zebrafish causes skeletal and cardiac muscle defects and presents a disrupted expression of myofibril organization [108]. SMYD3 has been mainly related with cancer cell proliferation [109]. Several findings indicate that endogenous expression of SMYD3 is present at very high levels in hepatocellular, colon and breast carcinoma, and silencing through siRNAs have an inhibitory effect in cell growth [110]. Similarly to smyd1a/b, smyd3 plays an important role in cardiac and skeletal muscle

development in zebrafish [110]. In the other hand, SMYD4 is significantly reduced in tumor cells and its re-expression dramatically decreases cancer cell growth [111]. Also, *Drosophila* SMYD4 homologue has been involved in muscle development [112]. Little is known so far about SMYD5. Unlike the rest of family members, SMYD5 does not present a C-terminal tetratricopeptide repeat (TPR) domain [113].

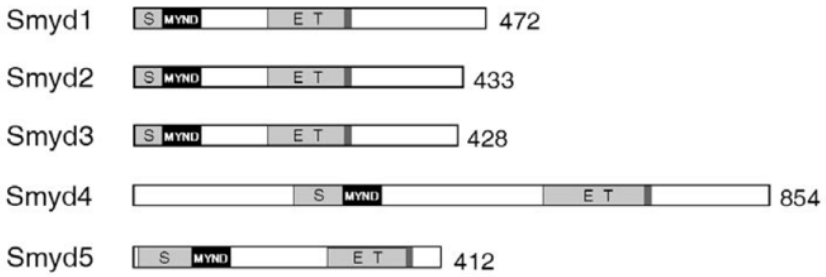


Figure 6. Schematic representation of mammalian SMYD family members. All family members present a split SET domain (light gray), a MYND domain (black) and a cysteine-rich post-SET domain (dark gray) [119].

1.4.2. SMYD2 structure

The overall structure of SMYD2 consists of 433 amino acid, and is composed of five structurally distinct domains, which together form two large lobes that are separated by a deep groove [114-116] (Fig. 7). The first lobe (1–282) comprises the first four domains and the N-terminal catalytic SET domain, and the second lobe (283-433) the C-terminal three helix-turn-helix domain, containing tetratricopeptide repeats (TPR). The core of the SET domain (split into 1-44 and 182-244) shares a characteristically conserved folding structure, formed by three sets of anti-parallel β- strands. The MYND zinc binding domain is inserted between the N- terminal core SET domain (45-99), and consists of a long bent α-helix αA and a few loops which are organized by two zinc ions coordinated by a typical C2HC motif composed of seven cystein residues. Between the MYND domain and the C-terminal portion of the catalytic SET domain we found the insertion SET domain (100–181) and finally, the cysteine-rich post-SET domain (246-282). As expected, all these four domains have been also observed in other protein lysine methyltransferases, but the split SET domain is restricted to the SMYD protein family [115]. Together with SMYD2, the structure of mouse SMYD1 and human SMYD3 present a similar architecture, but with different arrangements on the N- and C-

terminal lobes [117,118]. The interaction with the methyl donor SAM takes place at the bottom of the deep surface groove that separates the N- and C-terminal lobes of SMYD2, where the SAM bounds to the protein. Given the functional and structural importance of this region, removing the first 13 residues would clearly disrupt the structural integrity of the cofactor binding site [116].

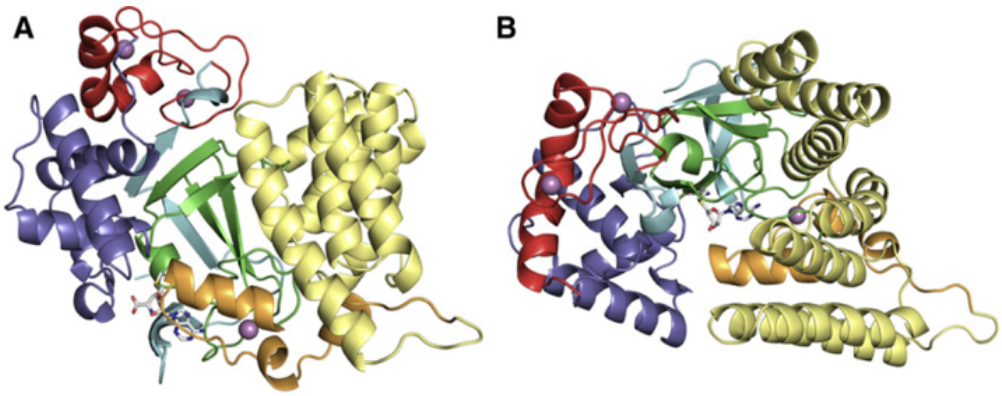


Figure 7. Overall structure of SMYD2. The view is given (a) directly above the pronounced surface groove that separates the N- and C-terminal lobes of SMYD2, and (b) rotated by 90°. The S-sequence is shown in cyan, the MYND domain is red, the I-SET domain is blue, the core SET domain is green, the post-SET domain is orange, and the C-terminal domain is yellow. Three coordinated zinc ions are also shown in purple [116].

1.4.3. SMYD2 activity

SMYD2 was first described as a histone lysine methyltransferase with specific H3K36 dimethylation activity, a mark associated with actively transcribed genes [119]. Moreover, SMYD2 has been described to interact with HDAC1 and the Sin3A repressor complex [119]. Although H3K36 methylation has been associated with transcriptional activation, there is evidence of a link between H3K36 methylation and the recruitment of repressive Rpd3 (the prototypical yeast HDAC) complex in yeast [70,120,121]. Conversely, SMYD2 was also described to specifically monomethylate H3K4 in the presence of HSP90α *in vitro*, with no activity for H3K36 [122]. Despite observing a weak activity of H3K36 methylation in the absence of HSP90α, *in vivo* experiments suggested that H3K4 is the predominant site of methylation by SMYD2 [122]. Recent studies determined that SMYD2 also methylates histones H2B and H4 more efficiently than H3 *in vitro* [123].

Several non-histone proteins have been also identified as substrates for SMYD2 methylation such as p53 and retinoblastoma (Rb) [124-128]. Monomethylation of p53 at K370 reduces its binding to target genes like p21 and mdm2, resulting in a decreased expression of these genes [124]. Unexpectedly, the adult hearts of cardiac-conditional Smyd2 knockout mice showed no changes in p21 and mdm2 expression and no significant effects in the global levels of H3K36 or H3K4 methylation [129]. Additionally, Smyd2 was dispensable for proper heart development in mouse [129]. Rb protein can be methylated by SMYD2 at K860 facilitating its interaction with the methyl-binding protein L3MBTL1 [126]. More recently, SMYD2 was also found to methylate Rb at K810, which increases Rb phosphorylation and promotes cell cycle progression [128]. The effects observed on p53 and Rb tumor supressors are in accordance with the high levels of SMYD2 expression found in human cancer [128,130], suggesting a possible role during tumor progression and proliferation. However, contrary to these observations where SMYD2 promotes cell proliferation, previous data has shown that exogenous expression of SMYD2 significantly reduces the proliferation rate of NIH3T3 [119]. It was also reported that SMYD2 is involved in maintaining self-renewal activity of MLL-AF9-induced acute myeloid leukemia [131]. Another novel non-histone substrate for SMYD2, cytoplasmic HSP90, is methylated at K209 and K615 [113]. HSP90 is known to regulate myosin stability and sarcomeric A-band organization in skeletal muscle. In mouse, Smyd2 methylates HSP90 to form a complex with the sarcomeric protein titin to protect myocyte organization [132,133].

1.4.4. SMYD2 expression patterns

SMYD2 has a wide expression along different tissues in mouse embryos, but is most highly expressed in heart, skeletal muscle and brain tissues, as demonstrated by *in situ* hibridization and RT-PCR [119,129].

In zebrafish, SMYD2 homologues *smyd2a* and *smyd2b*, present a high expression levels in somites and muscle cells, and *smyd2a* was also observed to be significantly, though weakly, expressed in heart primordium during early development [134]. Moreover, the same study identified *smyd2a* as a maternally expressed gene in zebrafish embryos.

In addition, SMYD2 was also found to be overexpressed in a wide range of human tumor samples and esophageal squamous cell carcinoma (ESCC) cell line [128,130].

1.5. EMBRYONIC STEM CELLS

Embryonic stem (ES) were first isolated in 1981 from mouse embryos [135]. ES cells are derived from the inner cell mass (ICM) at the early blastocyst stage of the pre-implantation embryo, although some researchers have shown that it is also possible to obtain ES cells from earlier stage embryos [136,137]. The main properties of ES cells are their ability to self-renew indefinitely *in vitro* maintaining their undifferentiated state, and their potential to give rise to derivatives of the three germ layers (endoderm, ectoderm and mesoderm). For these reasons, ES cells hold tremendous potential for regenerative medicine and tissue replacement. Recently, it has been reported that terminally differentiated somatic cells can be reprogrammed into induced pluripotent stem (iPS) cells by the overexpression of a defined set of transcription factors, as described by Takahashi and Yamanaka [138]. These iPS cells are morphologically and phenotypically indistinguishable from ES cells. The use of human ES cells has been controversial due to the need to destroy human embryos for their isolation. This and the fact that iPS cells can be generated for autologous therapies make iPS cells a promising alternative to ES cells in regenerative medicine. Given the strong impact of iPS cells discovery, Dr. Yamanaka, together with Dr. John B. Gurdon, was awarded with the Nobel Prize in Physiology or Medicine in 2012.

1.5.1. Maintenance of self-renewal

1.5.1.1 The pluripotency network

One of the main goals in stem cell biology is to decipher the molecular basis underlying ES cell physiology, which is crucial to understand the key components that regulate self-renewal and pluripotency. Key transcription factors have been identified to form a coordinated regulatory network that play an essential role in maintaining the pluripotent stem cell phenotype [139,140]. The first transcription factor identified as indispensable for pluripotent stem cells in the mammalian embryo was the POU transcription factor, OCT4 [141]. The precise control of OCT4 levels is crucial to determine ES cell identity

and cell fate. An increase in the levels of OCT4 causes the differentiation of ES cells into primitive endoderm and mesoderm, whereas a reduction initiates the differentiation into trophectoderm [140]. Another strong candidate, SOX2, has been implicated in the regulation and preservation of developmental potential [142]. The homeobox transcription factor NANOG was also identified as a master regulator in pluripotent cells. NANOG knockout embryos are not able to form epiblast and present problems with extraembryonic tissue, while overexpression experiments exhibit a LIF-independence condition to maintain pluripotent stem cells in culture [143]. Considerable evidence indicates that this "trinity" of transcription factors, OCT4, SOX2 and NANOG, function in combinatorial complexes that govern pluripotency *in vivo* and *in vitro* [144,145]. The three transcription factors co-occupy a large number of genes in ES cells, including genes involved in self-renewal and developmental regulators [139]. Moreover, OCT4 and SOX2 have been known to act synergistically to stimulate their own transcription [146]. The mechanisms by which these transcription factors activate the expression of self-renewal genes while participating in the repression of developmental regulators is not fully understood. Interestingly, OCT4 and NANOG have been reported to interact with proteins from multiple repression complexes, suggesting their involvement in transcriptional repression [147].

1.5.1.2. Bivalent domains

It has been reported that many developmental genes in ES cells present both active H3K4 and repressive H3K27 methylation marks on their regulatory regions, the so-called "bivalent domains" [148,149]. This histone modification pattern consists of large regions of H3K27 trimethylation harbouring smaller regions of H3K4 trimethylation around the transcriptional start site [150]. This bivalency keeps these genes silenced in ES cells, but "poised" to become activated during differentiation [151] (Fig. 8).

The molecular mechanisms involved in maintaining developmental genes poised in ES cells are not fully understood. Although H3K27 trimethylation has been proposed to participate in the repression of lineage-specific genes [152,153], self-renewing mouse ES cells can be derived from PRC2-deficient blastocysts, but these show defects on lineage specification [154]. [155]. Interestingly, H2A ubiquitination by PRC1 has been suggested to restrain poised RNA Pol II at bivalent domains [156]. On the other hand,

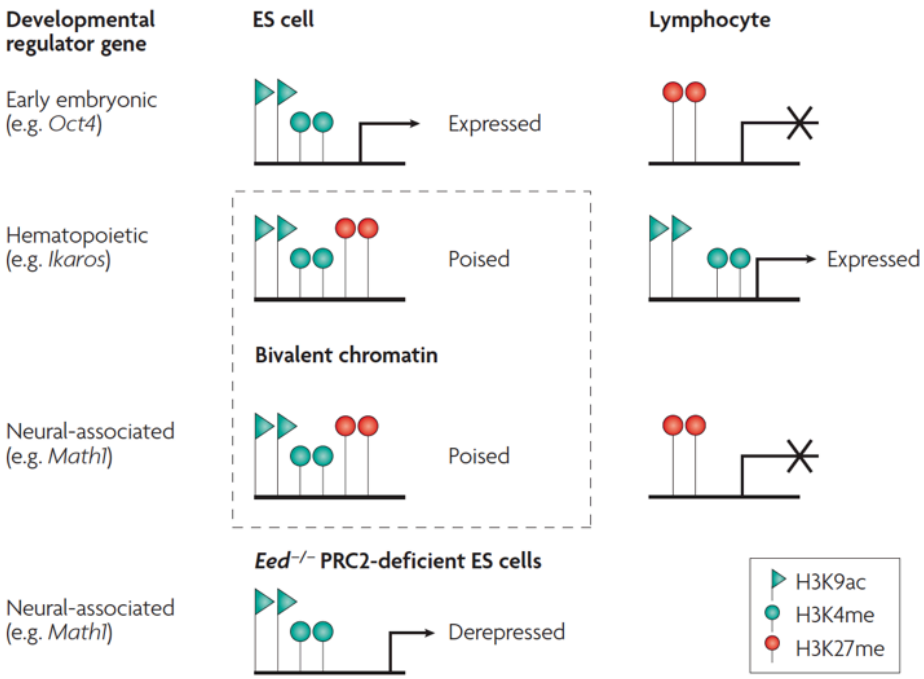


Figure 8. Bivalent domains in ES cells. The promoters of a range of non-transcribed developmental genes in ES cells present a combination of histone modifications associated with either active (H3K9ac and H3K4me) and inactive (H3K27) chromatin states. During differentiation, "bivalent" marks are resolved, leading to transcriptional activation of tissue-specific genes and silencing of loci associated with alternative developmental pathways [151].

depletion of a core subunit of the MLL complex results in a partial reduction of the levels of H3K4 trimethylation in mouse ES cells that alters the activation of genes upon differentiation but has no major implications for self-renewal [157]. However, LSD1 depletion in human ES cells causes the induction of developmental genes due to increase in the levels of H3K4me2 at the regulatory regions of these genes [158].

The mechanism by which PcG proteins are recruited to bivalent genes is still poorly understood. A recent study has shown that Oct4 can interact with components of the MLL and PRC complexes [159], suggesting that pluripotency factors may play a role in the establishment of bivalent domains in ES cells. Additionally, Jarid2, a JmjC family member without catalytic activity, has been described to facilitate the recruitment of PRC2 to target genes in mouse ES cells [160,161].

Interestingly, many inactive genes in ES cells have been found to be marked with H3K4me3 only. Despite this activating modification, these monovalent genes were neither expressed nor stably bound by RNA polymerase II nor transcriptional activators [162,163].

1.5.2. Differentiation of embryonic stem cells

The differentiation of embryonic stem cells is characterized by the silencing of the expression of pluripotency-related factors and the induction of developmental genes in a germ layer specific fashion. All these gene expression changes are orchestrated by a complex network of signalling events that respond to the presence of certain factors in the media.

During differentiation bivalent domains can be resolved into H3K27me3 only, H3K4me3 only or remain bivalent depending on cell fate decisions [164]. The activation of developmental genes during lineage commitment requires the removal of the H3K27me3 mark. UTX and Jmjd3, two histone H3K27me3/me2 demethylases, have been described to participate in this process [165,166]. For example, Jmjd3 is required for neural lineage commitment by resolving the bivalent domain at the *Nestin* promoter in mouse ES cells [86].

Importantly, the gene clusters repressed by PcG proteins in ES cells change during differentiation. In response to differentiation signals, PcG complex are recruited to other different promoters regions that are required to become silenced during differentiation [167-169]. Furthermore, some ES cell-specific genes such as *NANOG* are marked with H3K27me3 during differentiation, indicating that upon lineage commitment pluripotency genes are required to be silenced [170]. In order to avoid a possible reactivation stem cell-specific genes, DNA methylation occurs at the promoter region of key pluripotent transcription factors during differentiation [171].

1.6. ZEBRAFISH AS A MODEL SYSTEM

During the past three decades, the zebrafish (*Danio rerio*) has become an excellent model system for the study of developmental processes, physiology, and human disorders [172]. Compared to other vertebrate model systems, the zebrafish presents some advantages such as: short generation time (3-4 months), high fecundity rate (200-300 eggs/week), and embryos are transparent and develop outside the body, making them very accessible to visualization and manipulation during all stages of their development. In addition, zebrafish are easy and inexpensive to maintain.

Another advantage of using the zebrafish as an *in vivo* model is the highly degree of genetic conservation with higher vertebrates [173]. Thus, most of the molecular mechanisms that regulate embryonic development and many other physiological processes are shared between zebrafish and mammals [174]. For that reason, during the last years the zebrafish has gained a special interest for modeling human diseases including cardiovascular disorders, muscular dystrophy, neural disorders, and cancer therapies [175-178].

As a vertebrate model, the zebrafish presents a wide range of genetic tools originally developed for research on developmental biology and embryogenesis [179]. Recently, the zebrafish has emerged as a suitable system for large-scale forward genetic screens and gene inactivation studies. One efficient method for gene knockdown is the use of morpholinos, a chemically modified antisense oligonucleotides that transiently inhibits gene expression by blocking translation or splicing [180,180,181]. Morpholinos are injected at one- or few-cell stage embryos and can maintain their activity up to several days later. Considering that in zebrafish most of the organs are already developed and functional during the first five days after fertilization, this technology provides a rapid analysis of a specific gene function during the first stages of development. Another strategy to inactivate gene expression is known as TILLING (targeting-induced local lesions in genes), which generates stable zebrafish mutants. This method consists in a combination of standard mutagenesis using ethylnitrosourea (ENU), an alkylating mutagen that induces point mutations, together with genome sequencing in order to identify specific mutations [182]. There are alternative methods for the generation of stable zebrafish mutants including the use of zinc finger endonucleases, transcription activator-like effectors, and transposon-mediated systems [183-185].

1.6.1. Early development

Zebrafish embryogenesis can be classified in different periods as shown by Kimmel *et al*. [186,187] (Fig. 9). These divisions are based on the morphological features and major developmental processes that take place during the first three days after fertilization:

- *Zygote (0 - 0.75 h)*. The fertilized egg remains at the one-cell stage until the first cleavage division take place about 40 minutes after fertilization.

- *Cleavage (0.75 – 2.25 h)*. From one-cell stage, successive cell divisions of the blastomeres take place at the animal pole until reaching the 64-cell stage.

- *Blastula (0.75 – 5.25 h)*. The blastula period is comprised between the 128-cell stage and the onset of gastrulation. Around the 512-cell stage the embryo enters the midblastula transition (MBT) in which zygotic gene transcription is activated. Also at this point, the marginal cells release their cytoplasm and nuclei together into the adjoining cytoplasm of the yolk cell, forming the yolk syncytial layer (YSL). At the late blastula stage, epiboly appears and produces a blastoderm. The fraction of the yolk covered by the blastoderm is known as percent-epiboly. Blastula stage ends with 30%-epiboly.

- *Gastrula (5.25 – 10.33 h)*. During this period, morphogenetic cell movements produce the primary germ layers and embryonic axis. At 50%-epiboly appears the marginal region termed germ ring. Within the germ ring there are two germ layers: the epiblast (upper) and hypoblast (lower). Epiblast cells move towards the margin, and the cells from the margin move inward to enter the hypoblast. Then, convergence movements produce the local accumulation of cells along the germ ring, the so-called embryonic shield. Epiboly continues its expansion until cover the yolk cell completely and, just after the yolk closure, the posterior region of embryonic axis develops the tail bud. By the end of the gastrulation, the cells that remained in the epiblast correspond to the definitive ectoderm, whereas the cells from the hypoblast will give rise to the mesoendoderm.

- *Segmentation (10.33 – 24 h).* This period is characterized by the development of somites, the observation of primary organs, the prominent extension of the tail bud, and the embryo elongation. Somites appear sequentially in the trunk and the tail. Several somite divisions will take place during this period, from one- to 26-somite stage.

- *Pharyngula (24 – 48 h)*. The body axis begins to straighten and the head straightens out and lifts dorsally. At this stage embryos present a well-developed notochord, a nervous system expanded anteriorly, a functional circulatory system and a beating heart.

- *Hatching (48 -72 h)*. Morphogenesis of many of the rudimentary organs is now completed and slows down considerably. After this period, the zebrafish embryos are known as "larvae".

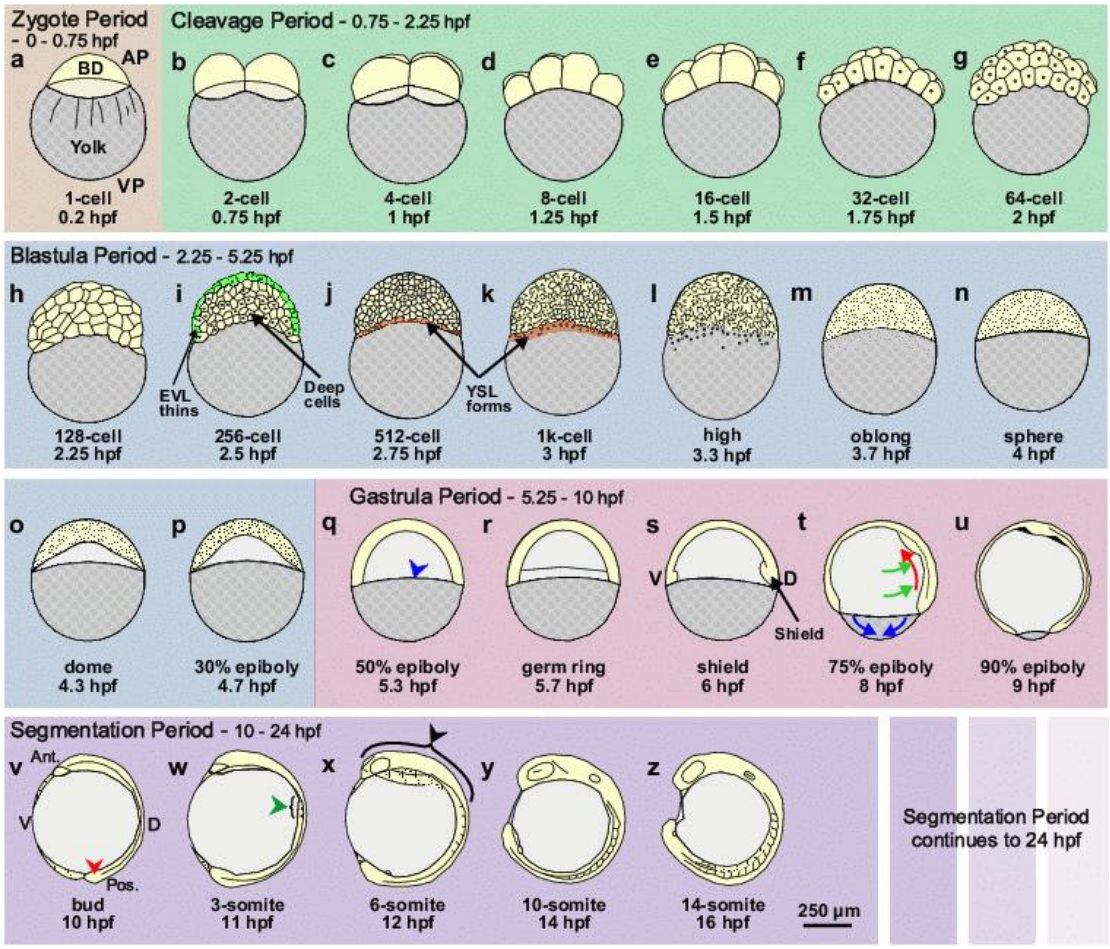


Figure 9. Schematic representation of zebrafish development from the Zygote Period to the mid-Segmentation Period. In the Zygotic Period, BD indicates the blastodisc at the one-cell stage, which develops into blastoderm during subsequent stages of development. The Cleavage Period runs from the two-cell stage to 64-cell stage. During the Blastula Period, which comprises from 128-cell stage to the 50% epiboly stage, formation of the EVL and YSL are indicated in panels I, and j and k, respectively. The Gastrula Period runs from the end of 50% epiboly stage through the bud stage. The leading edge of the blastoderm is indicated by a blue arrowhead in the panel q, and the simultaneous cell movements of epiboly (blue arrows), convergence (green arrows) and extension (red arrows) are shown in panel t. The Segmentation Period is only partially shown at early stages, from bud stage to the 14-somite stage [187].

1.6.2. Signaling events during gastrulation

The precursors of the different germ layers are already distributed along the animal-vegetal axis by the onset of gastrulation [188]. The mesoendodermal cells are located at the margin region adjacent to the yolk, the mesoderm is next to the mesoendodermal cells towards the animal pole, and the animal pole give rise to the ectoderm. There are different signaling pathways governing the specification of the three germ layers: Nodal, Bmp, Wnt and Fgf.

The ectoderm differentiates into epidermis, spinal cord and brain. Combinatorial Fgf and Bmp signaling patterns the gastrula ectoderm [189]. In the vegetal ectoderm, Fgf activity initiates the development of neural tissue that contributes to trunk and tail CNS. In the animal ectoderm, high levels of BMP activity induce non-neural fate whereas in vegetal ectoderm, differential levels of BMP promote the ability of cells to contribute to caudal neural ectoderm [189].

The mesoderm develops from the mesoendoderm ring away from the yolk and develops into notochord, head, trunk somites, tail somites, vasculature and erythroid cells. It has been found that the overlapping activity of Nodal, Bmp, Wnt and Fgf pathways lead to the formation of mesoderm [190]. The effect of the different signaling pathways on the mesodermal cells is dependent on their location. Nodal dosage received by mesodermal cells is higher at the dorsal and ventral region than the ones at the ventral and animal region. In the lateral mesoderm, medium levels of Nodal, Wnt and Fgf activate the expression of *tbx6*, *ntl* and *tbx6* genes, which are required for the trunk and tail mesoderm formation [191,192].

The endoderm is derived from the most marginal cells at the dorsal and lateral region of the blastoderm [193] and differentiates into pharynx, stomach/intestine and liver. It has been shown that endoderm segregates from mesoendodermal cells by a combinatorial mechanism of Nodal, Bmp, and Fgf signaling [194]. Apparently, Bmp (ventral site) and Fgf (dorsal site) pathways cooperate to restrict the number of endodermal progenitors induced in response to Nodal pathway. Fgf/ERK signaling phosphorylates Casanova, an important regulator of endoderm formation, and reduces its ability to induce *sox17* [194].

1.6.3. Signaling pathways

Nodal signaling pathway. Both Nodal and Bmp pathways are members of the Tgf-β superfamily. Tgf-β is activated through cell surface complexes by specific ligand-binding into type I and type II serine/threonine receptors [195]. Type II receptor phosphorylates the cytoplasmatic domain of the type I receptor, which become activated and phosphorylates a Smad factor. In zebrafish, Nodal ligands include Squint and Cyclops [196]. Type I and type II Activin receptors mediates Smad2/3 phosphorylation and binds to Smad4 co-factor. Smad complex translocate to the nucleus where they bind to specific DNA-binding factors and regulate the transcriptional activation of specific target genes [197,198]. Members of the Nodal family of Tgf-β signals are essential inducers of both mesoderm and endoderm.

Bmp signaling pathway. In zebrafish, the Bmp pathway comprises several ligands such as Bmp1a, Bmp2a, Bmp2b, Bmp4, Bmp5 and Bmp7 [190,199]. Type I and type II Bmp receptors mediate Smad1/5/8 phosphorylation and binds to Smad4 co-factor. The Bmp pathway present several inhibitors of Smad1/5/8 phosphorylation including Chordin, Noggin 1, and Follistatin [200-202]. During gastrulation, Bmp expression has been found highly expressed in the ventral region and is involved in patterning the mesoderm [190,203]. Several studies have demonstrated that Bmp is necessary for tail formation [204-206], and embryos with a mutant Bmp pathway exhibit expanded trunk muscle, abnormal tails and severe defects in ventral mesoderm such as vasculature and blood [207]. Recently, it has been shown that Bmp is necessary to establish the trunk-tail boundary and, in addition, it has also an important role to regulate the morphogenesis of cells on the ventral side of the embryo, so that they will end up in the tail bud [208].

Wnt signaling pathway. The Wnt pathway can be divided in two main branches: the canonical pathway (β-catenin-dependent) and the non-canonical pathway (β-catenin-independent). In the canonical pathway, Wnt ligand binds to Frizzled /Lrp receptor and mediates intracellular response by disrupting the called b-catenin protein complex. When β-catenin is not degraded, it translocates into the nucleus and associates with Tcf transcription factor to activate expression of downstream genes. In zebrafish, maternal β-catenin is necessary for dorsal YSL gene activity and organizer formation [209]. Gsk3 and Axin1, both components of the β-catenin protein complex, have been

required for the determination of dorsal fate [210,211].

Fgf signaling pathway. Binding of Fgf ligand to the Fgf receptor results in receptor dimerization and phosphorylation of the cytosolic domain. Phosphorylated Fgf receptor activates G-protein Ras, which activates Raf kinase. Raf phosphorylates and activates Mek, which subsequently phosphorylates and activates Mapk. Mapk enters the nucleus and activates the target transcription factors. In zebrafish, injection of *fgf17b* mRNA induces expression of *ntl* (mesodermal marker) and *chordin* (dorsal marker) [212]. Activation of Fgf pathway results in embryo dorsalization, by inhibiting ventral *bmp* gene expression [213].

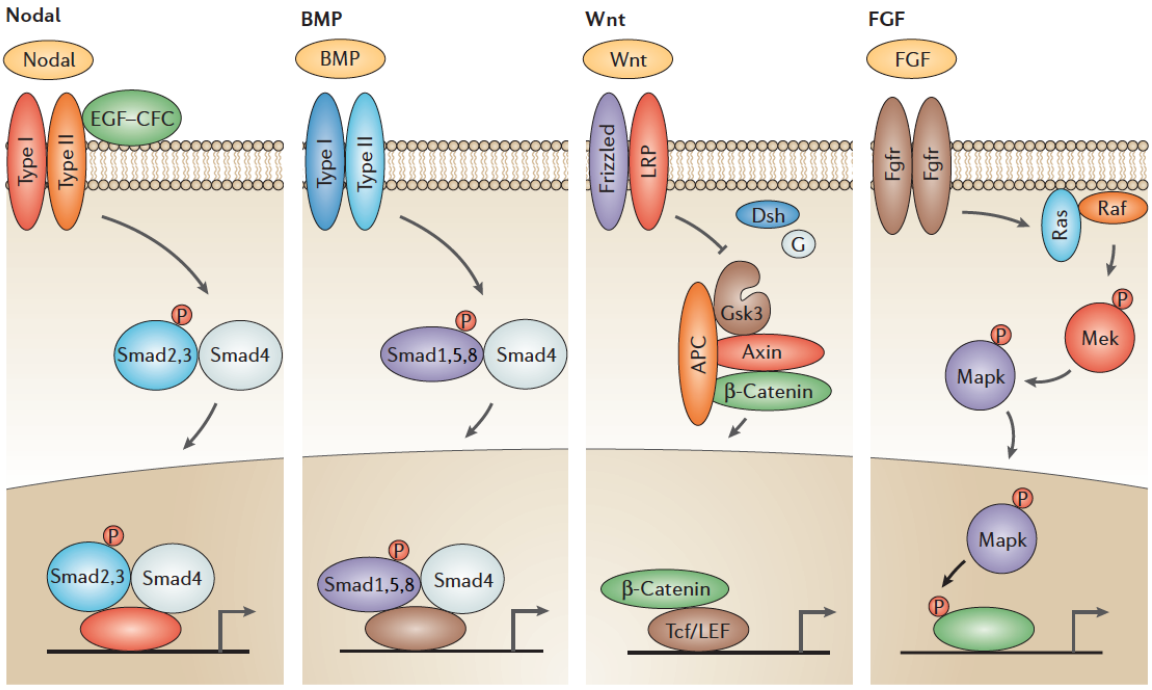


Figure 10. Intracellular signaling pathways. Essential details of the intracellular pathways that are used by the four signaling factors Nodal, Bmp, Wnt and Fgf [190].

2. SPECIFIC AIMS

1. Evaluate the expression of SMYD2 in undifferentiated and differentiated human ES cells and somatic cells.

2. Test the involvement of the identified candidate SMYD2 in the differentiation of human embryonic stem cells by gain- and loss-of-function strategies.

3. Determine the molecular mechanisms by which SMYD2 controls the differentiation of human ES cells.

3. RESULTS

3.1. SMYD2 EXPRESSION IS PREVALENT IN DIFFERENTIATED CELL TYPES

3.1.1. SMYD2 is highly expressed in somatic cells

In an effort to understand how the epigenetic landscapes are maintained in pluripotent and somatic cells we analyzed the potential differential expression of histone modifying enzymes between human pluripotent and differentiated cells. Preliminary data suggested that the histone methyltransferase SMYD2 could be differentially expressed in these cell types. To confirm this, we analyzed the expression levels of SMYD2 in several pluripotent and somatic cells by quantitative PCR (qPCR) (Fig. 11). Pluripotent cells included three different human ES cell lines (ES[4], ES[2] and ES[6]) and an iPS cell line derived from human keratinocytes (KiPS4F1); and differentiated cells included 293T, human foreskin fibroblasts (HFF), human keratinocytes (HK), a human liver hepatocellular carcinoma cell line (HepG2) and ES[4] cells after 15 days of *in vitro* differentiation (ES[4] D15). Our data shows that *SMYD2* mRNA is preferentially expressed in human somatic cells compared to pluripotent cells.

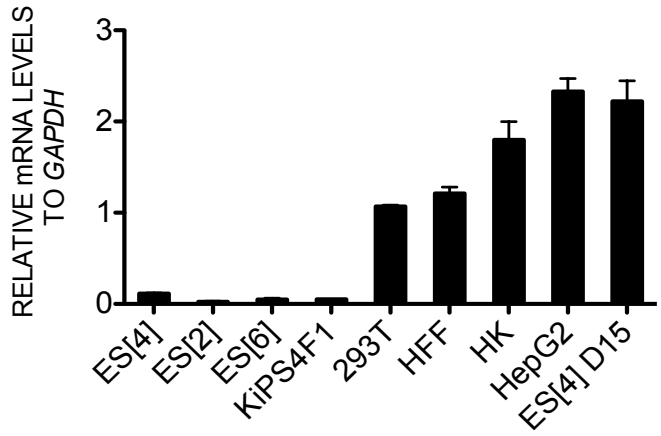


Figure 11. Expression of SMYD2 in different cell types. Comparison of *SMYD2* mRNA levels between pluripotent (ES[4], ES[2], ES[6] and KiPS4F1) and differentiated (293T, HFF, HK, HepG2 and ES[4] at day 15 of *in vitro* differentiation) cells. mRNA levels were quantified by qPCR and normalized to *GAPDH*. Means and standard deviations from two independent RNA extractions are shown.

Next we asked whether other SMYD family members could follow the same expression pattern as *SMYD2* in pluripotent and differentiated cells. For that, mRNA levels of all five SMYD family members were measured in human ES cells (ES[2] and ES]4]), induced pluripotent cells derived from human keratinocytes (KiPS4F1 and KiPS4F8), human fibroblasts (HFF) and keratinocytes (HK) (Fig. 12). Here, we show that SMYD2 is the most differentially regulated family member between somatic and pluripotent cells. *SMYD1* is expressed at hardly detectable levels in all cases, but showed higher levels of expression in keratinocytes. Other members of the SMYD family did not show consistent significant differences in expression levels between pluripotent and differentiated cells.

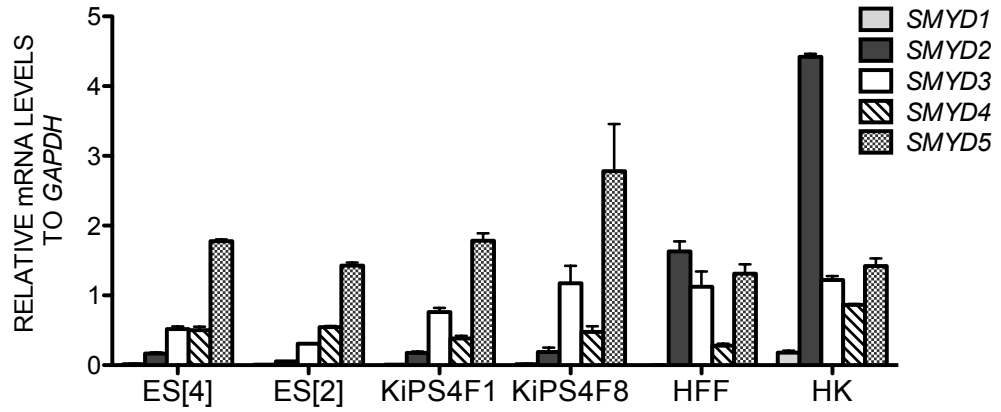


Figure 12. Expression of SMYD family members in different cell types. Comparison of *SMYD1-5* mRNA levels between pluripotent (ES[4], ES[2], KiPS4F1, and KiPS4F8) and differentiated (HFF and HK) cells. mRNA levels were quantified by qPCR and normalized to *GAPDH*. Means and standard deviation from two quantifications are shown.

3.1.2. SMYD2 is induced during the differentiation of human ES cells

Our previous results suggest that *SMYD2* could be induced during the differentiation of human ES cells. To confirm this we performed a time course experiment in which we collected samples at day 0, 4, 8 and 15 of human ES cells differentiation in the form of embryoid bodies (EB). Additionally, we analyzed the expression levels of the rest of the SMYD family members.

Fig. 13a shows the expression profile of *SMYD1-5* during the differentiation of two different human ES cell lines (ES[4] and ES[2]). After 15 days of differentiation

SMYD2 presents a remarkable induction of mRNA levels compared with the rest of family members. *SMYD2* was already induced at day 4 with considerable expression levels in both human ES cell lines. Also *SMYD3* presents a progressive induction during the differentiation process, but less accentuated. *SMYD1* was also induced during differentiation until day 8 and followed by a decrease at day 15. In the other hand, *SMYD4* and *SMYD5* show a slight downregulation during differentiation. Proper downregulation of the pluripotecy markers *OCT4*, *NANOG* and *SOX2* during differentiation is also shown (Fig. 13b). Taken together, these results show that *SMYD2* expression is rapidly induced during human ES cells differentiation to similar levels as found in somatic cells.

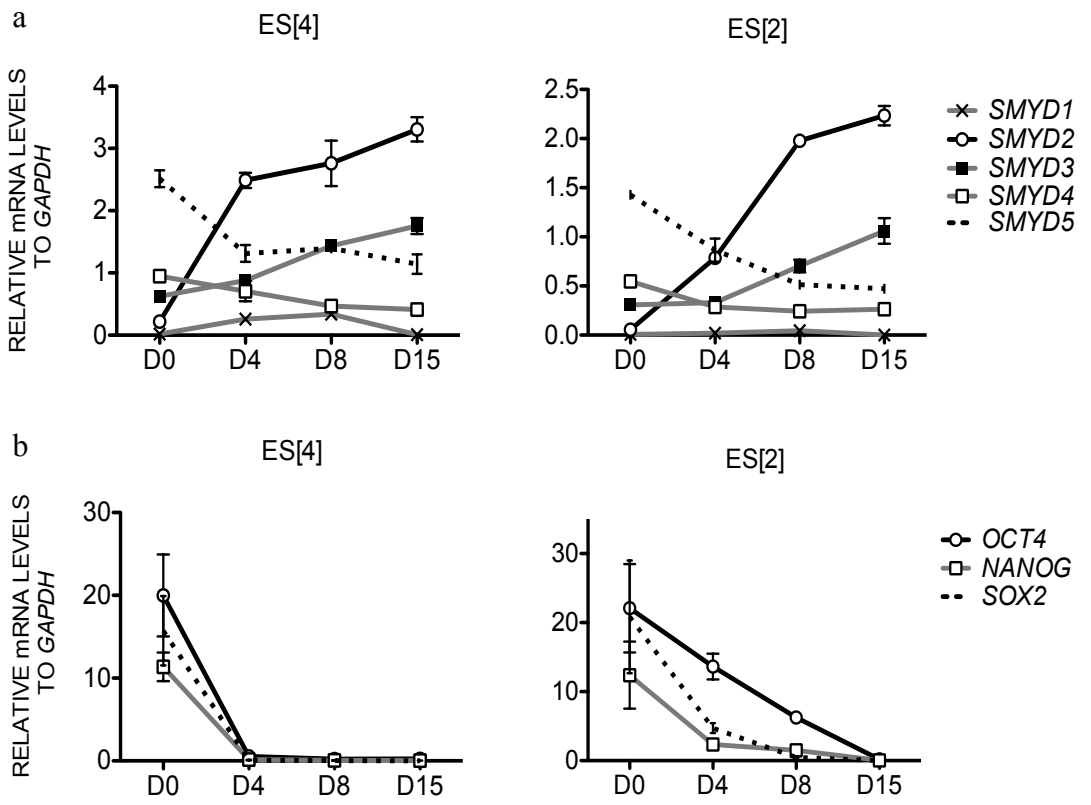


Figure 13. Expression of SMYD family members during human ES cell differentiation. (a) SMYD family members (*SMYD1-5*) and (b) pluripotency genes (*OCT4*, *NANOG* and *SOX2*) mRNA levels were measured in undifferentiated ES[4] and ES[2] cells (D0) and at day 4, 8 and 15 of EB differentiation. mRNA levels were quantified by qPCR and normalized to *GAPDH*. Means and standard deviation from two quantifications are shown.

3.1.3. SMYD2 is marked with H3K4me2/3 in human ES cells

Since *SMYD2* is expressed at low levels in human ES cells but rapidly induced during differentiation we considered the potential presence of bivalent domains on its regulatory regions in pluripotent cells. We performed ChIP assays to detect the presence of H3K4me2/3 and H3K27me3 in ES[4] and ES[2] cell lines (Fig. 14) Precipitated DNA was analyzed by qPCR for the presence of *SMYD2* regulatory regions. As controls, we tested the presence of histone marks at the pluripotency gene *OCT4* and two well described genes containing bivalent domains in human ES cells, *SOX17* and *FOXA2*. Surprisingly, despite being transcribed at low levels, *SMYD2* showed remarkable levels of H3K4me2/3 at its promoter. Levels of H3K27me3 showed variability between lines, being very low in ES[4] but significant in ES[2] while compared with the well known bivalent genes *SOX17* and *FOXA2*. The *OCT4* promoter shows active H3K4me2/3 marks with no repressive H3K27 mark. *SOX17* and *FOXA2* show clear bivalent domains on their promoters with significant signal of both active H3K4me2/3 and repressive H3K27me3 marks.

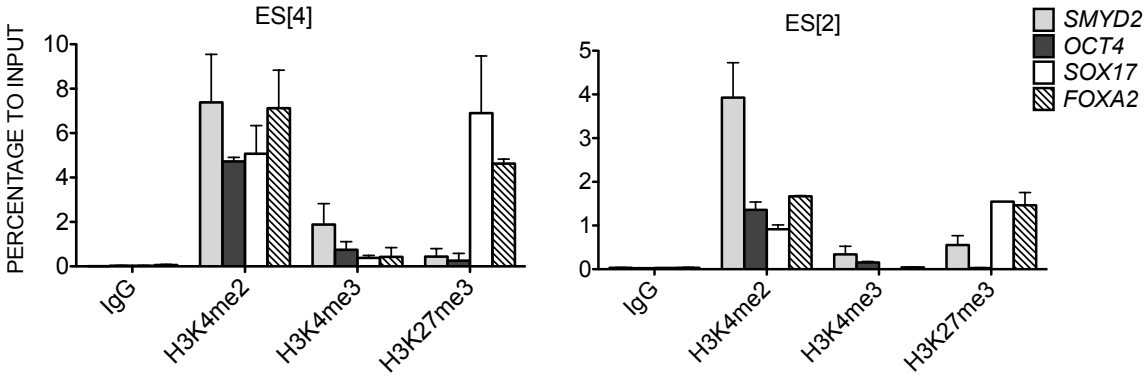


Figure 14. ChIP assay for histone marks on the *SMYD2* promoter. ChIP assay was performed using control IgGs, and antibodies against H3K4me2, H3K4me3 and H3K27me3 in ES[4] and ES[2] cells. Precipitated DNA was analyzed by qPCR for the presence of the regulatory regions of *SMYD2*, *OCT4*, *SOX17* and *FOXA2* genes. Values are represented as percentage to the input. Means and standard deviation from three independent experiments are shown.

In addition, we used publically available data from the ENCODE project to check the presence of histone marks on the *SMYD2* promoter in other lines of human ES cells. We compared the promoter region of two different human ES cell lines (H1 and H7) and human fibroblasts (NHLF) (Fig. 15). Surprisingly, the H1 cell line exhibited bivalent domains on the *SMYD2* promoter, whereas the H7 cell line only showed active

H3K4me3 marks, similar to NHLF cells. As previously shown, these results are in accordance with the differences found on the levels of H3K27me3 between ES[4] and ES[2] at the *SMYD2* promoter.

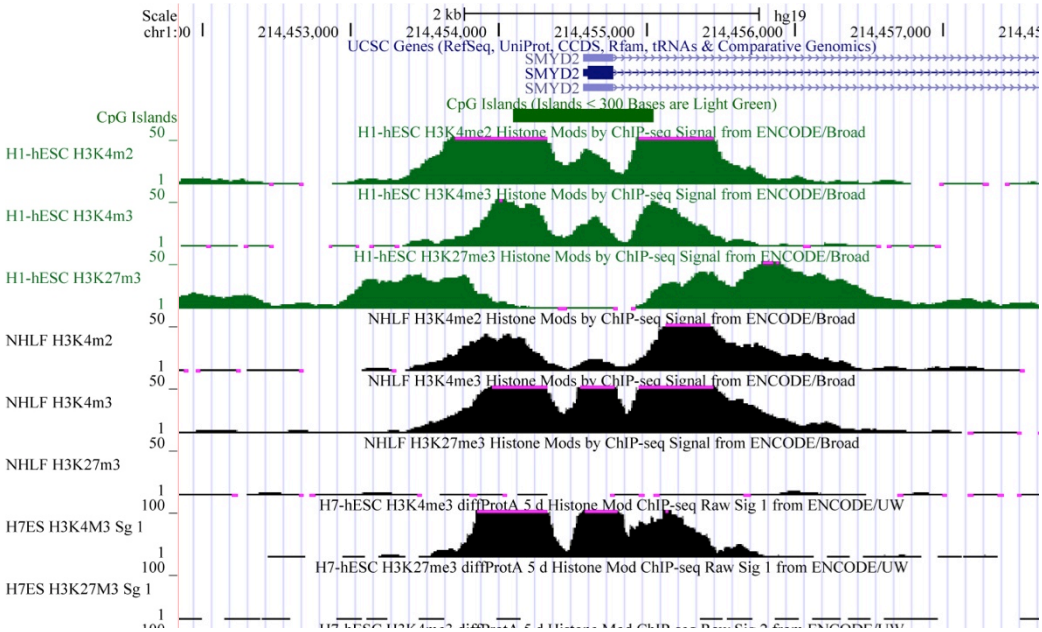


Figure 15. **ENCODE data on histone modifications in the *SMYD2* promoter.** Presence of active (H3K4me2/3) and repressive (H3K27me3) histone marks in the *SMYD2* promoter in human ES cell lines H1 and H7, and human fibroblasts NHLF according to the ENCODE project.

3.1.4. SMYD2 is a potential target of miR-302/367 in human ES cells

The presence of active chromatin marks in the *SMYD2* promoter suggests the existence of other mechanisms responsible for blocking its expression in human ES cells. Recently, it has been published that *SMYD2* is a high-confident target gene of miR-302/367 in human ES cells [214]. The miR-302 cluster is the most abundant microRNA transcript in human ES cells and its levels rapidly decline upon differentiation [215,216], and its overexpression in somatic cells mediate the reprogramming into pluripotent cells [217]. Thus, we suspected that miR-302/367 might target the *SMYD2* mRNA in undifferentiated ES cells for degradation. To determine whether *SMYD2* is regulated by miR-302/367 cluster we compared the expression levels of *SMYD2* in 293T cells overexpressing miR-302/367 (293T-miR-302), and 293T control (293T-GFP) (Fig. 16a). We detected slight but significant lower levels of *SMYD2* mRNA in cells overexpressing the miR-302/367 cluster. To further confirm this finding, we

cloned into the luciferase reporter vector pmirGLO the full length SMYD2 3' UTR region (pmirGLO-SMYD2.UTR) or a shorter form of SMYD2 3'UTR (pmirGLO-SMYD2.UTRshort) that does not contain the reported target region for the miR-302/367 cluster. We also cloned the full length MBD2 3' UTR region (pmirGLO-MBD2.UTR) as a positive control, since it has been recently found that overexpression of miR-302 cluster suppresses MBD2 expression during reprogramming of somatic to iPS cells [218]. Both constructs were transfected into control 293T cells and 293T overexpressing the mir-302/367 cluster and after 48 hours we measured the luciferase activity for each condition (Fig. 16b). Unfortunately, we found no significant changes in luciferase activity between 293T cells with and without overexpression of the miR-302/367 cluster. These results indicate that the miR-302/367 cluster does not target the 3' UTR region of *SMYD2*. However, other regions of the *SMYD2* mRNA might be targeted by this cluster.

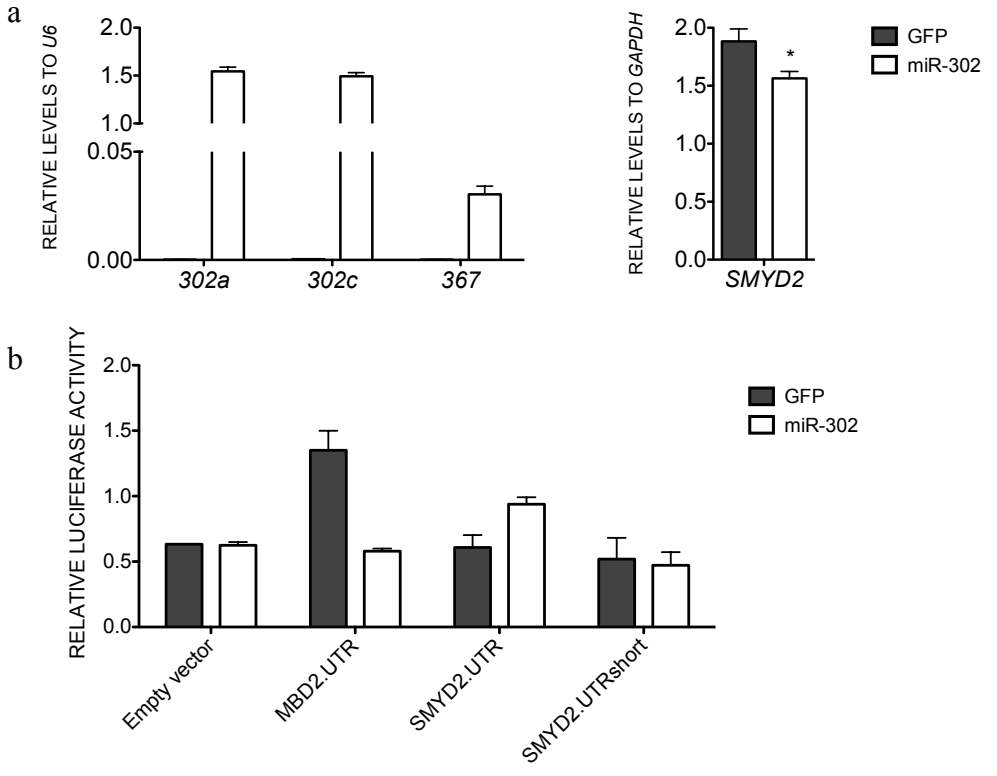


Figure 16. Luciferase assay for *SMYD2* 3'UTR regulatory region. (a) Levels of miR-302a, miR-302c and miR-367, as part of miR-302/367 cluster, normalized to *U6* expression (left panel), and levels of *SMYD2* in 293T-GFP and 293T-miR-302 stable cell lines normalized to *GAPDH* (right panel). Means and standard deviation from three quantifications are shown. (*) $P < 0.05$ compared with 293-GFP cell line. (b) pmirGLO-Empty vector, pmirGLO-SMYD2.UTR, and pmirGLO-SMYD2.UTRshort were transfected into 293T-GFP and 293T-miR-302 stable cell lines. At 48 h post-transfection, cells were harvested and assayed for luciferase activity. Values are normalized to *Renilla* activity. Means and standard deviation of three independent experiments are shown.

3.2. SMYD2 METHYLATES H3K4 AND H4K20 *IN VITRO*

SMYD2 has been described as a histone lysine methyltransferase, but there is controversy regarding its specific catalytic activity on histone tails. Brown *et al.* [119] first identified SMYD2 to methylate H3K36, but years later Abu-Farha *et al.* [122] found that it can also methylate H3K4. To decipher the specific residue methylated by SMYD2 on histone H3 N-terminal tail, we performed *in vitro* methylation assays using as a substrate the H3 N-terminal tail wild type or mutated at specific sites (K4 and K36) fused to GST, in the presence of [H^3] SAM as a methyl-donor. We tested a recombinant octamer reconstituted from independent core histones expressed and purified from *E. coli* (Oct [b]), an octamer purified from HeLa cells (Oct [H]), two independently purified wild type GST-H3 (1 and 2), GST-H3K4R, GST-H3K36R, GST-H3K4R/K36R and also a GST fused to the N-terminal tail of histone H4 (GST-H4) (Fig. 17). Among all different substrates, SMYD2 was able to strongly methylate both recombinant and HeLa core octamers and likely both histones H3 and H4. Wild type GST-H3, GST-H3K36R, and GST-H4, were also methylated, although wild type GST-H3 showed an unexpected weak methylation signal, in both protein preparations. Our results suggest that the main *in vitro* methylation site on histone H3 tail is K4.

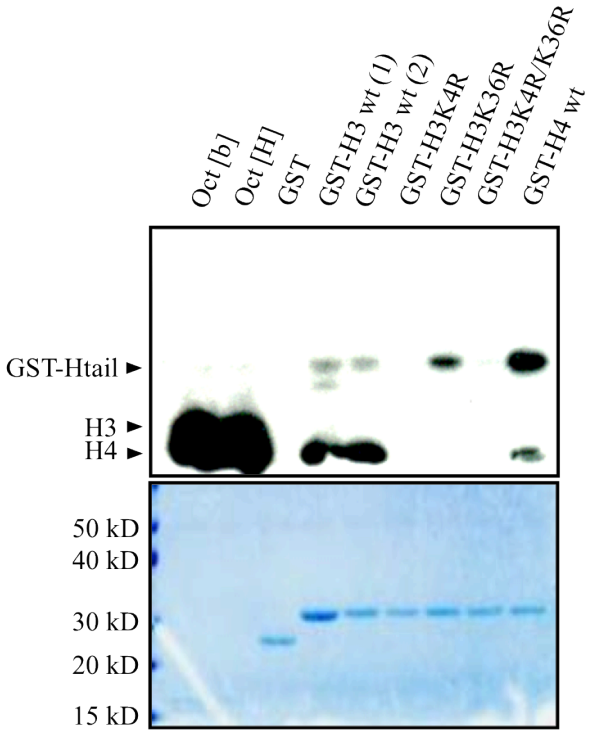


Figure 17. *In vitro* Methyltransferase activity of SMYD2 on histone substrates. *In vitro* methyltransferase assay using recombinant SMYD2 and different substrates including histone octamers (Oct[b] and Oct[H]) and GST tagged histone tails (GST, GST-H3 wt (1), GST-H3 wt (2), GST-H3K4R, GST-H3K36R, GST-H3K4R/K36R and GST-H4 wt). *Coomassie* staining shows the presence and amount of loaded proteins (bottom).

We next planned to address the specific residue of methylation of histone H4 tail. A well-described residue of histone H4 subject to methylation is K20, therefore, we tested the *in vitro* methylation of a GST-H4 tail construct mutated in K20. We also took advantage of several recombinant transcription factors available in our laboratory, including linker histone H1, histone variant macroH2A.1 and the transcription factors RXR, HNF4, AP2 and E2F6 (Fig. 18). Our results show that SMYD2 methylates histone H4 at K20 and that is also able to *in vitro* methylate linker histone H1.

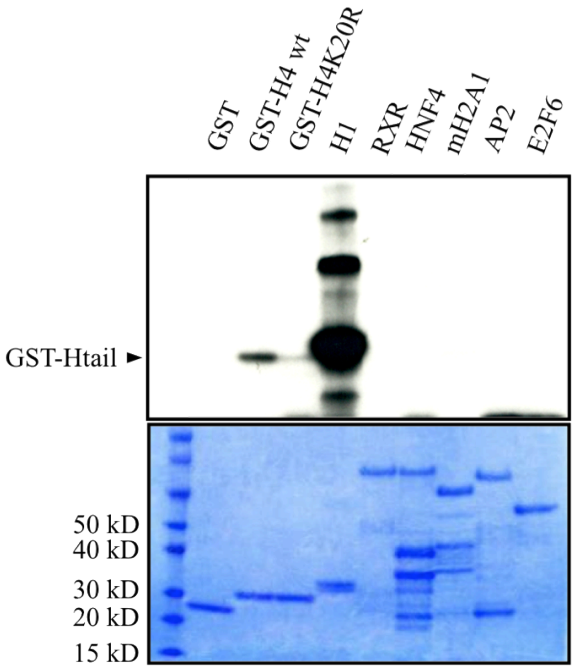


Figure 18. *In vitro* Methyltransferase activity of SMYD2 on histone H4 and other substrates. *In vitro* methyltransferase assay of recombinant SMYD2 with different substrates including GST tagged histone tails (GST, GST-H4 wt and GST-H4K20R) and several recombinant proteins (H1, RXR, HNF4, mH2A1, AP2 and E2F6). *Coomassie* staining shows the presence and amount of loaded proteins (bottom).

3.3. SMYD2 IS INVOLVED IN HUMAN ES DIFFERENTIATION

3.3.1. Knockdown of SMYD2 in human ES cells

To investigate whether SMYD2 could be involved in the differentiation of human ES cells we performed loss-of-function experiments using a lentiviral pLVTHM vector encoding short hairpin RNAs (shRNA) against *SMYD2* and GFP. The knockdown efficiency of six different shRNAs, including a previously published shRNA [124] (shSMYD2), was tested in 293T. Cells were infected with the different viral encoded shRNAs, including a non target shRNA (shControl), and GFP positive cells were sorted and the *SMYD2* mRNA levels were measured by qPCR (Fig. 19). All shRNA sequences

were able to knock down *SMYD2* expression. shSMYD2 was the most efficient shRNA with more than 90% of knockdown efficiency.

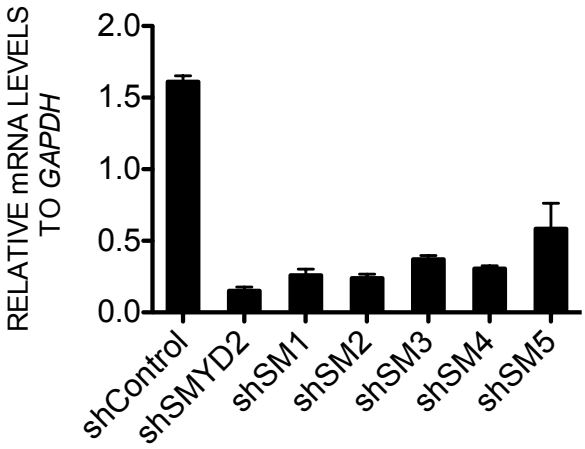


Figure 19. Knockdown efficiency of SMYD2 in 293T cells. *SMYD2* mRNA levels were measured in 293T cell lines with stable expression of different short hairpins against *SMYD2* (shSMYD2, shSM1, shSM2, shSM3, shSM4 and shSM5), and compared to a 293T control line (shControl). Expression levels were quantified by qPCR and normalized to *GAPDH*. Means and standard deviation of two quantifications are shown.

Therefore, we used shSMYD2 shRNA to knock down SMYD2 in ES[4]. We infected ES[4] with the lentiviruses expressing shSMYD2 (ES[4]-shSMYD2) and non-target shRNA (ES[4]-shControl), sorted the GFP positive cells and analyzed the efficiency of the knockdown by qPCR (Fig. 20). Although the levels of endogenous *SMYD2* in ES[4] was very low, we could observe a 67% of knockdown efficiency. In self-renewing cells, the knockdown of *SMYD2* did not show any morphological differences compared to the control, neither differences regarding the expression of selected pluripotency and differentiation genes (data not shown). This result was expected since *SMYD2* is expressed at very low levels in ES[4].

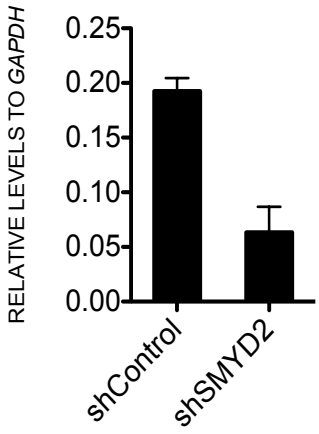


Figure 20. Knockdown efficiency of SMYD2 in ES[4] cells. SMYD2 mRNA levels were measured in ES[4]-shSMYD2 cell line and compared to ES[4]-shControl cell line. Expression levels were quantified by qPCR and normalized to GAPDH. Means of two independent quantifications are shown.

3.3.1.1. The knockdown of SMYD2 promotes the induction of endodermal genes during human ES cells differentiation

To evaluate the potential role of SMYD2 in the differentiation of human ES cells we performed *in vitro* differentiation in the form of EB and presence of 20% FBS of both ES[4]-shSMYD2 and ES[4]-shControl cell lines. We analyzed the mRNA levels of several pluripotent and differentiation markers by qPCR at different time points (day 0, 4, 8 and 15). *OCT4*, *NANOG* and *SOX2* were used as markers of pluripotency; *HNF4*, *FOXA2* and *SOX17* for endoderm; *BRACHYURY* for mesoderm and *PAX6* for ectoderm. mRNA levels of *SMYD2* were upregulated during differentiation with a knockdown efficiency of about 50%. Upon differentiation, the ES[4]-shSMYD2 cell line showed a strong induction of the endodermal genes *HNF4*, *FOXA2* and *SOX17* at day 4 compared with the ES4[4]-shControl line (Fig. 21) but no differences in *PAX6* or *BRACHYURY* expression. The pluripotency-related genes *OCT4*, *NANOG* and *SOX2*

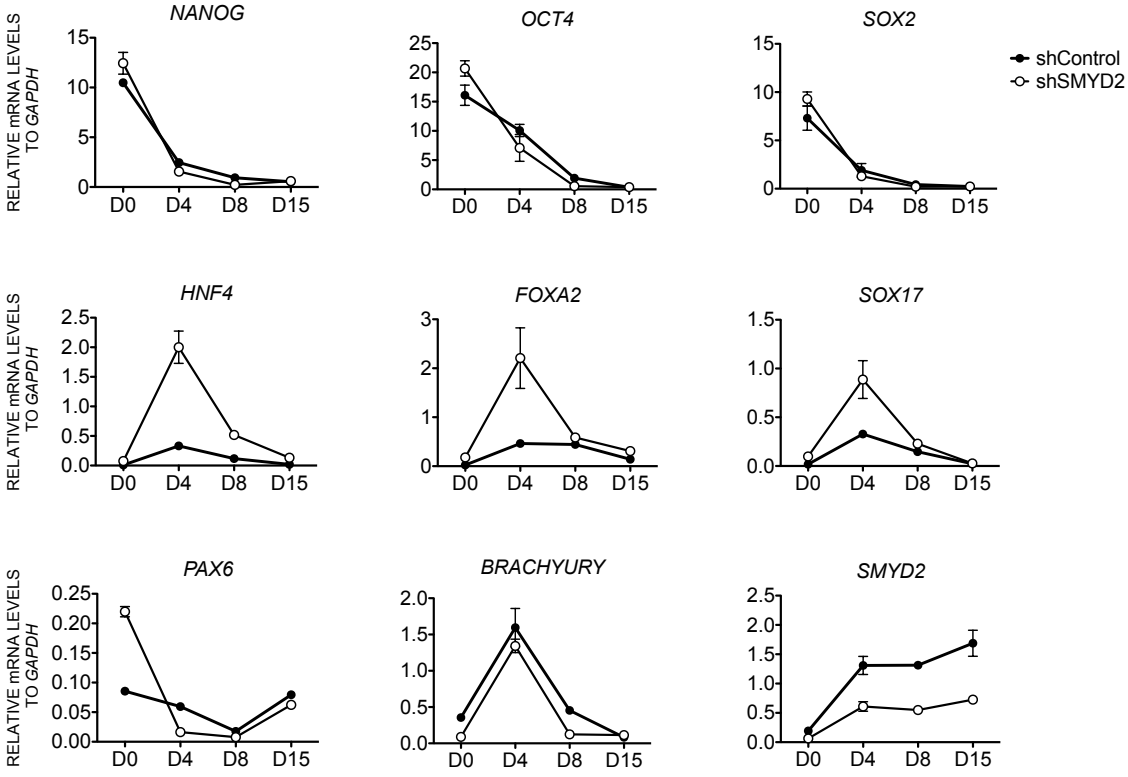


Figure 21. Knockdown of SMYD2 during human ES cell EB differentiation. EB differentiation of ES[4]-shControl and ES[4]-shSMYD2 cell lines. mRNA levels of pluripotency genes (*NANOG*, *OCT4*, *SOX2*), differentiation genes (*HNF4*, *FOXA2*, *SOX17*, *PAX6*, *BRACHYURY*), and *SMYD2* were measured by qPCR at the indicated time points and normalized to *GAPDH*. One representative experiment out of three independent experiments is shown. Mean and standard deviation of two quantifications are shown.

were similarly downregulated in both ES[4]-shSMYD2 and ES[4]-shControl lines. Together, the *SMYD2* knockdown causes an induction of endodermal genes during human ES differentiation, without affecting the silencing of the pluripotency network. Our results suggest that SMYD2 might act as a negative regulator of endodermal differentiation.

3.3.1.2. The knockdown of SMYD2 accelerates the silencing of OCT4 during monolayer differentiation of human ES cells

Next, we reevaluated the potential effects of SMYD2 depletion in the silencing of OCT4 during monolayer differentiation. Compared to EB differentiation, monolayer differentiation occurs much more slowly and might offer additional time windows to observe potential differences between the control and the SMYD2 depleted cell line. ES[4] cell lines were grown to 70% confluency in self-renewal conditions and then media containing 20%FBS was added. The percentage of OCT4 positive cells was measured by FACS, at different days after the addition of differentiation media (Fig. 22) shows that the knockdown of SMYD2 causes a faster reduction of the number of cells positive for OCT4 during monolayer differentiation, compared to the control. These results, together with previous EB differentiation experiments, indicate that the *SMYD2* knockdown stimulates the *in vitro* differentiation of human ES cells.

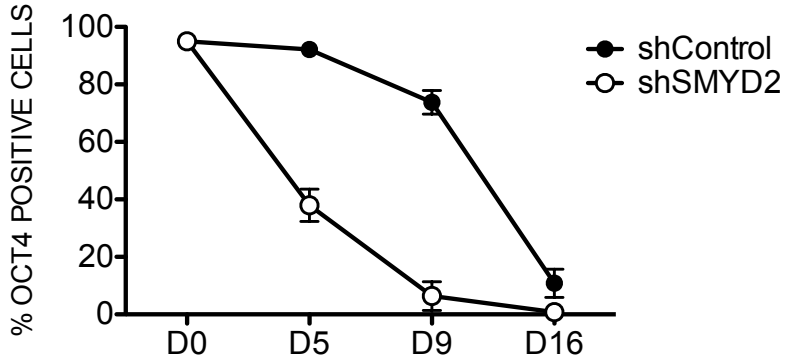


Figure 22. Knockdown of SMYD2 during human ES cell monolayer differentiation. Monolayer differentiation of ES[4]-shControl and ES[4]-shSMYD2 cell lines. The percentage of OCT4 positive cells was determined by FACS analysis at the indicated time points. Mean and standard deviation from three independent experiments are shown.

3.3.2. Overexpression of SMYD2 in human ES cells

To further confirm the involvement of SMYD2 in human ES cells differentiation we also performed SMYD2 overexpression experiments in human ES cells. The use of retroviral and lentiviral vectors in human ES cells present some difficulties for stable transgene expression since they very frequently become silenced after transduction and integration [219]. For that reason we used an expression vector (pTP6) that has been reported to support strong expression levels in ES cells [220]. The presence of overexpressed FLAG-SMYD2 was determined by immunofluorescence after transient transfection of ES[4] cells with pTP6-SMYD2, at undifferentiated state and after 14 days of differentiation, using antibodies against SMYD2 or FLAG (Fig. 23). Consistent with published data in 293T cells [119], SMYD2 is present in both cytoplasm and nucleus of human ES cells.

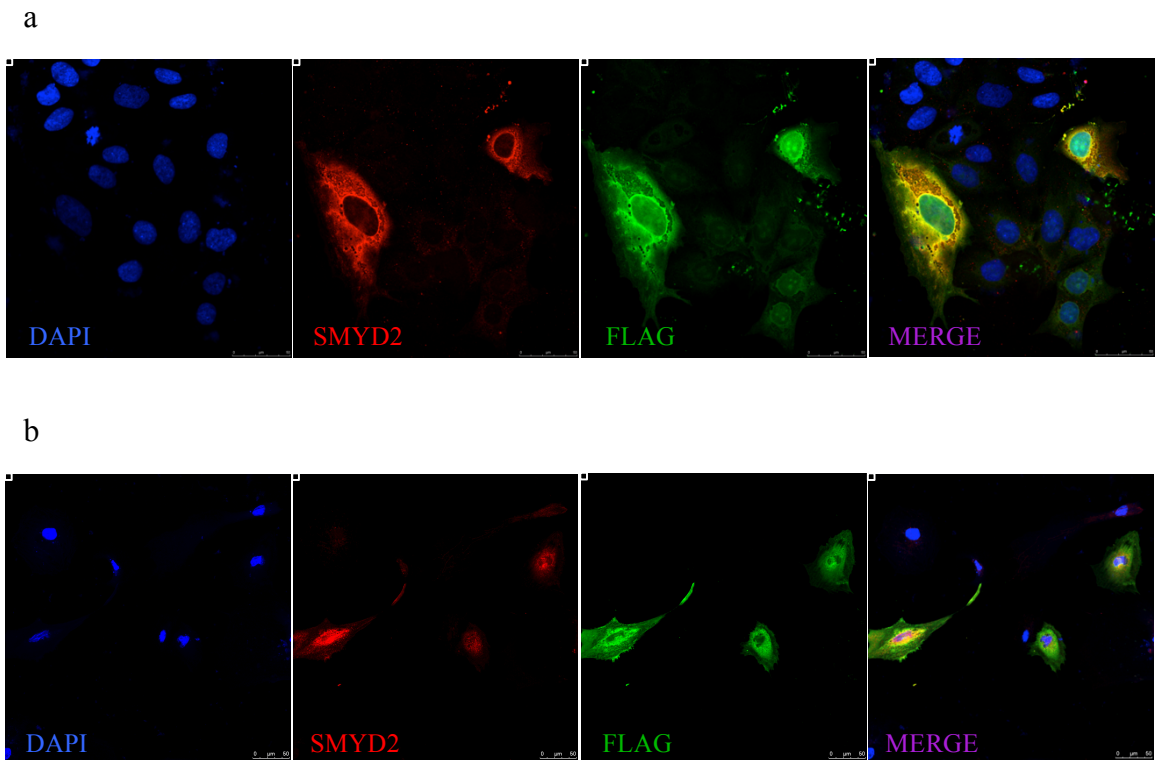


Figure 23. Immunolocalization of SMYD2 in human ES cells. (a) Undifferentiated ES[4] cells and (b) 14 days differentiated ES[4] cells were transiently transfected with pTP6-SMYD2 overexpression vector. Cells were fixed and immunolocalization with anti-SMYD2 (red) or anti-FLAG (green) antibodies was performed.

Next, we generated ES[4] stable cell lines overexpressing FLAG-SMYD2 (ES[4]-SMYD2) or GFP (ES[4]-GFP) as a control, by transfecting pTP6-SMYD2 or pTP6 and selecting with puromycin. We confirmed FLAG-SMYD2 overexpression by qPCR and western blot using antibodies against SMYD2 or FLAG (Fig. 24). Despite the high levels of SMYD2 overexpression, we could not find any phenotypical change compared to control cell lines.

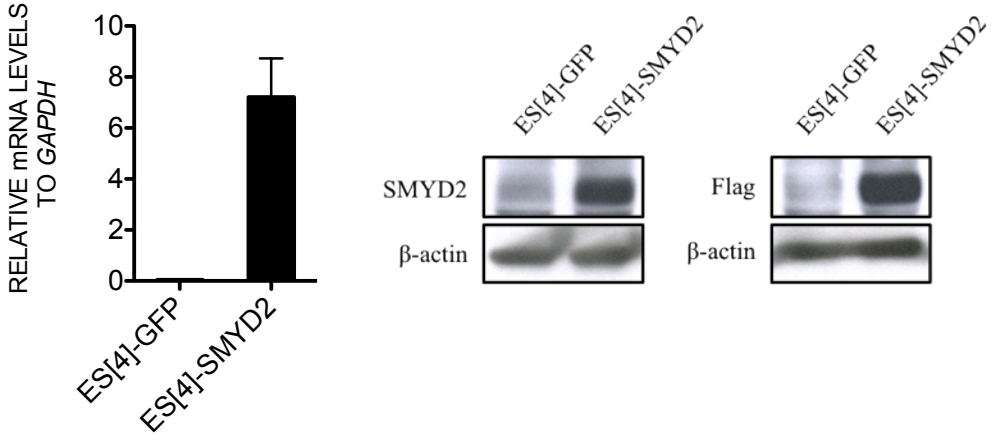


Figure 24. Overexpression of FLAG-SMYD2 in ES[4]. (A) SMYD2 mRNA levels were measured by qPCR in ES[4]-GFP and ES[4]-SMYD2 cell lines and normalized to GAPDH. Mean and standard deviation of two quantifications are shown. (B) SMYD2 protein levels were measured in ES[4]-GFP and ES[4]-SMYD2 cell lines using antibodies against SMYD2 (left) and FLAG (right). β-actin was used as a loading control.

3.3.2.1. The overexpression of SMYD2 impairs the differentiation of human ES cells

To further confirm the involvement of SMYD2 in human ES cell differentiation, we performed EB differentiation of ES[4]-SMYD2 and ES[4]-GFP cell lines. Contrary to the knockdown, cells overexpressing SMYD2 showed a reduced induction of endodermal (*HNF4*, *FOXA2* and *SOX17*) and mesodermal (*BRACHYURY*) genes but no differences in the silencing of pluripotency genes (*OCT4*, *NANOG* and *SOX2*) compared to the control line during differentiation (Fig. 25). However the ectodermal marker *PAX6* was more induced in ES[4]-SMYD2 cells than ES[4]-GFP cells during differentiation.

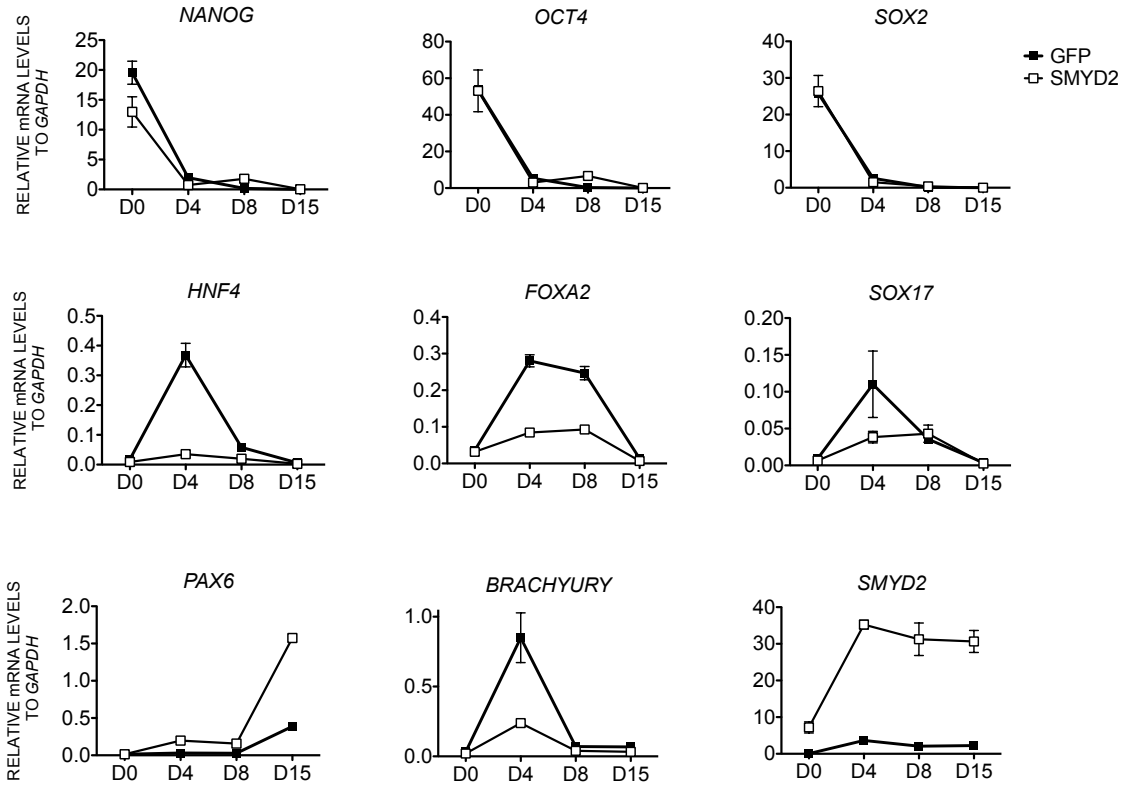


Figure 25. Effects of SMYD2 overexpression in human ES cells EB differentiation. mRNA levels of pluripotency genes (*NANOG*, *OCT4*, *SOX2*), differentiation genes (*HNF4*, *FOXA2*, *SOX17*, *PAX6*, *BRACHYURY*), and *SMYD2* were measured by qPCR in ES[4]-GFP and ES[4]-SMYD2 cell lines at the indicated days of EB differentiation and normalized to *GAPDH*. One representative experiment out of three independent experiments is shown. Mean and standard deviation from two quantifications are shown.

Additionally, ES[4]-SMYD2 cells showed differential morphology at days 12-15 of differentiation, which correlates with their inability to properly differentiate. After 15 days of differentiation, the number of cells in the ES[4]-SMYD2 line was smaller and displayed a more elongated shape and bigger size that resembled the senescent phenotype (Fig. 26).

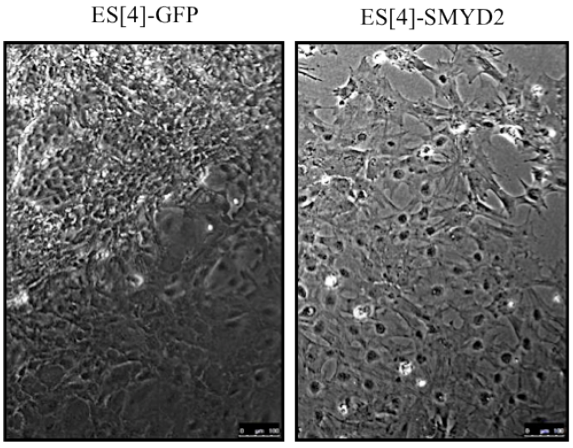
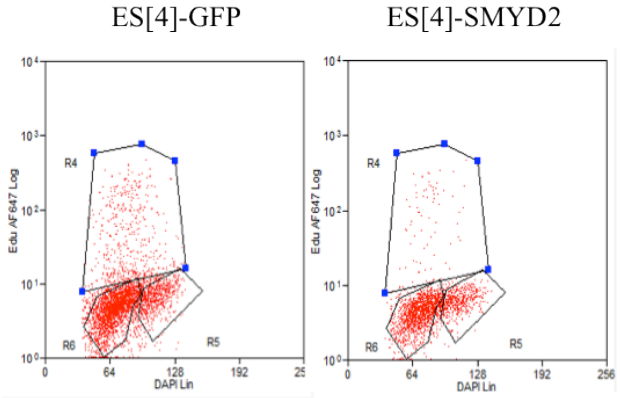


Figure 26. SMYD2 overexpression phenotype during human ES cell EB differentiation. Comparison of ES[4]-GFP and ES[4]-SMYD2 cell line phenotype after 15 days of EB differentiation.

To test if ES[4]-SMYD2 cells had defects in proliferation we performed EdU incorporation assays, to discriminate cells in G_0/G_1, S or G_2/M phases in ES[4]-SMYD2 and ES[4]-GFP cell lines. In accordance with the observed phenotype, at day 12 of differentiation ES[4]-SMYD2 cells showed a lower percentage of cells in S phase, and a larger percentage in G_2/M phase, compared to ES[4]-GFP cells, suggesting that the overexpression of SMYD2 causes cells to slow the cell cycle progression during differentiation (Fig. 27) .



	ES[4]-GFP	ES[4]-SMYD2
G_0/G_1	62,27 %	67,75 %
S	11,27 %	3,38 %
G_2	16,76 %	22,65 %

Figure 27. Cell cycle analysis of SMYD2 overexpression in human ES cells EB differentiation. Cell cycle analysis was performed in ES[4]-GFP and ES[4]-SMYD2 cell lines at day 12 of EB differentiation. DAPI (4', 6-diamidio-2-phenylindole) staining and EdU incorporation were measured. One representative experiment is shown out of two independent experiments.

To gain mechanistic insights into the observed cell cycle effects we measured the mRNA levels of two majors cyclin-dependent kinase inhibitors (CDKIs), p21 and p16, during EB differentiation of ES[4]-SMYD2 and ES[4]-GFP cell lines (Fig. 28). Although no changes were observed for *p21* expression between both lines, *p16* is more strongly induced in ES[4]-SMYD2 cell line, specially at day 15 of differentiation. Thus, the reduced proliferation and changes in morphology observed in ES[4]-SMYD2 cell line may be caused by p16 induction during EB differentiation, resulting in a cell cycle arrest.

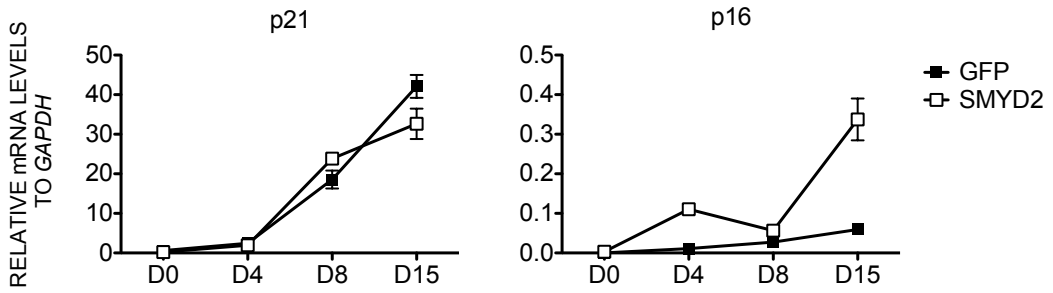


Figure 28. CDKIs levels during overexpression of SMYD2 in human ES cell EB differentiation. As shown in Fig. 15, mRNA levels of *p21* and *p16* were measured by qPCR at the indicated days of EB differentiation of ES[4]-GFP and ES[4]-SMYD2 cell lines and normalized to *GAPDH*. One representative experiment out of three independent experiments is shown. Mean and standard deviation from two quantifications are shown.

3.3.2.2. The overexpression of SMYD2 does not affect the neural differentiation of human ES cells

We have shown that the overexpression of SMYD2 in human ES cells causes defects in the induction of the expression of differentiation genes during EB differentiation, except for the ectodermal marker PAX6. Therefore, we hypothesized that SMYD2 overexpression may cause ES[4] cells to preferentially differentiate into ectodermal lineage during EB differentiation. To validate our hypothesis, we performed *in vitro* neural directed differentiation experiments based on the dual SMAD signalling inhibition protocol [221]. We collected several time points during the differentiation of for ES[4]-SMYD2 and ES[4]-GFP cell lines and analyzed the mRNA levels of *PAX6*, *OCT4* and *SMYD2* (Fig. 29). The endogenous expression of SMYD2 was induced during upon neural commitment in ES[4]-GFP cells, reaching similar mRNA expression levels to the differentiated cell types analyzed in Fig. 11. However, SMYD2 overexpression did not further promote the induction of *PAX6*, compared to control cells.

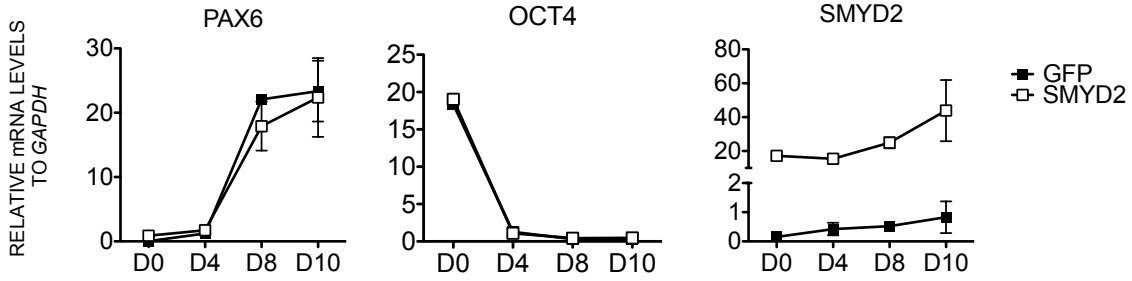


Figure 29. Overexpression of SMYD2 during the differentiation of human ES cells to neural precursors. ES[4]-GFP and ES[4]-SMYD2 cell lines were differentiated to neural precursors following the dual SMAD signaling inhibition protocol and mRNA levels of *OCT4*, *PAX6*, and *SMYD2* were measured by qPCR at the indicated days of differentiation and normalized to *GAPDH*. One representative experiment out of two independent experiments is shown. Mean and standard deviation from two quantifications are shown.

We additionally evaluated the expression of PAX6 and SMYD2 by immunohistochemistry at day 10 of differentiation (Fig. 30). Both cell lines showed a strong percentage of cells positive for PAX6, whereas ES[4]-SMYD2 cells were also positive for SMYD2, and ES[4]-GFP cells were positive for GFP. Taken together, the overexpression of SMYD2 has no effect on the efficiency of neural lineage differentiation using the dual SMAD signalling inhibition protocol.

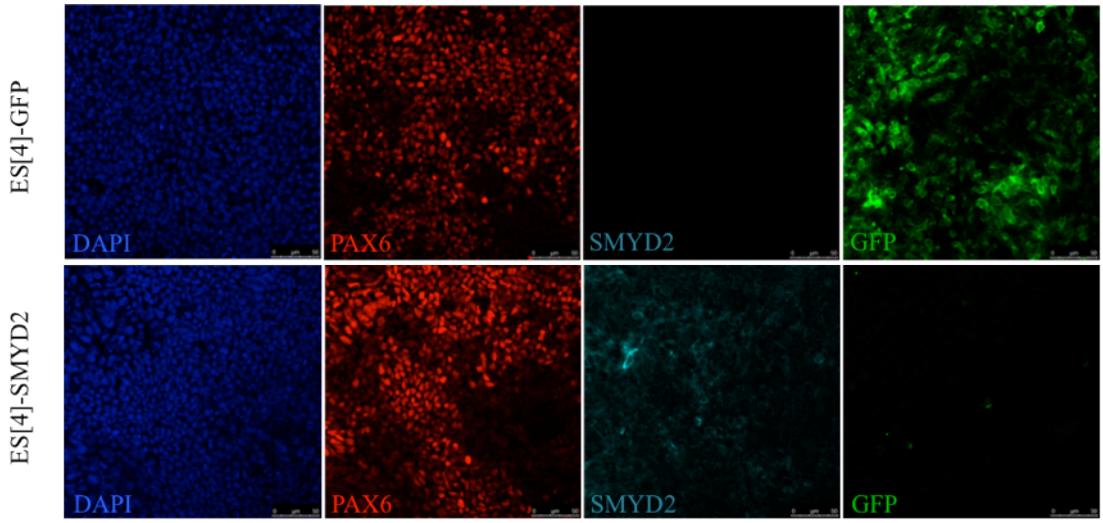


Figure 30. Immunolocalization of PAX6 in human ES cell lines differentiated to neural precursors. ES[4]-GFP and ES[4]-SMYD2 cell lines were fixed and analyzed at day 10 of neural differentiation. Images show the immunolocalization of SMYD2 (blue), the neural marker PAX6 (red) and GFP signal (green) and are representative of two independent differentiation experiments.

3.4. SMYD2 IS INVOLVED IN ZEBRAFISH DEVELOPMENT

3.4.1. *smyd2a* is induced during zebrafish gastrulation

The effects of SMYD2 knockdown and overexpression in human ES cells differentiation suggested that SMYD2 could be involved in regulating early development. To assess the potential involvement of SMYD2 in early stages of development we performed knockdown experiments in zebrafish. The zebrafish

presents two ortholog genes to human *SMYD2*: *smyd2a* and *smyd2b*. *SMYD2* shares a 75% identity and 90% similarity with *smyd2a*, and a 65% identity and less that 90% similarity with *smyd2b*. Moreover, they all present a similar exon-intron structure composed of 12 exons with the SET domain located between exons 1-8 [134].

We first analyzed the mRNA expression profile of *smyd2a* and *smyd2b* genes during zebrafish development from 0.2 to 48 hours post-fertilization (hpf) (Fig. 31). Interestingly, *smyd2a* was maternally expressed, rapidly degraded after fertilization and induced again during gastrulation (from 5 to 10 hpf). *smyd2b* expression levels remained low until 10 hpf and was dramatically induced after gastrulation. The induction of *smyd2a* during gastrulation suggests a potential role for this gene in germ layer specification.

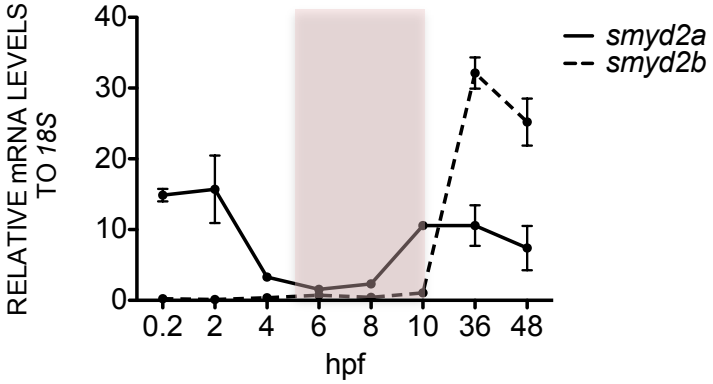


Figure 31. Expression of *smyd2a* and *smyd2b* during the first 48 hours of zebrafish development. 30 embryos were collected at the indicated hours post fertilization (hpf) and mRNA levels were measured by qPCR and normalized to *18S*. Coloured area represents the gastrulation stage, from 5 to 10 hpf. One representative experiment out of two is shown. Mean and standard deviation from two quantifications are shown.

3.4.2. The knockdown of *smyd2a* in zebrafish causes tail formation defects

Because *smyd2a* is induced during gastrulation, it is the most likely gene to play homologous functions to human SMYD2. In order to knock down the expression of *smyd2a* we designed a splice-blocking morpholino (smyd2a-MO) at the exon 1-intron 1 junction, to impair proper splicing of the zygotic transcripts without affecting the maternal mRNA. It has been described that targeting a splice junction of the first exon-intron generates a first intron inclusion [222]. In that sense, smyd2a-MO will block

smyd2a pre-mRNA splicing generating an intron 1 insertion, introducing several stop sequences on the reading frame.

In order to find the optimal morpholino concentration to observe potential effects, we injected three different concentrations (0.50 mM, 0.35 mM and 0.25 mM), and evaluated the resulting phenotype and mortality (Fig. 32). smyd2a-MO injection at 0.50 mM caused 70-80% of mortality and all surviving embryos exhibited a strong delay of development at 24 hpf. Mortality reached 100% at 48 hpf. Importantly, Control-MO injection at 0.50mM had only marginal effects on mortality or phenotype, suggesting that the effects caused by smyd2a-MO injection were specific. Injections of smyd2a-MO at 0.35 mM caused 50-55% of mortality and 70% of surviving embryos showed a delay in development and tail defects but were alive at 48 hpf. Injections at 0.25 mM caused no significant differences compared with Control-MO. Therefore, the 0.35 mM morpholino concentration seemed optimal to study the potential effects of smyd2a depletion during zebrafish development.

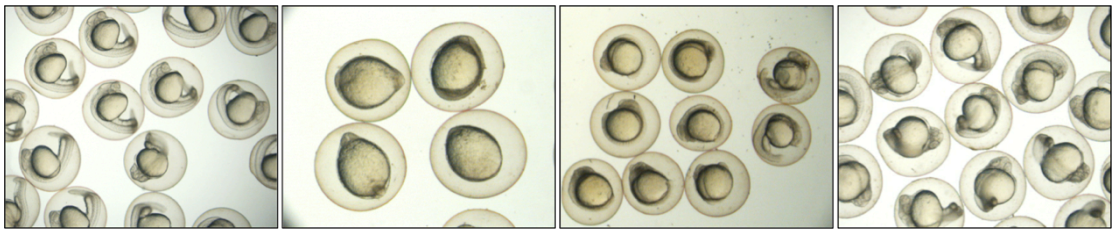


Figure 32. *smyd2a* morphant phenotypes corresponding to different doses of smyd2a-MO at 24 hpf. Zebrafish embryos were injected at 1-2 cell stage with smyd2a-MO and Control-MO at 0.50 mM, 0.35 mM and 0.25 mM. For the control 0.50mM injected embryos are shown.

After choosing the proper concentration of smyd2a-MO injection, we evaluated the *smyd2a* knockdown. We injected smyd2a-MO and Control-MO and measured the mRNA levels of *smyd2a* by RT-PCR at 24 hpf. As expected, the mature mRNA of *smyd2a* was undetectable in embryos injected with smyd2a-MO at 24 hpf compared to embryos injected with Control-MO, while *smyd2b* expression was unaffected by the injection of smyd2a-MO (Fig. 33). The resulting unspliced *smyd2a* fragment of 4473 bp generated in smyd2a-MO embryos was not detected probably due to its long size. In any case, we achieved a very efficient knockdown of *smyd2a* in zebrafish embryos.

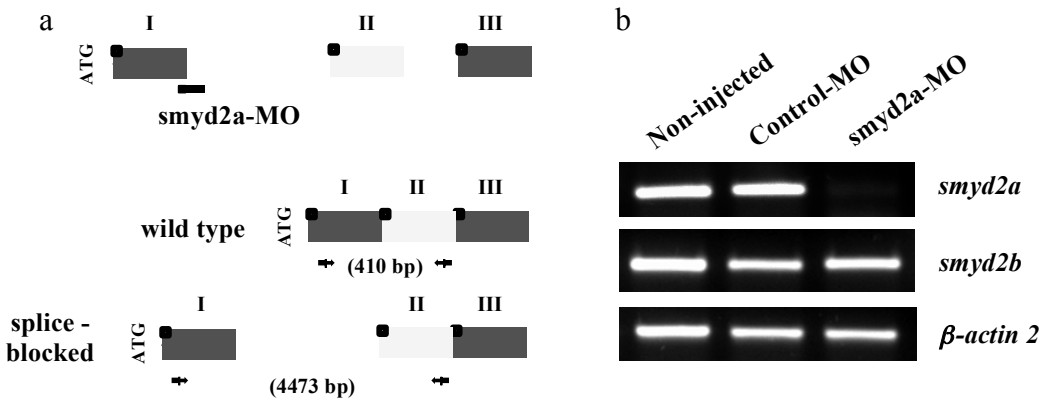


Figure 33. *smyd2a* knockdown efficiency of smyd2a-MO at 24 hpf. (a) Schematic representation of the splice-blocking morpholino smyd2a-MO at the exon1-intron 1 junction. Specific primers for *smyd2a* were designed at each side of exon 1 and intron 1 flanking a region of 410 bp in wild type mRNA, and 4473 bp in unspliced mRNA. (b) Absence of mature *smyd2a* mRNA in smyd2a-MO 0.35 mM compared to Control-MO 0.35 mM and non-injected embryos. 50 embryos were collected for each injection and mRNA was measured by RT-PCR using specific primers for *smyd2a* shown in left panel. Expression of *smyd2b* was unaffected by the smyd2a-MO injection. *β-actin 2* was used as control.

We injected zebrafish embryos again with 0.35 mM smyd2a-MO and Control-MO to evaluate the morphant phenotype at different time points during development. Injected embryos were examined morphologically up to 6 days post-fertilization (dpf) (Fig. 34). At 5 hpf, smyd2a-MO injected embryos could not be morphologically distinguished from the controls. At 24 hpf, we could detect some *smyd2a* morphant embryos with a tail defect (mild), and others with a strong delay in development (severe). At 48 hpf, we could observe several degrees of tail defects, including a morphant phenotype with a complete absence of the tail (very severe). Later on, at 6 dpf, all three *smyd2a* morphant phenotypes were distinguishable and still alive. The mild and severe morphants were able to swim with some difficulties while the very severe morphants showed very reduced motility but beating-heart activity. Our results show that the depletion of *smyd2a* results in a developmental delay and aberrant tail formation during fish development.

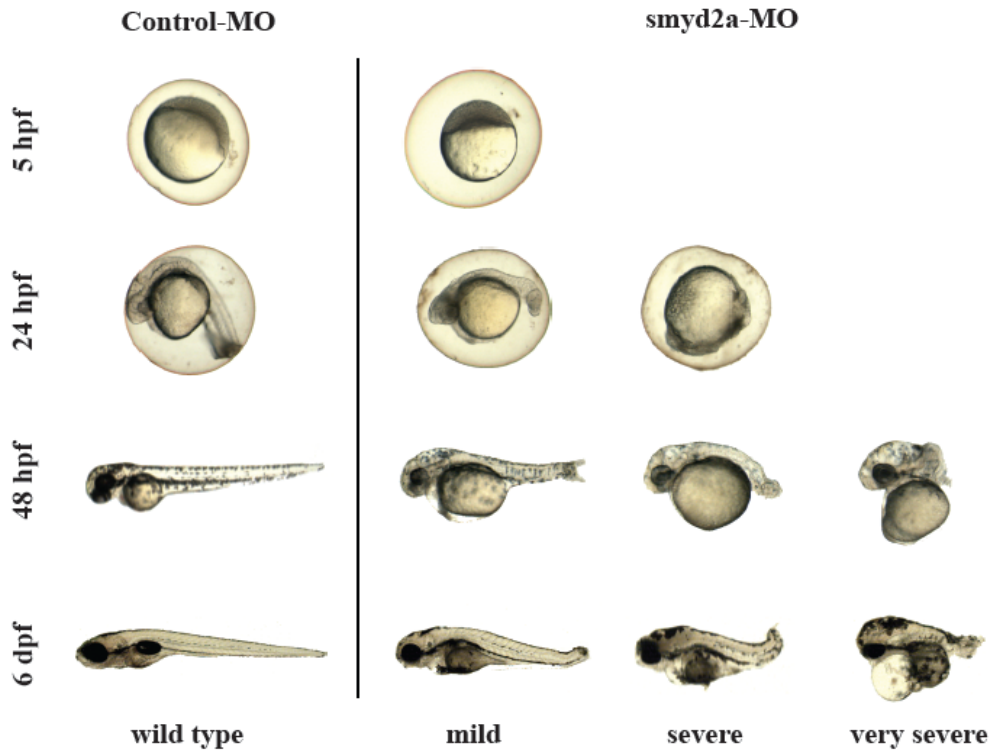


Figure 34. Different phenotypes of *smyd2a* morphant embryos during zebrafish development. Zebrafish embryos were injected at 1-2 cell stage with Control-MO (0.35 mM) and smyd2a-MO (0.35 mM). Images were taken at 5, 24, 48 hpf and 6 dpf. Different smyd2a-MO phenotypes were classified as mild, severe and very severe, corresponding to the observed tail defects.

3.4.3. The *smyd2a* knockdown phenotype in zebrafish can be rescued overexpressing human SMYD2

To further confirm the specificity of the smyd2a-MO, we performed rescue experiments by co-injecting with *in vitro*-transcribed human SMYD2 mRNA (smyd2a-MO + SMYD2 mRNA). In order to find the optimal amount of SMYD2 mRNA that rescues the smyd2a-MO phenotype, we performed co-injection of 0.35 mM smyd2-MO with different amounts of SMYD2 mRNA (Fig. 35). At 24 hpf, co-injection with 30pg SMYD2 mRNA showed the most significant rescue of morphant embryos, compared to 57% obtained with the smyd2a-MO alone. Co-injection of higher amounts of SMYD2 mRNA (200 pg and 300 pg) rendered no surviving embryos at 24 hpf. Next, we tested the effect of SMYD2 mRNA injection alone. Even when injecting 300 pg of SMYD2 mRNA alone, no significant differences on zebrafish embryos morphology or mortality

were detected. This indicates that the lethal effects observed at high doses of SMYD2 mRNA was only associated to its co-injection with smyd2a-MO.

	n.	Dead (%)	No. Surviving	wild type (%)
smyd2a MO	107	47	57	25
smyd2a MO + 15 pg SMYD2 RNA	99	47	52	28
smyd2a MO + 30 pg SMYD2 RNA	172	35	112	60
smyd2a MO + 100 pg SMYD2 RNA	176	59	73	0
smyd2a MO + 200 pg SMYD2 RNA	171	all dead	0	-
smyd2a MO + 300 pg SMYD2 RNA	161	all dead	0	-
30 pg SMYD2 mRNA	35	3	34	97
100 pg SMYD2 mRNA	78	0	78	100
300 pg SMYD2 mRNA	86	5	82	96

Figure 35. Percentage of mortality and rescue effect of SMYD2 mRNA co-injected with smyd2a-MO at 24 hpf. Zebrafish embryos were injected with different combinations of smyd2a-MO morpholino (0.35 mM) and SMYD2 RNA (15, 30, 100, 200 and 300 pg). Survival rates and percentages of wild type embryos are shown.

Next we quantified the effects of smyd2a-MO in zebrafish embryos and the rescue capacity of the SMYD2 mRNA at 24 hpf. For this experiment we injected 0.35 mM of morpholino and 30 ng of SMYD2 mRNA (Fig. 36). The smyd2a-MO injection resulted in about 50% of embryo death and 85% of embryos with mild or severe tail defects. Co-injection with SMYD2 mRNA rescued mortality by 36% and phenotype by 29% suggesting that the effects caused by the smyd2a-MO injection are specific.

	n.	Dead (%)	No. Surviving	wild type (%)	mild (%)	severe (%)
A) Control-MO	339	9 ± 3	307	94 ± 3	6 ± 3	0 ± 0
B) smyd2a-MO	545	57 ± 3	238	15 ± 7	25 ± 5	60 ± 12
C) smyd2a-MO + SMYD2 RNA	665	36 ± 9	436	71 ± 16	11 ± 4	18 ± 16
D) Control-MO + SMYD2 RNA	512	7 ± 2	475	97 ± 3	3 ± 2	0 ± 0

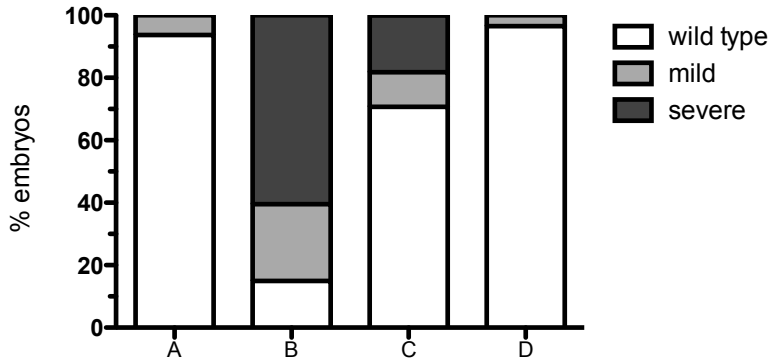


Figure 36. Percentage of phenotypes of morpholino injection and rescue experiment at 24 hpf. Zebrafish embryos were injected with four different combinations of morpholino (0.35 mM) and SMYD2 RNA (30 pg): A) Control-MO; B) smyd2a-MO; C) smyd2a-MO + SMYD2 RNA; and D) Control-MO + SMYD2 RNA. Upper panel shows the survival rates and percentages of wild type, mild and severe phenotypes. Mean and standard deviation from three independent experiments are shown. Lower panel represents the percentages of phenotypes shown in upper panel.

3.4.4. Gene expression analysis of *smyd2a* knockdown in zebrafish embryos

Once we made sure that the effects of *smyd2a* knockdown in zebrafish embryos were consistent and reproducible, we decided to identify possible changes in gene expression during gastrulation stage in smyd2a-MO injected embryos. To gain insights into the potential pathways affected by the *smyd2a* knockdown we investigated the expression pattern of the T-box gene *no tail* (*ntl*), a zebrafish Brachyury ortholog required for notochord formation and essential for tail development [223,224]. Interestingly, *ntl* mutants embryos exhibited a phenotype with tail defects, similar to smyd2a-MO injected embryos [225]. To determine the expression pattern of *ntl* gene on *smyd2a* morphants, we performed whole-mount *in situ* hybridization (ISH) assays at 6 hpf using a specific probe for *ntl* gene (Fig. 37). Because at this stage we could not yet differentiate the *smyd2a* morphant phenotype, we collected a pool of embryos for each condition: 8 embryos for the Control-MO, and 17 for the smyd2a-MO. As expected, all control embryos showed *ntl* expression at the margin cells as gastrulation proceeds while most s*myd2a* morphants showed a reduction or complete absence of *ntl* expression. Interestingly, the percentage of the three different *ntl* expression patters detected in smyd2a morphants is similar to the percentage of different phenotypes at 24 hpf. Therefore, the lack of *ntl* expression in *smyd2a*-depleted embryos may explain the problems of tail formation.

Next we analyzed different marker genes that have been described to play a role in gastrulation, such as *gata2* and *sox17*. We collected 14 embryos for Control-MO, and 20 for smyd2a-MO at 6 hpf and analyzed the expression of each gene (Fig. 37). *Gata2* was expressed in all control embryos and exhibited a ventral expression, but only 25% of smyd2a-MO embryos presented reduced expression. *Sox17* was expressed on the

margin in all control embryos, while about 50% of smyd2-MO-injected embryos showed abnormal distribution through the animal pole.

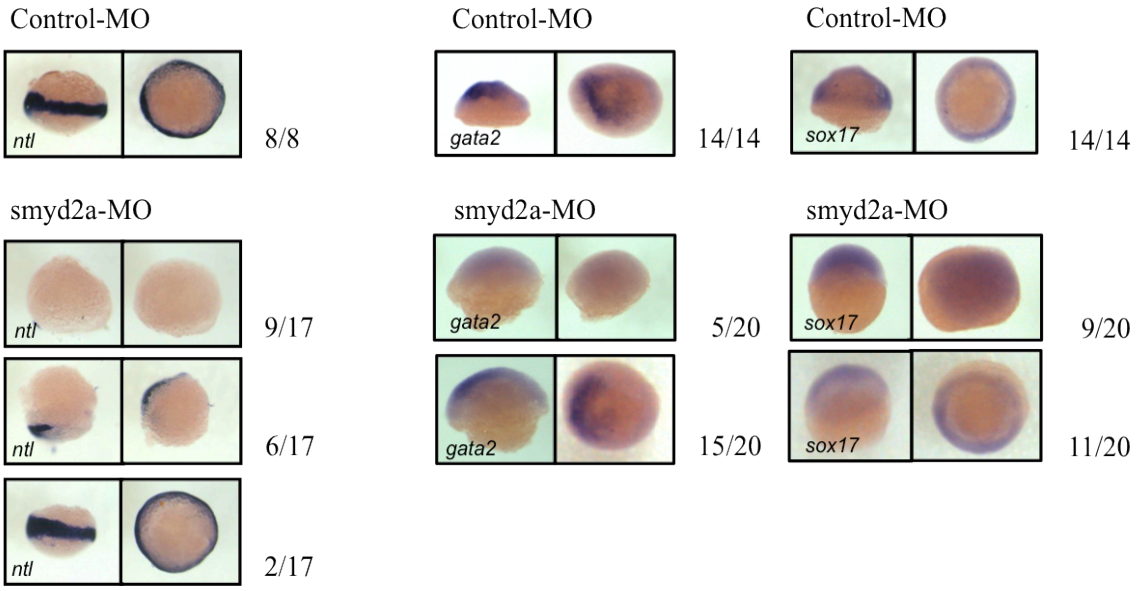


Figure 37. *In situ* hybridization of *ntl*, *gata2* and *sox17* in control and *smyd2a* morphant embryos. Zebrafish embryos were injected at 1-2 cell stage with Control-MO (35 mM) and smyd2a-MO (35 mM) and analyzed for the expression of the indicated genes by in situ hybridization at 6hpf. Embryos are shown in lateral views (left), and animal pole views (right).

In order to identify potential defects in muscle development in smyd2a-MO injected embryos we next checked the expression of the early myogenic marker *myoD*. Since the smyd2a-MO-injected embryos showed a delay in development we compared the expression of *myoD* in smyd2a-MO-injected embryos at 48 hpf to control-MO-injected embryos at 24 hpf, when the tail structure and somites have been already formed. We selected two smyd2a-MO embryos with different tail defects and one Control-MO embryo as a control (Fig. 38). In accordance with published data [226], Control-MO embryos showed *myoD* expression in lateral cells of all somites along the trunk and the tail. In contrast, some smyd2a-MO-injected embryos showed *myoD* expression but it reflected a disorganized structure of somites in the tail, whereas some others exhibited a weak *myoD* expression. In addition, all smyd2a-MO embryos presented *myoD* expression on the tail somites, but not in the trunk somites.

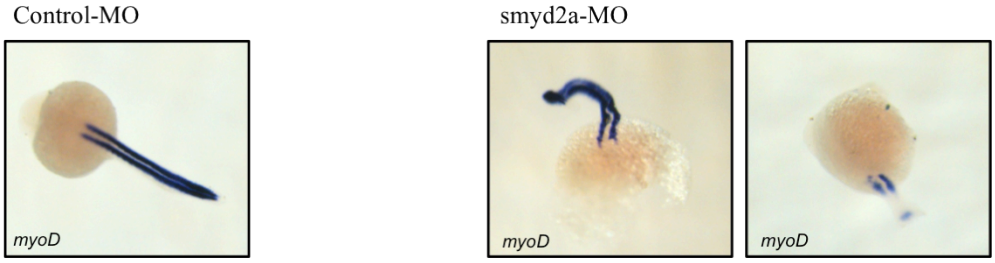


Figure 38. *In situ* hybridization of *myoD* in control and smyd2a morphant embryos. Zebrafish embryos were injected at 1-2 cell stage with Control-MO (35 mM) and smyd2a-MO (35 mM). Different expression of *myoD* in one representative embryo for Control-MO at 24 hpf (left), and two different representative embryos for smyd2a-MO at 48 hpf (right).

To further identify specific transcripts affected by *smyd2a* knockdown, we analyzed the expression profile of a set of genes involved in gastrulation from 4 hpf to 10 hpf by qPCR (Fig. 39). For this analysis we selected the Nodal target genes *chordin* (*chd*), *goosecoid* (*gsc*), *bonnie and clyde*/*Mixer* (*bon*), and *casanova*/*sox32* (*cas*); the ventral markers *gata2* and *bmp2a*; and the Wnt pathway genes *β-catenin-1* and *dkk1b*. All four Nodal target genes (*chd*, *gsc*, *bon*, *cas*) were up-regulated in smyd2a-MO-injected embryos compared to control embryos. In addition to their strong induction, *gsc* and *bon* also showed a different profile of expression with at peak of expression at 8 hpf, while the control-MO-injected embryos showed already reduced expression at this time point. Compared to Control-MO-injected embryos, the ventral marker *gata2* was less

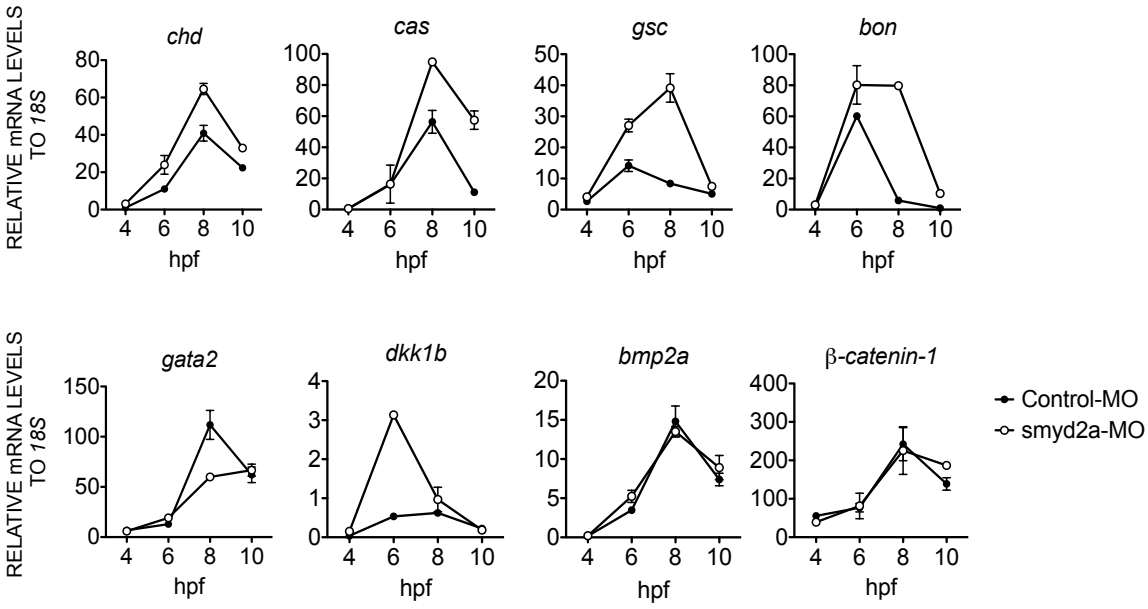


Figure 39. Changes in gene expression during gastrulation in *smyd2a* morphant and control embryos. Zebrafish embryos were injected at 1-2 cell stage with Control-MO (0.35 mM) and smyd2a-MO (0.35 mM) and 50 embryos were collected at each time point for mRNA extraction. mRNA levels of Nodal related genes (*chd*, *cas*, *gsc* and *bon*), ventral genes (*gata2* and *bmp2a*) and Wnt genes (*dkk1b* and *β-catenin-1*) were measured by qPCR and normalized to *18S*. Mean and standard deviation of triplicate quantifications are shown.

expressed in smyd2a-MO-injected embryos at 8hpf. Moreover, *dkk1b*, a negative regulator of Wnt signaling, was highly induced at 6 hpf in smyd2a-MO-injected embryos. *Bmp2a* and *β-catenin-1* are not affected by *smyd2a* knockdown, suggesting that genes of the Bmp2 and Wnt pathways remain unaffected. These results indicate that *smyd2a* depletion causes an induction of Nodal-regulated genes during gastrulation events in zebrafish. This effect is similar to the induction of endodermal genes in human ES cells differentiation caused by SMYD2 knockdown.

3.5. SMYD2 PROMOTES BMP SIGNALING IN HUMAN ES CELLS

3.5.1. SMYD2 enhances the expression of BMP2 target genes in human ES cells

The results obtained in the zebrafish model support the idea that SMYD2 has an early effect during development, possibly during gastrulation, as suggested also by our previous experiments of knockdown and overexpression in human ES cells. It has been described that BMPs play an essential role during tail formation in zebrafish [204]. In addition, tail defects observed in zebrafish mutant embryos are often related with disrupted members of the BMP pathway [227]. Therefore, we addressed the potential mechanistic connection between SMYD2 and BMP signaling.

To address the potential involvement of SMYD2 in the BMP signaling, we asked whether SMYD2 would be able to promote the induction of BMP target genes in human ES cells. We treated ES[4]-SMYD2 and ES[4]-GFP cell lines with BMP2 during four days and measured the expression of BMP2 target genes by qPCR (Fig. 40). These genes were selected based on previous transcriptomic analysis carried out in our laboratory after BMP2 treatment of human ES cells [158]. After BMP2 treatment, most BMP2 target genes were more strongly induced in the presence of overexpressed

SMYD2, compared to the control. Accordingly, OCT4 was more strongly downregulated in the ES[4]-SMYD2 cell line than in the control.

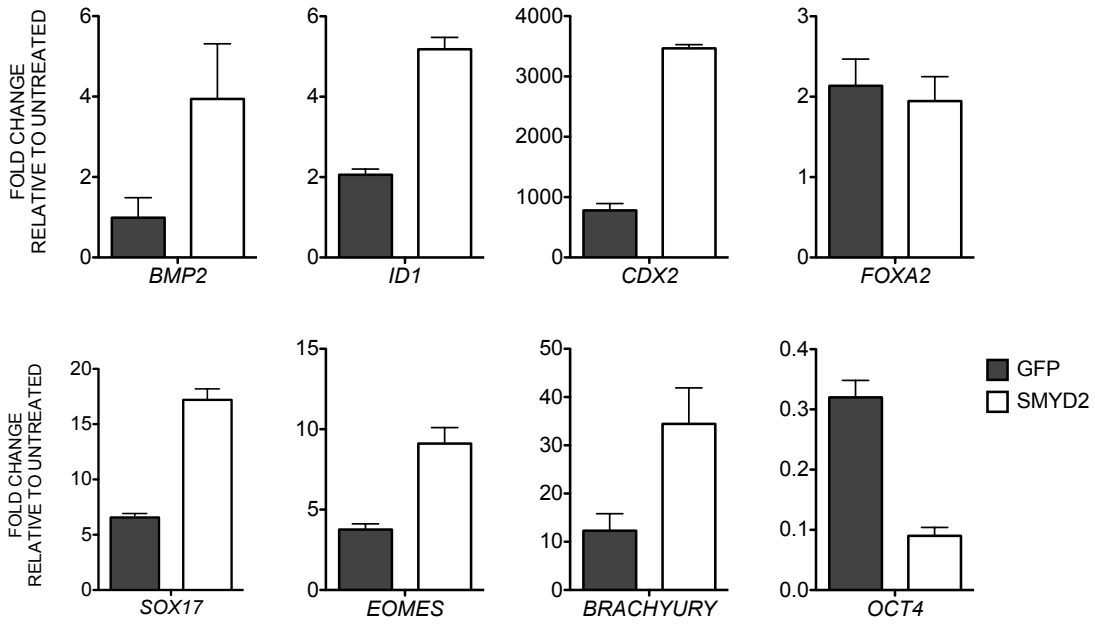


Figure 40. Induction of BMP2 target genes in human ES cells control and overexpressing SMYD2. ES[4]-GFP and ES[4]-SMYD2 cell lines were untreated and treated during 4 days with BMP2 (200 ng/ml). The mRNA levels of selected BMP2 target-genes (*BMP2, ID1, CDX2, FOXA2, SOX17*, *EOMES* and *BRACHYURY*) and a marker for pluripotency (*OCT4*) were measured by pPCR and normalized to *GAPDH*. Mean and standard deviation from two quantifications are shown. Expression levels are plotted as a fold change relative to untreated cells.

In order to validate the specificity on BMP2 target genes we treated both cell lines with high doses of Activin A (100 ng/ml), (Fig. 41). *EOMES*, *BRACHYURY* and *FOXA2* were induced after Activin A treatment but no differences were observed between ES[4]-SMYD2 and ES[4]-GFP cell lines.

To gain insight into the potential regulation of *SMYD2* expression by BMP2 or Activin A we quantified its mRNA levels in ES[4]-GFP after BMP2 or Activin A treatment. Endogenous *SMYD2* was strongly induced after BMP2 treatment and reached expression levels similar to the ones found in differentiated cells (Fig. 42). However, *SMYD2* was barely induced upon Activin A treatment.

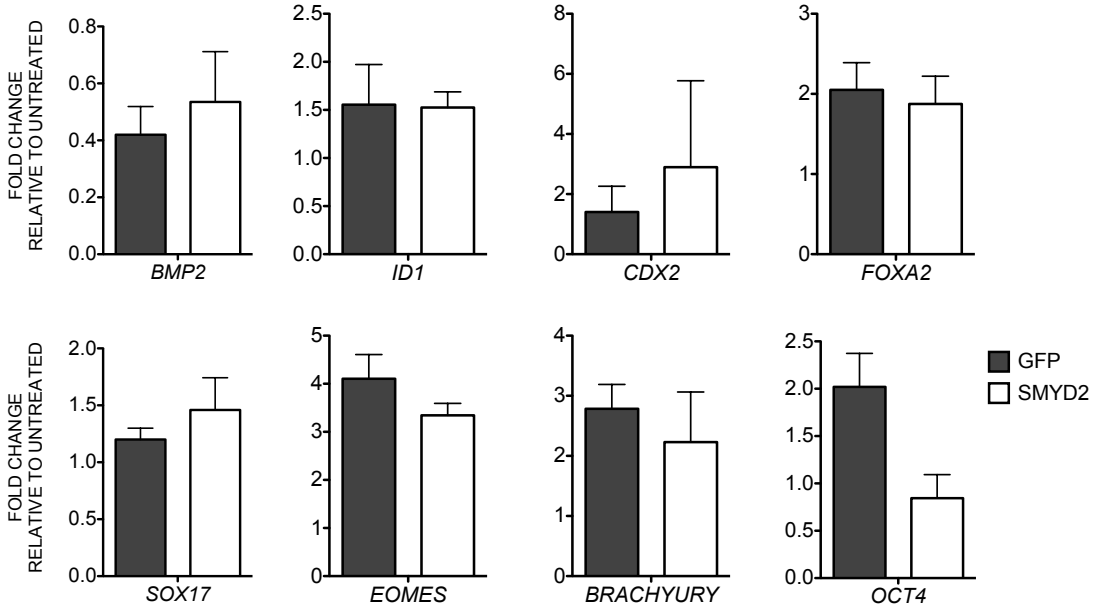


Figure 41. Induction of Activin A target genes in human ES cells control and overexpressing SMYD2. ES[4]-GFP and ES[4]-SMYD2 cell lines were untreated and treated during 4 days with Activin A (100 ng/ml). The mRNA levels of selected BMP2 target-genes (*BMP2, ID1, CDX2, FOXA2, SOX17*, *EOMES* and *BRACHYURY*) and a marker for pluripotency (*OCT4*) were measured by pPCR and normalized to *GAPDH*. Mean and standard deviation from two quantifications are shown. Expression levels are plotted as a fold change relative to untreated cells.

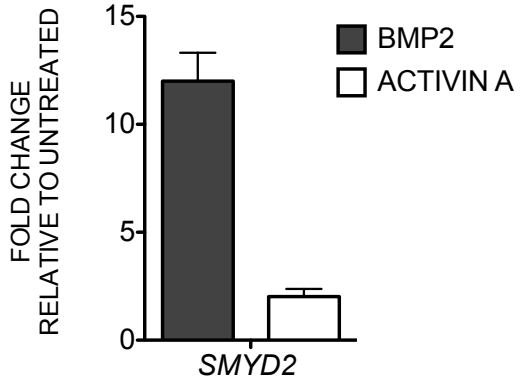


Figure 42. SMYD2 expression levels after BMP2 and Activin A treatment. Induction of *SMYD2* mRNA levels in ES[4]-GFP cell line after BMP2 (200 ng/ml) or Activin A (100ng/ml) treatment. mRNA levels were normalized to *GAPDH*. Mean and standard deviation from two quantifications are shown. Expression levels are plotted as a fold change relative to untreated cells.

Therefore, our results show that SMYD2 promotes the induction of BMP2 target genes in human ES cells, but has no significant effects on Activin A treatment. Additionally, SMYD2 is a BMP2 target gene in human ES cells.

3.5.2. SMYD2 stimulates BMP4 signaling in human ES cells

Next, we measured potential effects of SMYD2 overexpression in the phosphorylation of SMAD1/5/8 induced by BMP4. ES[4]-SMYD2 and ES[4]-GFP cell lines were treated with recombinant BMP4 for 1 hour and phospho-SMAD1/5/8 protein levels were measured by western blot (Fig. 43). The BMP4 treatment caused increased levels of phospho-SMAD1/5/8 in the ES[4]-SMYD2 cell line compared to the ES[4]-GFP cell line. These results further confirm that SMYD2 promotes BMP signaling.

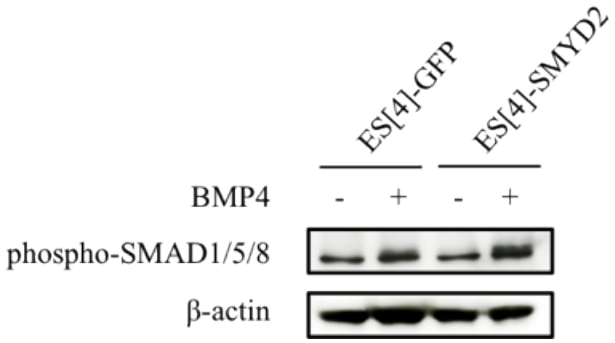


Figure 43. SMAD1/5/8 phosphorylation after BMP4 treatment of human ES cells overexpressing SMYD2. ES[4]-GFP and ES[4]-SMYD2 cell lines were untreated or treated with BMP4 (20 ng/ml) for 1 hour and the levels of phospho-SMAD1/5/8 were measured by western blot. β-actin was used as a loading control.

3.5.3. SMYD2 does not interact with SMADs family members *in vitro* or *in vivo*

It has been described that some histone methyltransferases, such as Suv39h, can interact with SMAD proteins and that this interaction can contribute to the transcriptional regulation mediated by BMP signaling [228]. Given that SMYD2 can modulate the BMP response in human ES cells, we tested its potential interaction with SMADs. We performed *in vitro* pull down assays with purified SMYD2 and several SMAD family members fused to GST: SMAD1 as a member of the BMP pathway, SMAD3 as a member of the TGF-β/Activin pathway, and SMAD4 for both signaling pathways. Figure 44 shows FLAG-tagged SMYD2 did not co-precipitate with any of the GST fusion proteins (Fig. 44).

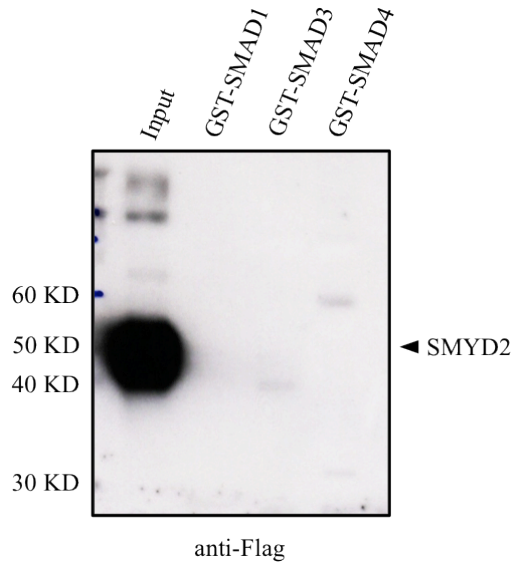


Figure 44. *In vitro* interaction assay between SMYD2 and SMAD1, SMAD3 and SMAD4. Purified FLAG-tagged SMYD2 was incubated with purified SMADs fused to GST (GST, GST-SMAD1, GST-SMAD3, and GST-SMAD4). GST fusion proteins were pulled down with glutathine-sepharose beads, washed, eluted and analyzed by western blot using an anti-FLAG antibody. 30% of input FLAG-SMYD2 was loaded.

A 293T cell line overexpressing FLAG-tagged SMYD2 (293T-SMYD2) was used to immunoprecipitate SMYD2 from the whole cell extract using an anti-FLAG antibody, and a line without SMYD2 over expression as a control (293T-GFP). Next, we tested the potential co-immunoprecipitation of phospho-SMAD1/5/8 in each eluted fraction by western blot (Fig. 45). However, we did not find any co-precipitation of phospho-SMAD1/5/8 with SMYD2. Taking together, these results indicate that SMYD2 do not interact with SMAD1, SMAD3, and SMAD4 *in vitro*, neither with phosphorylated SMAD1, SMAD5, and SMAD8 in 293T cells.

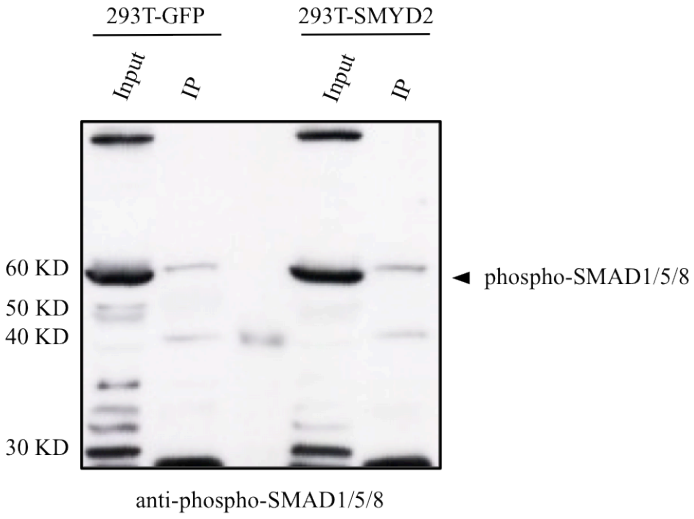


Figure 45. Co-IP of SMYD2 in 293T cells. Whole cell extracts from 293T-GFP and 293T-SMYD2 were immuniprecipitated using an anti-FLAG antibody and immunoblots were probed with an anti-phospho-SMAD1/5/8 antibody. 10% of input was loaded.

4. DISCUSSION

4.1. SMYD2 IS MAINLY EXPRESSED IN DIFFERENTIATED CELLS

The interest of our laboratory has been focused on the epigenetic mechanisms that contribute to establish and maintain the identity of differentiated cells. We hypothesized that critical chromatin regulators involved in this role might have differential expression between pluripotent and differentiated cells. Therefore, we tested the differential expression of several chromatin related enzymes, with a focus on methyltransferases and demethylases, between human pluripotent and differentiated cells.

We found that *SMYD2* is expressed at high levels in somatic cells compared with pluripotent cells (Fig. 11). This result confirmed our previous identification of SMYD2 as a candidate among a screening of differentially expressed genes in human ES cells. Moreover, we also compared the expression levels of *SMYD2* with the rest of SMYD family members between pluripotent and differentiated cells and *SMYD2* was the most differentially expressed member, whereas the rest of family members are not so differentially expressed (Fig. 12). Little is known about SMYD2, and more specifically in ES cells. In mouse adult tissue, *Smyd2* has been shown to be mostly expressed in heart and brain, although presenting levels of expression also in liver, kidney, thymus and ovary [119]. In neonatal mouse tissues, *Smyd2* expression also presents a broader organ distribution, but is highly expressed in heart, brain and skeletal muscle [129]. Additionally, *Smyd2* was also expressed in embryonic tissue, as well as *Smyd1* and *Smyd3* [119]. In mouse embryos at E13.5 stage, *Smyd2* expression is restricted to the heart and hypothalamus of the brain [119]. Regarding all expression experiments performed in mouse, it seems that Smyd2 is ubiquitously expressed after birth.

During human ES cells differentiation, *SMYD2* is also the most induced member of the SMYD family (Fig. 13a) both ES[4] and ES[2] cell lines. From very early at day 4, *SMYD2* is already highly induced compared with pluripotent cells. *SMYD1* and *SMYD3*, which share the highest degree of sequence homology with SMYD2 [113], are also induced during differentiation but not so remarkably. *SMYD3* shows a progressive

induction but only around 3-fold change compared with undifferentiated cells. Even at low levels, *SMYD1* was induced during differentiation, reaching its peak of expression at day 8. This indicates that SMYD1 might be involved also in early development, consistent with previous studies involving Smyd1 in mouse developing heart as a direct target of MEF2c [229]. Diehl *et al.* [129] followed the expression of *Smyd1*, *Smyd2* and *Smyd5* during heart development from embryonic stages through adult mouse, and both *Smyd1* and *Smyd2* present a pick of expression at neonatal stages in the mouse heart. As we mentioned before, Smyd1-3 are all widely expressed in adult mouse tissue, so maybe they share a similar function in certain specific tissues. For example, *in vivo* experiments have shown that Smyd1-3 play a role in cardiac and skeletal muscle regulation [107,108,110,132,133,229,230]. In addition, a SMYD4 homologue has been described also as a muscle-specific regulator in *Drosphila* development [112].

We also confirmed that during directed differentiation of ES[4] to NPCs endogenous *SMYD2* is induced and reaches similar levels of mRNA expression as during EB differentiation and as in somatic cells (Fig. 29) and that SMYD2 is induced in response to BMP treatment (Fig. 42). Despite the potential redundant role of the SMYD family members in differentiated cells, the fact that *SMYD2* is the most differentially expressed member of the family suggests specific roles for this member in the early stages of human ES cells differentiation.

4.2. SMYD2 PROMOTER IS MARKED WITH H3K4me2/3 IN HUMAN ES CELLS

During the last decade, genome-wide analyses using ChIP technology to study histone modifications revealed that many key developmental genes in ES cells show both active (H3K4me3) and repressive (H3K27me3) marks on their regulatory regions [149,151,231]. These bivalent domains have been proposed to keep those genes repressed, but ready for activation during ES differentiation. Given the low levels of expression of *SMYD2* in pluripotent cells and its rapid induction upon differentiation, we interrogated the presence of bivalent domains on its promoter. We found active H3K4 methylation marks on the *SMYD2* promoter in all studied cell lines (Fig. 14) but we could detect presence of H3K27me3 in the ES[2] cell line only. However, compared to the levels of H3K27me3 at SOX17 and FOXA2, two well-described bivalent genes in

ES cells, the levels of H3K27me3 are lower at the SMYD2 promoter. Analysis of two additional human ES cell lines in public data available through the ENCODE project also showed that the *SMYD2* promoter contains only H3K4me3 mark in the human ES cell line H7, whereas the H1 cell line shows both H3K4me3 and H3K27me3 marks (Fig. 15).

Since in several lines of human ES cells the promoter of *SMYD2* did not show detectable levels of H3K27me3, we looked for additional mechanisms that could be contributing to block the expression of *SMYD2* in these cells. Recently, Lipchina *et al.* [214] identified SMYD2, together with other 145 genes, as a high-confidence target of the miR-302/367 cluster in human ES cells. miR-302/367 is expressed in undifferentiated conditions and is downregulated upon neuronal differentiation, allowing the expression of its target genes [214]. Therefore, miR-302/367 might be regulating the levels of SMYD2 in pluripotent cells. Although 293T cells overexpressing miR-302/367 show lower mRNA levels of SMYD2 than control 293T we did not find any evidence that the miR-302/367 cluster targets the SMYD2 3' UTR region (Fig. 16a and 16b).

Interestingly, there are several studies that have identified transcriptionally inactive genes marked with H3K4me3 but not H3K27me3 on their promoters [163,232,233]. Vastenhouw *et al.* [163] have shown that among inactive genes after maternal-zygotic transition in zebrafish, 36% have H3K4me3 and H3K27me3, 28% have only H3K4me3, and 36% have no marks. Moreover, Vastenhouw *et al.* [163] concluded that according to published histone modifications [234-237] and gene expression data 21% of inactive genes show monovalent H3K4me3 marks in human ES cells and 20% in mouse ES cells. The role of H3K4me3 in transcriptional activation is controversial, since it is not clear if this mark participates in transcription or its deposition is a consequence of transcription. For example, H3K4me3 can be selectively recognized by the transcription factor TFIID [238,239]. Depletion of different subunits of the MLL complex in ES cells has been described to have different outcomes. Depletion of Dpy-30 resulted in a reduction of total H3K4me3 levels and impeded the activation of lineage-specific genes upon differentiation of mouse ES cells, whereas the expression of pluripotency-related genes was unaffected [157]. On the other hand, depletion of Wdr5, a core subunit of the MLL complexes, resulted in loss of self-renewal in ES cells and reduced the expression

of pluripotency-related genes [240]. Additionally, H3K4me3 may mark genes for future transcription under certain circumstances, but might not imply that genes carrying this mark are being transcribed. Importantly, it has been shown that PRC1-mediated histone H2AK119 ubiquitination restrains poised RNA pol II at developmental genes in mouse ES cells [156], and recent studies have shown that PRC1 complexes can be recruited to target genes and mediate the ubiquitination of H2A independently of the presence of H3K27me3 [63,66].

Certainly, additional experiments will be needed in order to determine if the SMYD2 gene is actually being transcribed in human ES cells or if its expression is post-transcriptionally regulated.

4.3. SMYD2 CAN METHYLATE SEVERAL HISTONE TARGETS *IN VITRO*

The specific enzymatic activity of SMYD2 as a histone lysine methyltransferase remains unclear. On one hand, Brown *et al.* [119] first described SMYD2 as a H3K36 methyltransferase whereas, on the other hand, Abu-Farha *et al.* [122] proposed that SMYD2 specifically methylates H3K4. This information is contradictory because, although both marks are associated with gene activation, they have different epigenetic roles in the transcription process [59]. In fact, it is possible that SMYD2 methylates both H3K4 and H3K36 depending on the presence of other proteins or preexisting histone modifications. Moreover, another report showed that SMYD2 can methylate *in vitro* several peptides corresponding to H3 and H4 tails (H3K4, H3K36me1, and H3K36me2, H4K20 and H3K27) [115]. This wide range of histone substrates for SMYD2 could be explained by the use of small peptides and the use of large amounts of purified SMYD2, which may facilitate the access of SMYD2 to the corresponding lysines. Finally, other reports suggested H4K20 as a possible candidate for SMYD2 methylation [115,241]

Our *in vitro* methylation assays using purified SMYD2 and recombinant or HeLa core histone octamers showed that SMYD2 methylates both H3 and H4 *in vitro* (Fig. 17). Introduction of mutations in histone tails fused to GST showed that the main *in vitro* activity is towards H3K4 and H4K20. Contrary to H3K4, methylation of H4K20 has been associated to transcriptional repression [75,76,78]. Interestingly, this H4K20

methylation activity of SMYD2 could be closely related with its previous described role as a transcriptional repressor mediated through its interaction with the repressive HDACI and Sin3A complex [119].

Besides core histones, we found that SMYD2 can methylate linker histone H1 *in vitro* (Fig. 18). We have recently shown that histone H1 variants have differential expression in pluripotent and differentiated human cells [242]. Moreover, H1 variants are recruited to the promoters of pluripotency-related genes during ES differentiation, suggesting that the dynamic expression of histone H1 variants may play a decisive role in gene silencing during the differentiation of human ES cells[242]. These findings, together with the involvement of SMYD2 in human ES cell differentiation, led us to speculate that SMYD2 methylation of histone H1 may contribute to the functional role of histone H1 in silencing the expression of pluripotency genes during differentiation.

Nevertheless, our methylation assay has some limitations. First, proteins were purified in bacteria, so they lack most of the eukaryotic post-translational modifications that could be decisive for the methylation process. Second, the histone substrates used consist of a short histone tail region fused to a GST protein of 30 kD in size, which differs notably from the full-length histone structure found in the nucleosome. Third, the absence of some potential SMYD2 co-activators, such as HSP90 [122], may have a negative effect on the methyltransferase activity.

I anticipate that several additional strategies could be performed in the future in order to unravel the specific methylation substrates of SMYD2 *in vivo*: (1) The use of nucleosomes or nucleosomal templates in our *in vitro* methylation assays might represent a more physiological substrate than the histone octamer or histone tails. (2) Correlate the presence of SMYD2 and histone modifications genome wide by ChIP-Seq (3) Correlate changes in histone modifications after depletion or overexpression of SMYD2 (4) identify proteins that interact with SMYD2 and that might be methylation substrates.

4.4. SMYD2 IS LOCALIZED IN BOTH THE CYTOPLASM AND THE NUCLEUS OF ES CELLS

Transient overexpression of FLAG-tagged SMYD2 in ES[4] provided us with a tool to observe the subcellular localization of SMYD2. SMYD2 was expressed both in the nucleus and the cytoplasm in undifferentiated and differentiated ES[4] (Fig. 23). Although we cannot rule out potential artifacts caused by the high ectopic expression of SMYD2, the same expression pattern was described in 293T cells by Brown et al. [119]. A recent publication has shown that endogenous SMYD2 is mainly expressed in the cytoplasm of 293 and C2C12 cells and human skeletal muscle tissue with only minor expression in the nucleus (19%) [132]. In addition, some other members of the SMYD family, SMYD1 and SMYD3, have been localized both in the nucleus and the cytoplasm. SMYD1 was described to be expressed in the nucleus and cytoplasm in C2C12 myoblast cells, but during myoblast differentiation the distribution of expression changes, and SMYD1 was mostly found in the cytoplasm with very low expression in the nucleus [243]. SMYD3 in mainly present in the cytoplasm of Huh7 cells arrested in G_0/G_1, but during S and G2/M phases it starts to accumulate in the nucleus [109]. Furthermore, a model for SMYD3 translocation was proposed by Ruden *et al.* [244] in which in the absence of estrogen or stress, SMYD3 interacts with Hsp90 in the cytoplasm and facilitates its activation and translocation to de nucleus where it participates in the transcription of WNT target genes. Possibly, given the closer similarities between SMYD1-3, SMYD2 might be also translocated between the cytoplasm and the nucleus, depending on the cell type or the cellular program. Interestingly, SMYD2 was also described to interact with Hsp90 [132,133], so it is easy to speculate that Hsp90 could also facilitate the translocation of SMYD2 into the nucleus.

4.5. GAIN AND LOSS OF FUNCTION OF SMYD2 AFFECTS DIFFERENTIATION OF HUMAN ES CELLS

Loss-of-function approaches revealed that SMYD2 is not essential for self-renewal of human ES cells. This observation is not surprising since SMYD2 is already expressed at very low levels in human ES cells. However, the knockdown of SMYD2 promoted the induction of endodermal genes during EB differentiation of human ES cells, compared to control lines (Fig. 21). *In vitro* differentiation experiments performed in monolayer

showed that the depletion of SMYD2 causes the accelerated silencing of the pluripotency gene OCT4, further suggesting that the absence of SMYD2 promotes differentiation (Fig. 22). This was an unexpected result since the pattern of SMYD2 expression suggests that its presence would be needed for human ES cell differentiation, similarly to histone H1.0 [242]. H1.0 is strongly induced during human ES cell differentiation and its absence impairs differentiation.

Overexpression of SMYD2 in ES[4] did not affect self-renewal, suggesting that the molecular machinery required for SMYD2 activity is not yet present in pluripotent ES cells. Overexpression of SMYD2 impaired the induction of mesendodermal genes and promoted the induction of the neuronal gene PAX6 during EB differentiation (Fig. 25). Despite this, the overexpression of SMYD2 did not have significant effects on the differentiation of ES cells towards NPCs (Fig. 29). However, we found that the endogenous levels of SMYD2 are induced during neuronal commitment, suggesting that SMYD2 may play a role in this process. In this regard, it would have been interesting to test whether the knockdown of SMYD2 reduces the differentiation efficiency into the neural lineage.

These experiments suggest that SMYD2 could be repressing the induction of endoderm. Since, SMYD2 has been previously involved in transcriptional activation through monomethylation of H3K4 we speculate that it might participate in the induction of repressors of the endodermal fate. Alternatively, SMYD2 might act as a transcriptional repressor itself through its association with HDACI and the Sin3A repressive complex [119]. Overexpression also causes the lack of induction of mesodermal genes suggesting that SMYD2 might be regulating a common pathway of endoderm and mesoderm induction, such as Nodal signaling [245]. Since we did not detect any effect of SMYD2 overexpression in the differentiation to NPCs we speculate that the overinduction of PAX6 during EB differentiation might be the result of an increase in the ratio of ectodermal versus mesendodermal fates. In the future, it would be interesting to test if SMYD2 overexpression affects the efficiency of definitive endoderm directed differentiation using available protocols published by D'Amour *et al.* [246] and Xu *et al.* [247], in which they obtain a strong induction of endodermal-related genes such as *MIXL1*, *GSC*, *SOX17* and *FOXA2*. Ultimately, determining the

genomic sites occupied by SMYD2 performing ChIP assays would reveal which genes are directly regulated by SMYD2.

Another potential mechanism that might be regulated by SMYD2 is cell proliferation and apoptosis. Both processes can have a physiological impact in differentiation. Published literature describes contradictories roles of SMYD2 in cell proliferation: SMYD2 has been described to induce cell proliferation through Rb methylation [128] and to suppress cell proliferation in NIH3T3 cells [119]. In addition, SMYD2 has been reported to repress p53 activity through Lys 370 methylation [124]. Although the potential role of p53 in differentiation is controversial, recent findings show that p53 knockdown causes a delay in differentiation of human ES cells while overexpression induces differentiation [248]. In any case, we could not observe any clear effects on the proliferation of 293T cells knockdown or overexpressing SMYD2, compared to their controls. However, overexpression of SMYD2 in human ES cells rendered a lower number of cells in S-phase at day 12 of differentiation, which correlates with an overinduction of the cell cycle inhibitor p16 (Fig. 27 and 28).

4.6. *smyd2a* DEPLETION CAUSES DORSALIZATION AND TAIL FORMATION DEFECTS IN ZEBRAFISH

We further investigated the effect of SMYD2 depletion *in vivo* using the zebrafish as a model. Over the past years, the zebrafish has emerged as a powerful model to study vertebrate development due to its transparency of embryos, which allows a direct visualization of tissue morphogenesis, and its rapid embryonic development. There are two homolgous genes identified for SMYD2, *smyd2a* and *smyd2b*. It is common to find a couple of orthologous genes in the zebrafish that are related to one single gene in human or mouse. This phenomena is well explained by the hypothesis of genome duplication that took place in the ancestor of teleost fishes [249,250]. Our results show that both *smyd2a* and *smyd2b* were induced during development, but only *smyd2a* was induced sharply during the gastrulation period (Fig. 31), which is comparable to the induction of SMYD2 during human ES cell EB differentiation into the three embryonic layers. Therefore, we analyzed the effects of blocking *smyd2a* expression during zebrafish embryonic development. *smyd2a* was maternally expressed in zebrafish embryos. This findings are supported by published data of *smyd2a* and *smyd2b* levels

by ISH assays [134]. Maternal gene products are generated during oogenesis and accumulated in the oocyte cytoplasm, to sustain the embryo during the first cleavage stages until the embryonic genome is activated [251,252]. In general, maternal mRNA participates during the very early steps of embryogenesis including fertilization, egg activation, the first cell divisions, and the initiation of zygotic transcription. To evaluate the involvement of SMYD2 in gastrulation, we designed a specific morpholino sequence that blocks the splicing of immature *smyd2a* mRNA, and therefore only affects the zygotic *smyd2a* transcript and not the maternal one.

Zebrafish embryos depleted of *smyd2a* suffered delays in development and tail deformities, and were present in more than 80% of surviving embryos at 24 hpf (Fig. 34 and 36). It has been well described that signaling pathways including Nodal, Bmp and Wnt8 are required for the tail organizer in zebrafish [204]. Furthermore, it has been proposed that both BMP and the non-canonical Wnt pathways regulate tail morphogenesis by controlling cell migration and cell adhesion within the tailbud [253]. For example, several studies have shown that disrupting Bmp signaling by specific morpholino injection against *bmp4* lead to a strong dorsalized embryos, with severe trunk and tail defects [207,254,255]. In addition, this tail-defective phenotype resemble to zebrafish mutant embryos *snailhouse* (*bmp7*) [256], and *swirl* (*bmp2b*) [257]. Similar to loss of Bmp signaling, embryos injected with specific morpholino against *wnt8* also exhibited problems with trunk and tail formation [258]. So, due to the similarities between smyd2a-MO injected embryos and the mentioned Bmp and Wnt8 morphant phenotypes, we first speculated that *smyd2a* may be involved in either the Bmp or Wnt8 pathways. Intriguingly, another group recently described that *smyd2a* knockdown generates severe skeletal and cardiac muscle defects in zebrafish embryos, and curled tails. [132,133]. We did not appreciate any cardiac abnormalities, even in the very severe morphant phenotype at 6 dpf, which in spite of its impairment for mobility given the severe absence of a tail structure, still presents a functional-beating heart (Fig. 34). One possible explanation for the differences in phenotype observed by us in zebrafish is that Donlin *et al.* used two distinct non-overlapping morpholinos to block *smyd2a* expression, one against the 5'UTR region, and the other one against the splice donor site of intron 4 [132], which blocks both maternal and zygotic *smyd2a* transcripts. Moreover, our observations are consistent with mice conditional-knockout experiments in which Smyd2 was found dispensable for heart development [129].

Co-injection of smyd2a-MO and SMYD2 mRNA, partially rescued the *smyd2a* morphant phenotype at 24 hpf (Fig. 36). This result reveals two main insights. First, it validates the specificity of the splice-blocking morpholino against *smyd2a* and second, it demonstrates the conserved function of *smyd2a* and SMYD2 among two evolutionary separated species such as zebrafish and humans, respectively. Unexpectedly, SMYD2 mRNA injected embryos showed no developmental defects (Fig. 35). Maybe SMYD2 mRNA is not stable enough to ensure the high levels of SMYD2 protein, but sufficient to compensate the lack of *smyd2a* and rescue the morphant phenotype.

4.7. *smyd2a* DEPLETION IMPAIRS THE INDUCTION OF *ntl* DURING GASTRULATION

We found that *ntl*, the zebrafish homologue of the mouse Brachyury, is mainly expressed in the developing margin of zebrafish embryos at 6 hpf, but practically absent in smyd2a-MO embryos [224]. It has been reported that *ntl* is an immediate early gene required for mesoderm development and that *ntl* mutant embryos lack differentiated notochord and exhibit aberrant tail formation [223,224]. The embryonic dorsalization observed in the *smyd2a* knockdown embryos resembles the phenotype of *ntl* mutant embryos [225]. Given the well-established implication of *ntl* in tail morphogenesis, it is easy to speculate that this absence of *ntl* gene expression in *smyd2a* morphant embryos may cause the defective tail formation. Nevertheless, other mechanisms besides *ntl* misexpression can generate dorsalization and tail formation defects. For example, Yao *et al.* [259] described that transcription factor *kzp* controls the zygotic expression of *wnt8* by directly binding to its promoter during zebrafish development. Depletion of *kzp* induces a strong embryo dorsalization, with a very similar phenotype than smyd2a-MO embryos. Importantly, *kzp* morphant embryos have no changes in *ntl* expression, suggesting that *kzp* regulates dorsoventral patterning through Wnt8 signaling independently of *ntl* expression.

Our results show that a percentage of smyd2a-MO injected embryos present reduced expression of *gata2* at 6 hpf and that expression continues to be lower at 8 hpf. Considering that *gata2* is expressed in the ventral region of wild type embryos, its misexpression might be also contributing to the dorsalized phenotype of smyd2a-MO injected embryos. Supporting this observation, it has been described that injection of

Bmp2b leads to embryo ventralization and results in the expansion of *gata2* expression into dorsal regions at 5 hpf [260]. *gata2* has been described as a non-neural marker that functions predominantly in hematopoietic development [261]. The endodermal gene *sox17* was misexpressed in some smyd2a-MO embryos compared to control expression around the margin region (Fig. 37). Apparently, the overinduction of *SOX17* observed during the differentiation of ES cells knockdown for SMYD2 does not correlate with the expression defects detected in zebrafish.

The expression of *myoD*, described to be expressed in the somites of the trunk and tail, was considerably reduced in some *smyd2a* depleted embryos, whereas others presented significant *myoD* expression levels, but they exhibited an unstructured somite organization (Fig. 38) It is possible that *smyd2a* and *smyd2b* play a role in the process of somitogenesis, since they have been described to be expressed in somites and muscle cells at 18-72hpf [134]. This would be in agreement with recent findings that associate the role SMYD2 with regulatory functions in skeletal muscle [132,133]. However, the observed defects might also be the consequence of defective germ layer specification during the early stages of gastrulation.

4.8. *smyd2a* DEPLETION MIGHT AFFECT MULTIPLE SIGNALING PATHWAYS IN ZEBRAFISH

We observed an overinduction of the Nodal target genes *chd*, *cas*, *gsc*, and *bon* in the *smyd2a* knockdown during gastrulation (Fig. 39). Published data by Chang et al. [262] on *foxd3* gene, an upstream positive regulator of Nodal signaling in zebrafish, suggests the involvement of Nodal signaling in dorsalization. *foxd3*-injected embryos exhibited induction of Nodal target genes and suppression of bmp7 expression at the gastrula stage, resulting in strongly dorsalized embryos at 24 hpf. Also, in *Xenopus*, it has been described that *goosecoid* and *Mix.1*, the homologous genes of *gsc* and *bon* respectively, act together to promote endodermal differentiation and suppress expression of mesodermal *Xbra*, the *ntl* homologous gene [263]. Eventually, high levels of *gsc* and *bon* at 6 and 8 hpf in smyd2a-MO embryos might have repressive effects over mesoderm induction, resulting in subsequent tail formation defects. This would explain the lack of *ntl* expression observed by ISH assays in smyd2a-MO embryos (Fig. 37). Additionally, *bon* and *cas* have been reported to trigger endoderm formation in

zebrafish [264,265], suggesting that similar to human SMYD2, *smyd2a* might be blocking the expression of endodermal genes in zebrafish.

Additionally, during the onset of gastrulation, Bmp and Wnt8 signaling play a crucial role in mesoderm and tail formation and are responsible for nodal-independent induction of *ntl* in the ventral margin of zebrafish embryo [207,266,267]. Moreover, disruption of Bmp signaling has been reported to cause the loss of *gata2* expression at early stages of development [260]. Regarding the Wnt pathway, the expression of *dkk1b*, a Wnt inhibitor, is highly overinduced in smyd2a-MO embryos at 6 hpf. Given the implications of Wnt8 pathway in mesoderm formation [267], the inhibition of Wnt8 signaling due to high *dkk1b* expression could be responsible for the defective tail formation in smyd2a-MO embryos. In accordance with this, Shinya et al. [268] have shown that zebrafish embryos injected with *dkk1* RNA do not present notochord and somite formation, resulting in similar tail defects as the smyd2a-MO embryos. However, the expression of *bmp2a* and *β-catenin-1*, two Bmp and Wnt pathways related genes respectively, was not affected by the *smyd2a* depletion. In any case, it would be interesting to check potential changes in the expression pattern of these genes in the zebrafish embryo, such as changes in the translocation of *β*-catenin to the nucleus, as a consequence of the *smyd2a* knockdown.

At the beginning of gastrulation, Bmp signaling is blocked by several antagonists produced by the organizer, such as Chordin, in order to induce dorsal fates by inhibiting ventralizing signals [227]. Meanwhile, Bmp also acts as a negative regulator of *chordin* expression to induce different ventral-cell fates along the dorsal/ventral axis. Considering that Nodal signaling patterns the organizer, we hypothesized that Smyd2a might participate as a negative regulator of the expression of Nodal-related genes through the Bmp signaling during gastrulation.

4.9. SMYD2 PROMOTES BMP SIGNALING IN HUMAN ES CELLS

BMPs are members of the TFG-β superfamily, which are involved in the regulation of a wide range of biological processes, including cell fate establishment during ES cell differentiation. Maintaining low BMP signal is necessary for human ES cells to keep self-renewal [269,270]. On the contrary, in mouse ES cells BMP has a key role in the maintenance of ES cell self-renewal through regulation of the expression of inhibitor of differentiation (Id) genes [271]. In spite of these differences between species, BMP signaling inhibits neural commitment in both human and mouse ES cells [272,273]. The mechanism of action behind the role of BMP in cell fate determination is still not fully understood and increasing evidence suggests that BMP signaling participates intimately with other signaling pathways including extrinsic signals, intrinsic transcription factors and epigenetic regulators [270]. Thus, the different functions of BMP signaling may be dependent on which context the BMP signals are received.

The effects of SMYD2 depletion on gastrulation and ES differentiation encouraged us to test the potential involvement of SMYD2 in BMP signaling. Our results show that SMYD2 overexpression stimulates the induction of almost all selected BMP2 target genes after treatment, as well as the silencing of OCT4 (Fig. 40). Importantly, SMYD2 appears to be a BMP2 target gene itself (Fig. 42). Activin A is another member of the TFG-β superfamily that has been described to participate in definitive endoderm differentiation, as well as in the maintenance of self-renewal in human ES cells, depending on concentration [246,247,274]. However, our treatment with Activin A had very minor effects on the expression of selected genes and was not affected by SMYD2 overexpression (Fig. 41). The absence of strong effects on gene expression after Activin A treatment might be explained by its dual ability to promote both self-renewal and differentiation of human ES cells [274]. Taken together, our findings show that SMYD2 overexpression potentiates the induction of BMP signaling in human ES cells, but has no significant effects on Activin A signaling. Given that BMP and Activin A pathways regulate gene expression by different signal-transduction cascades, such as activated SMAD1/5/8 and SMAD2/3 respectively, we speculated that SMYD2 may interfere with some specific effectors of the BMP signaling pathway only.

In agreement with the previous observations, we found that the overexpression of SMYD2 increases the levels of phospho-SMAD1/5/8 after BMP4 treatment (Fig. 43). SMYD2 overexpression had no significant effects on phospho-basal levels, which correlates with the absence of effects on gene expression in the absence of BMP. The effects of SMYD2 on SMAD1/5/8 phosphorylation suggest the possibility of SMYD2 interaction with the SMAD family members. Our results show that SMYD2 is not able to interact with SMAD1, SMAD3 or SMAD4 *in vitro* and no interaction was detected between SMYD2 and phospho-SMAD1/5/8 *in vivo* (Fig. 44 and 45). However, we cannot discard that SMYD2 may interact with some SMAD family member under specific conditions or may be cell type specific, such as in human ES cell differentiation, but not in 293T cells. Eventually, the *in vitro* interaction might require the presence of additional proteins or the presence of certain post-translational modifications in SMAD or SMYD2. Alternatively, the increase in SMAD1/5/8 phosphorylation caused by SMYD2 overexpression could be caused by indirect mechanisms, such as the repression of SMAD inhibitors.

Interestingly, some other histone methyltransferases have been reported to cooperate and participate in SMAD signaling. It has been shown that H3K9 methyltransferase Suv39h2 associates with Smad5, leading to the silencing of genes involved in developmental processes [228]. Moreover, the H3K9 methyltransferases Setdb1 and Suv39h1 interact with Smad3 and suppress IL-2 promoter activity [275]. More recently, Setdb1 was shown to interact directly with Smad1 in the presence in a BMP-dependent manner [276]. Therefore, despite we were not able to find a direct interaction between SMYD2 and SMAD proteins in the conditions used, it is possible that SMYD2 may act as a SMAD coactivator during human ES cells differentiation.

Some of the tested BMP-responsive genes show bivalent domains on their promoters in human ES cells [164,277,278]. In mouse ES cells, genome-wide mapping of target gene promoter occupancy by SMAD1/5 and SMAD4 revealed a large group of overlapping target genes, enriched in developmental regulators marked with bivalent marks [279]. Despite the differences in BMP signaling between mouse and human ES cells, it is likely that poised developmental genes in human ES cells are also co-occupied and regulated by SMAD1/5 and SMAD4. Based on this assumption, the potential regulation

of the BMP pathway by SMYD2 might influence the resolution of certain bivalent domains during the differentiation of human ES cells.

In summary, our results point to a potential role of SMYD2 as a repressor of endodermal genes during human ES cells differentiation, possibly by blocking Nodal signaling pathway as shown during gastrulation stage in zebrafish experiments. This repressive activity of SMYD2 may be conducted through methylation of H4K20 of Nodal-target genes, or methylation of H3K4 of Nodal antagonists. Alternatively, SMYD2 activation of BMP signalling may contribute to the blockade of Nodal-signaling pathway, given the BMP/Nodal antagonism. Induction of SMYD2 expression during human ES cells differentiation may be necessary to maintain optimal levels of Nodal signaling and ensure proper formation of mesendodermal lineage. Genome wide mapping of the sites occupied by SMYD2 as well as identifying interacting proteins might reveal the mechanisms of action of SMYD2 during differentiation.

5. CONCLUSIONS

1. SMYD2 expression is almost absent in human pluripotent cells but is strongly induced during human ES cells differentiation and prevails in adult somatic cell types.

2. SMYD2 presents specific methyltransferase activity for histone H3K4 (active mark) and histone H4K20 (repressive mark) *in vitro*.

3. Gain- and loss-of-function experiments revealed that SMYD2 plays an early decisive role during human ES cells differentiation:

    a) Knockdown of SMYD2 promotes the induction of endodermal genes and accelerates the process of differentiation.

    b) Overexpression of SMYD2 leads to a blockade of differentiation by impairing the induction of endodermal and mesodermal genes.

4. *smyd2a* expression is induced during zebrafish gastrulation.

5. Loss of *smyd2a* expression causes problems with mesoderm formation in zebrafish developing embryos:

    a) *smyd2a*-deficient embryos exhibit tail formation defects during the first days of development.

    b) Knockdown of *smyd2a* promotes the induction of Nodal-target genes during the gastrulation stage.

6. SMYD2 promotes the activation of the BMP pathway in human ES cells.

6. MATERIAL AND METHODS

6.1. MATERIALS

6.1.1. Bacterial strains

Name	Description
DH5α	F-φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(rK-, mK+) *pho*A *sup*E44 λ- *thi*-1 *gyr*A96 *rel*A1. Used for the amplification of plasmids and cloning
BL21(DE3)pLys	F– *omp*T *hsd*SB(rB–, mB–) *gal dcm* (DE3) pLysS (CamR). Used for the overexpression and purification of recombinant proteins

6.1.2. Cell lines

Name	Description
293T	Stable cell line from human embryonic kidney expressing the SV40 large T antigen. Used for transfection assays, protein-protein interactions and as packaging cells
293T-GFP	293T cell line transduced with pWPI vector stably expressing GFP
293T-SMYD2	293T cell line transduced with pWPI-SMYD2 vector stably expressing FLAG-SMYD2
293T-miR-302	293T cell line transduced with pWPI-miRNA302/367 vector stably expressing the human miR302/367 cluster (Established by Dr. Bernd Kuebler)
HFF	Primary cell line of human fibroblasts isolated from juvenile foreskin
HK	Primary cell line of human keratinocytes isolated from juvenile foreskin
HepG2	Stable cell line derived from human liver tissue with differentiated hepatocellular carcinoma
MEF	Primary cell line of Mouse Embryonic Fibroblasts established from dissociated C57BL/6 mouse embryos at 13.5 d gestation
ES[2]	Human ES cell line derived at the Stem Cell Bank of CMR[B] from a cryopreserved donated human embryo at day +2 of development
ES[6]	Human ES cell line derived at the Stem Cell Bank of CMR[B] from a cryopreserved donated human embryo at day +2 of development
ES[4]	Human ES cell line derived at the Stem Cell Bank of CMR[B] from a cryopreserved donated human embryo at day +2 of development
ES[4]-shControl	ES[4] cell line transduced with pLVTHM-shControl vector stably expressing a non-target random shRNA and GFP

ES[4]-shSMYD2	ES[4] cell line transduced with pLVTHM-shSMYD2 vector stably expressing a shRNA against SMYD2 and GFP
ES[4]-GFP	ES[4] cell line transduced with pTP6 vector stably expressing tau-GFP
ES[4]-SMYD2	ES[4] cell line transduced with pTP6-SMYD2 vector stably expressing FLAG-SMYD2
KiPS4F1	iPS cell line derived from HK infected with retrovirus encoding for Oct4, Sox2, Klf4 and c-Myc
KiPS4F8	iPS cell line derived from HK infected with retrovirus encoding for Oct4, Sox2, Klf4 and c-Myc

6.1.3. Vectors

Name	Description
pmirGLO	Mammalian reporter vector used to quantify miRNA activity by the insertion of miRNA target sites downstream of the firefly luciferase gene (*luc2*) (Promega)
pmirGLO-SMYD2.UTR	pmirGLO vector containing a SMYD2 full length 3' UTR region (+1336 to +1680) cloned into NheI and XbaI
pmirGLO-SMYD2.UTRshort	pmirGLO vector containing a SMYD2 3' UTR truncated region (+1336 to +1443) cloned into NheI and XbaI
pmirGLO-MBD2.UTR	pmirGLO vector conaining a MBD2 full length 3' UTR region (+1475 to +1860) cloned into NheI and XbaI
pRL-CMV	Mammalian reporter vector encoding for *Renilla* luciferase used for the normalization of transfection efficiency (Promega)
pBS-SK+	Phagemid plasmid used to reach optimal DNA concentration for co-transfection with pmirGLO constructs and pRL-CMV (Agilent Technologies)
pGEX-6P-1	Bacterial vector used for the expression and purification GST fusion proteins in *E.coli*. It contains an IPTG inducible promoter and ampicillin resistance (GE Healthcare)
pGEX-H3	pGEX-6P-1vector containing aminoacids 1-41 of *Xenopus laevis* histone H3 fused to GST (Cloned by Dr. Julio Castaño)
pGEX-H3K4R	pGEX-6P-1vector containing aminoacids 1-41 of *Xenopus laevis* histone H3 with Lys-4 mutated to Arg (K4R) fused to GST
pGEX-H3K36R	pGEX-6P-1vector containing aminoacids 1-41 of *Xenopus laevis* histone H3 with Lys-36 mutated to Arg (K36R) fused to GST
pGEX-H3K4RK36R	pGEX-6P-1vector containing aminoacids 1-41 of *Xenopus laevis* histone H3 with Lys-4 and Lys-36 mutated to Arg (K4RK36R) fused to GST
pGEX-H4	pGEX-6P-1vector containing aminoacids 1-36 of *Xenopus laevis* histone H4 fused to GST (Provided by Dr. Sohail Malik)
pGEX-H4K20R	pGEX-6P-1vector containing aminoacids 1-36 of *Xenopus laevis* histone H4 with Lys-20 mutated to Arg (K20R) fused to GST
pGEX-SMAD1	pGEX-6P-1vector containing full length human SMAD1 cloned into BamHI and XhoI

pGEX-SMAD3	pGEX-6P-1vector containing full length human SMAD3 cloned into BamHI and XhoI
pGEX-SMAD4	pGEX-6P-1vector containing full length human SMAD3 cloned into BamHI and XhoI
pGEX-4T-RXRα	pGEX-4T containing full length RXR alpha (provided by Dr. Diego Haro)
pGEX-HNF4	pGEX-4T containing full length human HNF4 (provided by Dr. Sohail Malik)
pGEX-mH2A1	pGEX-6P-1vector containing full length human macroH2A.1 (Cloned by Dr. Julio Castaño)
pGEX-AP2	pGEX-6P-1vector containing full length human AP2 (Cloned by Dr. Julio Castaño)
pGEX-E2F6	pGEX-6P-1vector containing full length human E2F6 (Cloned by Dr. Julio Castaño)
pLVTHM	Lentiviral vector used for the expression of shRNAs. Also encodes for GFP. All sequences were cloned into ClaI and MluI restriction sites (Provided by Didier Trono)
pLVTHM-shControl	pLVTHM vector containing a non-target random shRNA
pLVTHM-shSMYD2	pLVTHM vector containing a shRNA against SMYD2
pLVTHM-shSM1	pLVTHM vector containing a shRNA shM1 against SMYD2
pLVTHM-shSM2	pLVTHM vector containing a shRNA shM2 against SMYD2
pLVTHM-shSM3	pLVTHM vector containing a shRNA shM3 against SMYD2
pLVTHM-shSM4	pLVTHM vector containing a shRNA shM4 against SMYD2
pLVTHM-shSM5	pLVTHM vector containing a shRNA shM5 against SMYD2
pWPI	Lentiviral vector used for overexpression in mammalian cells. Also encodes for GFP (Provided by Didier Trono)
pWPI-SMYD2	pWPI vector containing human FLAG-SMYD2 cloned into PmeI. SMYD2 cDNA was obtained from pCR-BluntII-TOPO-SMYD2 (cDNA clone MGC:119302 IMAGE:40005280)
pWPI-miRNA302/367	pWPI vector expressing the human miRNA302/367 cluster (Cloned by Drs. Maria Barrero and Josipa Bilic)
pTP6	Mammalian expression vector encoding for tau-GFP and provides puromycin resistance (kindly provided by T. Pratt)
pTP6-SMYD2	pTP6 vector containing human FLAG-SMYD2 cloned into EcoRI-blunted restriction sites
pCS2+	Expression vector used for *in vitro* mRNA synthesis under the SP6 promoter
pCS2-SMYD2	pCS2+ vector containing human FLAG-SMYD2 cloned into XhoI restriction site
pCMV-VSVG	Envelope vector used for lentiviral production (kindly provided by D. Trono)
psPAX2	Packaging vector used for lentiviral production (kindly provided by D. Trono)

6.1.4. Oligonucleotides

Quantitative Real Time PCR		
Name	**Forward (5'- 3')**	**Reverse (5'- 3')**
SMYD2	CAATCCGAGACATGGTCAGA	GCCCTCCGGAACTCTTCAA
GAPDH	GCACCGTCAAGGCTGAGAAC	AGGGATCTCGCTCCTGGAA
SMYD1	CACCAGGTCCCACGTTTGTT	TCATCATGTGGTGAGGATGGA
SMYD3	GCCGCGTCGCCAAA	TGTGGTCTGGCCAAGCTTTT
SMYD4	TGCCACCACCGCATCA	AATAGAGGTCCCCATTGGTGACT
SMYD5	TGTGCACTGTGCGCAAAGA	TGCAGTACATCACTTGGCAATG
OCT4	GGAGGAAGCTGACAACAATGAAA	GGCCTGCACGAGGGTTT
NANOG	ACAACTGGCCGAAGAATAGCA	GGTTCCCAGTCGGGTTCAC
SOX2	TGCGAGCGCTGCACAT	TCATGAGCGTCTTGGTTTTCC
HNF4	CTGCAGGCTCAAGAAATGCTT	TCATTCTGGACGGCTTCCTT
FOXA2	CTGAAGCCGGAACACCACTAC	CGAGGACATGAGGTTGTTGATG
SOX17	TGGCGCAGCAGAATCCA	CCACGACTTGCCCAGCAT
PAX6	GCTTCACCATGGCAAATAACC	GGCAGCATGCAGGAGTATGA
BRACHYURY	CAATGAGATGATCGTGACCAAGA	GCCAGACACGTTCACCTTCA
p21	TGGAGACTCTCAGGGTCGAAA	GCGTTTGGAGTGGTAGAAATCTG
p16	CCCTCAGACATCCCCGATT	TCTAAGTTTCCCGAGGTTTCTCA
BMP2	CCAACACTGTGCGCAGCTT	CCCACTCGTTTCTGGTAGTTCTTC
ID1	GCTGGACGAGCAGCAGGTA	GCGTGAGTAACAGCCGTTCA
CDX2	CTCGGCAGCCAAGTGAAAA	GGTCCGTGTACACCACTCGAT
EOMES	CCACTGCCCACTACAATGTGTT	CTGGAAGCGCCAGTGGTT
smyd2a	AGAGCTGTCTCATTTGGGATCTG	TGGGCTACAGCTGTGATTCATC
smyd2b	CAAATCTATGGATGTGGTGAAGATG	TCTCCTTTTCCCCGATTTCAT
chd	CCGCAGGCCGATAACG	GTGAGGTTTCGGCACATTTTTC
cas	GGAGAACACTGACCTCAGCAAA	GCTTATCTGCCAGAGACATTGCTT
gsc	CAATCGGTGAATGGAAGGATAG	GTTTCGTGCCAGTTGTTCTC
bon	GGAGGACGGGCATGCA	GCACCATCCAGGAACTTCATC
gata2	GGGCTAGAGTATCGTACGGACAA	TGGCCGACAGACCTGACTTC
dkk1b	GCGAGAGTGATGAGGAATG	CTTTCGGCATGAGAGCAGT
bmp2b	CAGCAGAGCAAACACGATACG	GCTGGACAGTGCCTCGAAA
β-catenin-1	GGCACCCTACACAATCTTTC	CAAGGCCACCATTTTCTGC
SMYD2.PROM	AACCCTCTGCACACCAAACTTC	GGAACGCCAAGGAGAAAGC
OCT4.PROM	CAGTTGTGTCTCCCGGTTTTC	CGAAGGATGTTTGCCTAATGG
FOXA2.PROM	CAGCAGCGGGCGAGTTA	TGCAAAGCGCTGTCCTATTTAG
SOX17.PROM	CCGTCTGTCCAGTCTTGCTTATT	TGAAGGCCAGGAGTTTGCA

Semiquantitative RT-PCR		
Name	**Name**	**Name**
smyd2a	GTGTATGTATGTGTGTGATTGCC	GCTTGTTTACATTTCCCACAC
smyd2b	GTCAGAGATTCTGCTCTTACTGG	GAAGTTTTCCATGGAGTTCCG
β-actin 2	TGTGGCCCTTGACTTTGAGCAG	TAGAAGCATTTGCGGTGGACGA
Insert amplification		
Name	**Forward (5'- 3')**	**Reverse (5'- 3')**
SMYD2 (pWPI)	ACGTTTAAACGCCGCCGCCATGGACTACAAGGACGACGATGACAAGAGGGCCGAGGGCCTCGGCGGCC	ACCAAATTTGTCAGTGGCTTTCAATTTCCTGTTTGATC
SMYD2 (pCS2)	ACGTTTAAACTCGAGGCCGCCGCCATGGACTACAAG	ACGTTTAAACTCGAGTCAGTGGCTTTCAATTTCCTG
SMYD2.UTR	ACGCTAGCAACTATGCAGCATTTCAGTTTTC	ACTCTAGACTGATATCTCTTTAATACTTTC
SMYD2.UTR short	ACGCTAGCAACTATGCAGCATTTCAGTTTTC	ACTCTAGACCAATTATTTTACAGAGATGCTG
SMAD1	ACGGATCCATGAATGTGACAAGTTTATTTTCCTTTACA	ACCTCGAGTTAAGATACAGATGAAATAGGATTATGA
SMAD3	ACGGATCCATGTCGTCCATCCTGCCTTTC	ACCTCGAGCTAAGACACACTGGAACAGCGG
SMAD4	ACGGATCCATGGACAATATGTCTATTACGAATACACC	ACCTCGAGTCAGTCTAAAGGTTGTGGGTCTG
Mutagenesis		
Name	**Forward (5'- 3')**	**Reverse (5'- 3')**
H3K4R	GATCCATGGCCCGTACCAGGCAGACCGCCCGTAAATC	GATTTACGGGCGGTCTGCCTGGTACGGGCCATGGATC
H3K36R	CTGCTACCGGCGGAGTCAGGAAACCTCACCGTTAG	CTAACGGTGAGGTTTCCTGACTCCGCCGGTAGCAG
H4K20R	CTAAACGTCACCGTAGAGTTCTGCGTGACAACATC	GATGTTGTCACGCAGAACTCTACGGTGACGTTTAG
Short-hairpins		
Name	**Forward (5'- 3')**	**Reverse (5'- 3')**
shControl	CGCGTCCCCCTTGCTATGAGAACAAATTTTCAAGAGAAATTTGTTCTCATAGCAAGTTTTTGGAAAT	AGGGGGAACGATACTCTTGTTTAAAAGTTCTCTTTAAACAAGAGTATCGTTCAAAAACCTTTAGC
shSMYD2	CGCGTCCCCGGATTGTCCAAATGTGGAAGACGAATCTTCCACATTTGGACAATCCTTTTTGGAAAT	CGATTTCCAAAAAGGATTGTCCAAATGTGGAAGATTCGTCTTCCACATTTGGACAATCCGGGGA
shSM1	CGCGTCCCCCCGGTTAAGAGATTCTTATTTCAAGAGAATAAGAATCTCTTAACCGGTTTTTGGAAAT	CGATTTCCAAAAACCGGTTAAGAGATTCTTATTCTCTTGAAATAAGAATCTCTTAACCGGGGGGA
shSM2	CGCGTCCCCGCCGTACTTTGAGGTTAGTTTCAAGAGAACTAACCTCAAAGTACGGCTTTTTGGAAAT	CGATTTCCAAAAAGCCGTACTTTGAGGTTAGTTCTCTTGAAACTAACCTCAAAGTACGGCGGGGA
shSM3	CGCGTCCCCCCCAACGGAAGATAGAAATTTCAAGAGAATTTCTATCTTCCGTTGGGTTTTTGGAAAT	CGATTTCCAAAAACCCAACGGAAGATAGAAATTCTCTTGAAATTTCTATCTTCCGTTGGGGGGGA
shSM4	CGCGTCCCCGTCTGAATCTTGAACTTTATTCAAGAGATAAAGTTCAAGATTCAGACTTTTTGGAAAT	CGATTTCCAAAAAGTCTGAATCTTGAACTTTATCTCTTGAATAAAGTTCAAGATTCAGACGGGGA
shSM5	CGCGTCCCCGGAGACTGTAAGACTAACATTCAAGAGATGTTAGTCTTACAGTCTCCTTTTTGGAAAT	CGATTTCCAAAAAGGAGACTGTAAGACTAACATCTCTTGAATGTTAGTCTTACAGTCTCCGGGGA

Morpholinos	
Name	**Forward (5'- 3')**
smyd2a-MO	TTATAAGGAGCGCTGACCTGGTAAA
Control-MO	CCTCTTACCTCAGTTACAATTTATA

6.1.5. Antibodies

Western blot		
Name	**Organism**	**Provider**
SMYD2 (Y-22)	Rabbit	Santa Cruz Biotechnology
FLAG(M2)	Mouse	Sigma
β-actin	Mouse	Sigma
phospho-SMAD1/5/8	Rabbit	Cell Signaling Technology
anti-Rabbit (2^{nd})	Goat	Santa Cruz Biotechnology
anti-Mouse (2^{nd})	Goat	Santa Cruz Biotechnology
Immunohistochemistry		
Name	**Organism**	**Provider**
SMYD2 (Y-22)	Rabbit	Santa Cruz Biotechnology
PAX6	Rabbit	Covance
anti-Rabbit (CY3)	Donkey	Jackson ImmunoResearch
anti-Mouse (CY2)	Donkey	Jackson ImmunoResearch
anti-Mouse (CY5)	Goat	Jackson ImmunoResearch
FACS		
Name	**Organism**	**Provider**
OCT3/4	Mouse	BD Biosciences

6.1.6. Reagents

Molecular biology	
Name	**Provider**
SYBER® Safe DNA Gel Stain	Invitrogen
Anctartic Phosphatase	New England Biolabs
Klenow enzyme	Roche
Taq DNA polymerase	Roche
T4 DNA Ligase	Roche

Phusion® Hot Start	Thermo Scientific
BIGDye Kit	Invitrogen
TRIzol® Reagent	Invitrogen
Cloned AMV Reverse Transcriptase	Invitrogen
SYBER® Green	Invitrogen
Bio-Rad Protein assay	Bio-Rad Laboratories
ECL Plus detection kit	GE Healthcare Europe
H^3SAM	PerkinElmer España SL
Recombinant SMYD2	Active Motif
Histone H1 (calf thymus)	Santa Cruz
HeLa core histones	Laboratorios Conda, S.A.
Amplify solution	Amersham
Glutathione-sepharose	Amersham
FLAG peptide	Sigma
M2 FLAG beads	Sigma
Dynabeads Protein A	Invitrogen
Proteinase K	Sigma
RNase A	Sigma
Polybrene	Millipore
mMESSAGE mMACHINE kit	Invitrogen
DIG RNA Labeling Kit SP6/T7	Roche
Passive Lysis Buffer	Promega
Dual-Luciferase® Reporter Assay System	Promega
Cell culture	
Name	**Provider**
Matrigel	BD Biosciences
Knockout DMEM	Gibco
DMEM	Gibco
DMEM/F12	Gibco
FBS	Gibco
KSR	Gibco
Non-essential amino acids	Gibco
2-mercaptoethanol	Gibco
Penicillin	Gibco
Streptomycin	Gibco
GlutaMAX	Gibco
bFGF	Peprotech
0.25% trypsin-EDTA	Invitrogen

0.05% trypsin-EDTA	Invitrogen
Epilife® medium	Invitrogen
TrypLE Express	Invitrogen
Fugene6® reagent	Roche
FIX and PERM® kit	Invitrogen
Click-iT® EdU AlexaFluor647® Flow Cytometry Assay kit	Invitrogen
mTeSR1 media	Stemcell Technologies
Accutase	Lab Clinics SA
SB431542	Tocris
Noggin	R&D Systems
N-2 supplement	Invitrogen
BMP2	Abnova
Activin A	Abnova

6.2. METHODS

6.2.1 DNA manipulation techniques

6.2.1.1. Bacterial growth conditions and transformation

E. coli was grown in liquid LB media (casein 10 gr, yeast extract 5 gr, NaCl 10gr and bacto-agar 15 gr, in a final volume of 1 liter) and solid LB plates. In addition, different antibiotics have been added according to the different *E. Coli* strands and to the specific plasmids resistances: kanamycin (30 μg/ml), cloramphenycol (20 μg/ml), or ampicillin (100 μg/ml).

- **Preparation of DH5alpha heat shock competent cells**

On single colony of *E.coli* from a LB-agar plate was inoculated into 5 ml of LB media and grown at 37°C for 8 hours. Bacterial culture was diluted 1:50 with SOB media (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$) and grown at 18°C to OD_{600} = 0.6. Next, Bacterial culture was placed on ice for 10 minutes and centrifuged at 2,500 g for 10 min at 4°C. Supernatant was discarded

and the pellet was resuspended in 80 ml ice-cold TB media. Again, bacterial suspension was centrifuged at 2,500 g for 10 min at 4°C and the pellet was resuspended in 20 ml ice-cold TB media. DMSO was added to a final concentration of 7% and placed on ice for 10 minutes. Several aliquots were prepared and frozen immediately in liquid N_2 and stored at -80°C.

- **Transformation using the heat shock method**

A suitable volume of DNA was added into 100 µl of competent bacteria and mixed gently. Incubation of DNA-bacteria mixture was performed for 30 min on ice. Next, the mixture was incubated for 90 seconds at 42°C, followed by 2 minutes on ice. LB media was added and incubated at 37°C shaking for 1 hour. The mixture was spread into an LB plates, containing the corresponding antibiotic, and incubated for 16-24 hours at 37°C.

6.2.1.2. DNA extraction from bacterial cells

The extraction of plasmidic DNA from *E. coli* cells was performed using the commercial QIAprep Spin Miniprep Kit (QIAGEN). The procedure consists in the lysis of the cells, the precipitation of the unsoluble cell debrees and bacterial genomic DNA and the binding of the plasmidic DNA to a positively charged membrane that retains the DNA while proteins and other components are washed away. The purified DNA was eluted into 25-50 µl of TE buffer or water.

6.2.1.3. Analysis of DNA fragments by agarose gel

The analysis, identification and isolation of DNA fragments were performed through agarose gel electrophoresis using non-denaturating conditions. The percentage of agarose used was variable, between 0.8 % and 2 %, depending on the size of the DNA fragments to analyze. The electrophoresis buffer used was TAE 1X and the detection of DNA fragments was performed by SYBER® Safe DNA Gel Stain (Invitrogen). To estimate the size of the DNA fragments we used molecular weight markers

6.2.1.4. Plasmid construction

6.2.1.4.1. Cloning

➢ Vector preparation

Digestion with restriction enzymes

We used a standard protocol for the digestion reaction of plasmidic DNA to ensure the total digestion of the DNA substrate without generating unspecific cuts.

DNA concentration in the final solution was between 20 and 200 ng/µl. Glycerol never exceeded 10 %. We used one unit on enzyme per 1µg of DNA. We digested 2-3 µg of DNA into a final volume of 20-50 µl. The amount of restriction enzyme, the buffer, the temperature and the time of the reaction were specific for each enzyme used. Restriction enzymes were supplied by New England Biolabs.

Vector Dephosphorylation

To dephosphorylate 1 µg of vector DNA we added 1 unit of Antarctic Phosphatase (New England Biolabs), and incubated during 1 hour at 37°C. Next, phosphatase was inactivated after incubation at 65°C during 10 minutes. DNA was purified using the PCR Purification Kit (QIAGEN), following the manufacturers protocol and eluted in a suitable volume of TE buffer or water.

➢ Insert preparation

A) Restriction enzyme digestion

We followed the same strategy as for the vector digestion with restriction enzymes. Digested products were isolated by agarose gel and purified using the Gel Extraction Kit (QIAGEN).

Generation of DNA blunt ends: Klenow fragment

If appropriate, purified DNA from digestion was treated with 2 U/µg DNA of Klenow enzyme (Roche) and 0.5 mM of dNTPs at 37°C during 1 hour in a final volume of 50 µl. This is used to fill the 5′ protruding DNA ends allowing the formation of blunted ends. The resulting blunt-DNA insert was purified using the PCR Purification Kit (QIAGEN).

B) PCR amplification

The reaction was performed in the presence of 10-50 ng of DNA template, 1X buffer reaction containing 1 mM $MgCl_2$, 0.5 mM of dNTPs, 0.3 µM of each complementary primers, and 1-2 U of a thermostable Taq DNA polymerase (Roche) in a final volume of 50 µl. The stardard protocol for PCR amplification was performed as follows: 95°C for 2 minutes; 20 to 25 cycles of 95°C for 15 seconds, 60°C* for 30 seconds (* temperature depends on the annealing temperature of primers), 68°C for 1minute per Kb; final extension of 68°C for 10 minutes.

C) Short-hairpin oligonucleotides annealing

We used a specific short-hairpin sequence to knock down SMYD2 in human ES cells. The oligonucleotides annealing reaction contained 10 µl of each 100 µM forward and reverse primers, 3 µl of 5M NaCl and 77 µl of TE. The mixture was heated at 95°C during 5 minutes and subsequently cooled slowly overnight (O/N) at room temperature (RT), to allow the annealing of the two complementary oligonucleotides. For cloning, we used 5 µl of annealing mixture and 50 ng of the host vector.

➢ Ligation

Once obtained purified vector and insert, ligation was performed in a general molar ratio vector:insert of 1:3 for sticky ends, or 1:10 for blunt ends. The reaction was usually performed in 10 µl of final volume and 1 U of T4 DNA Ligase (Roche) O/N at 16°C or during 1-2 hours at RT.

6.2.1.4.2. Site-directed mutagenesis

In order to introduce point mutations in histone tails we carried out site-directed mutagenesis reactions. The reaction was performed using 20 ng of DNA template, 1X reaction buffer containing 1 mM $MgCl_2$, 0.5 mM dNTPs, 0.3 µM each primer, 1.5 µl DMSO, and 1 U of Phusion® Hot Start (Thermo Scientific) in a final volume of 50 µl. PCR amplification was performed as follows: 98°C for 30 seconds; 25 cycles of 98°C for 30 seconds, 60°C for 30 seconds, 72°C for 4 minutes; final extension of 72°C for 10 minutes. After PCR amplification, DNA template was digested with 1 µl of DpnI during 1 hour at 37°C. About 1µl of this reaction was used to transform DH5α competent cells. Screening of positive colonies was done by sequencing.

6.2.1.5. Sequencing

We used the BIGDye Kit (Invitrogen) to amplify 300 ng of template DNA with 1 µM of specific primer in a final volume of 10 µl. The PCR amplification was performed as follows: 96°C for 2 minutes; 35 cycles of 96°C for 10 seconds, 50°C for 5 seconds, 60°C for 4 minutes. The product of the reaction was analyzed at the Pompeu Fabra University genomics Genomics Core Facility using an automated sequencer that discriminates DNA fragments that differ from a single base pair.

6.2.1.6. Real Time PCR analysis

RNA extraction

For real time RT-PCR experiments cells were collected from 10 cm tissue culture plates, centrifuged at 500 g for 5 min, and rinsed with PBS. Total cellular RNA was extracted using TRIzol® Reagent (Invitrogen) following the manufacturer instructions. Final RNA was resuspended in 30-50 µl of water.

cDNA synthesis

1 µg of purified RNA was converted into cDNA using the Cloned AMV Reverse Transcriptase (Invitrogen) following the manufacturer instructions in a final volume of 20 µl.

Real Time PCR

The mixture of the PCR reaction consists of 0.5-2 µl of cDNA, 10 µl SYBER® Green (Invitrogen), 0.5 µM each primer in a final volume of 20 µl. The PCR amplification was performed as follows: 50°C for 2 minutes, 95°C for 10 minutes, 40 cycles of 95°C for 15 seconds, 60°C for 1 minutes. To obtain a dissociation curve the following steps were also included at the end of the reaction: 95°C for 15 seconds, 60°C for 30 seconds, 95°C for 15 seconds. The relative expression of Real Time PCR products was determined using the fold induction method = $2^{-\Delta Ct}$, where Ct is the threshold cycle. Using this procedure the value of expression of each gene analyzed is normalized to the expression of a housekeeping gene such as *GAPDH*.

6.2.2. Protein manipulation techniques

6.2.2.1. Electrophoretic techniques for protein analysis

- **Protein extraction**

Total soluble protein extracts were obtained from the different cell lines using the RIPA extraction method. Cells were harvested from culture dishes and collected by centrifugation at 500 g for 5 minutes at 4°C. The pellet was resuspended in a volume of RIPA lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 1 mM EDTA, 0.1% SDS, 1mM PMSF and a mix of proteases inhibitors) and incubated during 20 minutes at 4°C. Soluble proteins were collected from the supernatant after centrifugation at 14000 rpm during 15 minutes at 4°C and protein concentration was determined using the Bio-Rad Protein Assay (Bio-Rad Laboratories). Samples were stored at -80°C.

➢ SDS-PAGE electrophoresis

Protein analysis was performed by electrophoresis in polyacrylamide gels under denaturant conditions (SDS-PAGE). We used gels with 1.5 mm thickness with 10-13% of acrylamide for the resolving gels (depending on the molecular weight of the proteins to resolve), and with 5% of acrylamide for the stacking gel. In all cases we kept the ratio acrylamide: bisacrylamide 37.5: 1.

Samples were prepared adding SDS loading buffer 2X (100 mM Tris-HCl pH 6.8, 4% SDS, 20% glycerol, 0.2% bromophenol blue and 200 mM 2-mercaptoethanol) and heated for 5 minutes at 100°C. Boiled samples were loaded in the acrylamide gel and resolved under a constant voltage during 1-2 hours. The molecular weight of the proteins was estimated by loading the protein marker Novex® Sharp Unstained Protein Standard (Invitrogen).

➢ *Western blot*

Proteins were transferred from the acrylamide gel to a PVDF membrane using a Semi-dry Transfer Apparatus (BioRad Laboratories) and transfer buffer (25 mM Tris Base, 200 mM glycine and 20% methanol). The protein transference was performed during 1 hour at 10 V.

➢ Immunodetection

Protein detection was performed by the sequential incubation of the PVDF membrane with a specific primary antibody and a subsequent secondary antibody conjugated with horseradish peroxidase (HRP). First, non-specific protein binding was blocked with TBS-Tween (10 mM Tris-HCl pH 8, 150 mM NaCl and 0.1% Tween 20) containing 5% milk powder for 1 hour at RT. Next, the membrane was incubated with the primary antibody diluted in TBS-Tween O/N at 4°C. The next day, the membrane was washed with TBS-Tween and incubated with the secondary anti- rabbit/mouse/goat antibody for 1 hour at RT. The membrane was washed with TBS-Tween and the HRP activity was detected using ECL Plus detection kit (GE Healthcare Europe) and exposed to an Amersham ECL Hyperfilm (VWR international Eurolab).

6.2.2.2. Immunohistochemistry assay

Cells were grown on plastic cover slide chambers (170920, NUNC) and fixed with 4% paraformaldehyde (PFA) (Sigma-Aldrich Quimica) during 20 minutes at RT. Slides were washed three times with TBS and blocked with TBS+ (0.5% Triton X-100 and 6% donkey or goat serum) during 30 minutes at RT. Slides were incubated overnight with the primary antibody in TBS++ (0.1% Triton X-100 and 6% donkey serum) at 4°C. The next day, the slides were washed with TBS++ and incubated with CY-conjugated secondary anti-rabbit/mouse/goat during two hours at 37°C. Finally, the slides were washed with TBS++ for 5 minutes at RT and counterstain with DAPI (1:10,000 diluted in water from stock solution) for 10 minutes at RT in the dark. Slides were covered with mounting media and a 24 x 40 mm Knittel Glass coverslip. Images were taken using Leica SP5 confocal microscope and Leica AF6000 software.

6.2.2.3. Recombinant protein expression and purification in *E. coli*

- **Protein expression**

The corresponding pGEX-6P-1 constructs were transformed into competent BL21(DE)pLys by the heat shock method and plated in ampicillin plates. A BL21(DE3)pLys colony was picked and inoculated in 50 ml of ampicillin/LB media and incubated O/N shaking (220 rpm) at 37°C. The next day the cultures were diluted 1:10 in 200 ml of ampicillin (2x)/LB media and incubated shaking at 30°C until the D.O reached 0.6. Then cultures were induced with 0.5 mM IPTG and kept growing in the same conditions for 4 hours. After that cultures were centrifuged at 8000 rpm for 15 minutes at 4°C.

- **Protein purification**

Pellet was resuspended in 10 ml of fresh NTEN buffer (0,5% NP-40, 1 mM EDTA, 20 mM Tris-HCl pH 8,0, 150 mM NaCl, proteases inhibitors and PMSF 1mM) and sonicated four times for 30 seconds. Samples were centrifugated at 12,000 rpm for 20 minutes at 4°C. Next, supernatant was collected and added to 0.5 ml of 50% glutathione-sepharose slurry equilibrated in NTEN buffer. Samples rotated for two

hours at 4°C and centrifuged at 1,000 rpm for one minute. Pellets were washed three times with BC300 (20 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 0.01% glycerol, 0.1% NP40 and 0.3 M KCl) and once with BC 150 (20 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 0.01% glycerol, 0.1% NP40 and 0.15 M KCl). Elution was performed with 0.5 ml of elution buffer (30 mg of glutathione and 500 µl of 1 M Tris-HCl pH 8.0 in 10 ml of final volume) per sample for 1 hour at 4°C. Finally, samples were centrifuged at 1,000 rpm and supernatant was recovered. A second elution was performed with 0.25 ml of elution buffer for 30 minutes at 4°C. Protein concentration was measured in both eluates and stored at - 80°C.

The efficiency of purification was determined by SDS-PAGE followed by 20 minutes staining with *Coomassie* blue solution (40% methanol, 10% acetic acid and 0.1% Brilliant Blue R) and several washes with destaining solution (40% methanol and 10% acetic acid). The amount of purified protein was estimated with a concentration curve of BSA, included in the same gel.

6.2.2.4. Methyltransferase assay

The *in vitro* methylation assays contained 4 µl of 5X assay buffer (250 mM Tris-HCl pH 9, 750 mM NaCl and 2.5 mM dTT), 1 µl H^3SAM (PerkinElmer España SL) as a methyl donor, 100-500 ng of the methyltransferase SMYD2 (Active Motif), 1 µg of protein substrate and up to 20 µl of H_2O. Reactions were incubated for 1 hour at 30°C.

Next, 4 µl of NuPage loading buffer (6X) and 1 µl 0.1 M dTT were added per sample and denatured at 95°C for 5 minutes. Samples were resolved on a NuPAGE® Novex® 4-12% Bis-Tris Gel (Invitrogen) at 100V for 2 hours and stained with *Coomassie* blue solution for 20 minutes and destained until sharp protein bands were visible. Then, about 20 ml of Amplify solution (Amersham) was added for 10-15 minutes. Gel was dryed in a gel dryer at 80°C under vacuum for 1 hour and exposed at -80°C O/N.

6.2.2.5. Protein-protein interaction assay

A) *In vitro* protein-protein interaction: GST pull down

In order to purify SMYD2 from mammalian cells, total soluble extracts were prepared from a 293T cell line that overexpresses SMYD2 (293T-SMYD2) using RIPA buffer. 10 mg of extract were incubated with 100 µl of 50% ANTI-FLAG M2 agarose beads (Sigma), previously equilibrated in RIPA buffer, in a total volume of 10ml. Samples were rotating O/N at 4°C. Next day, the agarose beads were extensively washed and FLAG-SMYD2 was eluted with 400 µl of BC180 (20 mM Tris-HCl pH 7.5, 0,1 mM EDTA, 0.01% glycerol, 0.1% NP40 and 0.18 M KCl) containing 8ul of 5 mg/ml FLAG peptide (Sigma).

GST fused proteins were expressed and purified from *E.Coli*. Cellular extracts were prepared as described in the protein purification section and incubated with 50 µl of 50% glutathione-sepharose for 1 hour at 4ºC and extensively washed. The purified GST fusion protein bound to the sepharose beads was finally resuspended in 50 µl NTEN buffer and 90 µl of purified FLAG-SMYD2 were added. Samples were kept rotating for 2 hours at 4°C, washed with BC180 buffer, resuspended in SDS sample buffer 2X and boiled for 5 minutes at 100°C. Proteins were resolved in a 8.5% acrylamide gel and immunoblot was performed by using an anti-FLAG antibody.

B) Co-IP of SMYD2

293T-SMYD2 cells were treated with BMP4 (10 ng/ml) for 1 hour, collected and lysed in 200 µl of IP buffer (20 mM Tris-HCl pH 7.5, 180 mM NaCl, 1 mM EDTA, 0.01% Triton X-100, 1 mM dTT, 1X PhosphoSTOP (Roche) and proteases inhibitors). About 1 mg of soluble extract was incubated with 100 µl of 50% ANTI-FLAG M2 agarose beads rotating for 2-3 hours at 4°C. and washed with IP buffer. The immunoprecipitated complexes were disrupted by boiling in SDS sample buffer and resolved in a 10% acrylamide gel. Immunoblot was performed by using an anti-phospho-SMAD1/5/8 antibody.

6.2.2.6. Chromatin immunoprecipitation (ChIP) assay

Briefly, 1×10^6 human ES cells were used for each immunoprecipitation. Cells were fixed using 1% formaldehyde, collected, resuspended in ChIP lysis buffer (1% SDS, 10 mM EDTA and 50 mM Tris-HCl, at pH 8.1) and sonicated using the Branson Digital Sonifier to generate fragments of 150–500 bp. Sonication conditions used were 50% power / 5 seconds pulse / 4 times. Soluble chromatin was diluted eightfold in ChIP RIPA buffer (10 mM Tris-HCl, at pH 7.5, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% SDS and 0.1% Na-deoxycholate) and incubated with Dynabeads Protein A (Invitrogen) coupled to the specific antibody. After incubation, the immunocomplexes were washed sequentially with low-salt buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, at pH 8.1 and 150 mM NaCl), high-salt buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, at pH 8.1, and 500 mM NaCl), LiCl buffer (0.25 M LiCl, 1% NP-40, 1% deoxycholate, 1 mM EDTA and 10 mM Tris-HCl, at pH 8.1) and TE buffer (10 mM Tris-HCl, at pH 7.5, and 1 mM EDTA). Immunocomplexes were eluted in ChIP elution buffer (1% SDS and 0.1 M $NaHCO_3$) and the crosslinking was reversed O/N at 65°C. Samples were treated with Proteinase K and RNase A, and purified using the Qiagen PCR purification kit.

6.2.3. Cell culture

6.2.3.1. Cell culture conditions

Human ES and human KiPS cell lines were derived and characterized at the CMR[B] [280-282]. All these cell lines were cultured in Matrigel (BD Biosciences) coated dishes with HES media (Knockout DMEM supplemented with 20% KSR, 1% non-essential amino acids, 50 μM 2-mercaptoethanol, 100 U/ml penicillin, 100 μg/ml streptomycin, 2 mM GlutaMAX and 10 ng/ml bFGF) conditioned for 24 hours with irradiated mouse embryonic fibroblasts (MEFs). Human ES and human KiPS cell lines were subcultured by trypsinization with 0.05% trypsin-EDTA (Invitrogen). Conditioned HES media was changed daily.

MEFs were grown in DMEM supplemented media (supplemented with 10% FBS, 100

U/ml penicillin, 100 µg/ml streptomycin, and 2 mM GlutaMAX) and mitotically inactivated by gamma irradiation.

Human foreskin fibroblasts (HFF), 293T and HepG2 cells were cultured in DMEM supplemented media and subcultured using 0.25% trypsin-EDTA (Invitrogen). Human keratinocytes (HK) were cultured in serum-free low calcium Epilife® medium (Invitrogen) and subcultured using TrypLE Express (Invitrogen).

All cell culture lines used in this work were grown and maintained in the incubator at 37°C and 5% CO_2 atmosphere conditions.

6.2.3.2. Transfection of eukaryotic cells

Transfection of DNA into 293T and ES[4] cell lines was performed using Fugene6® reagent (Roche), according to the manufacturer's protocol.

6.2.3.3. Flow cytometry analysis

- **FACS analysis**

For the OCT4 analyses by FACS we used the FIX and PERM® kit (Invitrogen). Human ES cells were trypsinized from a 10 cm plate, washed and resuspended in 100 µl of PBS-1% BSA. Then, 100 µl of Fixation medium A was added to the cell suspension and incubated for 15 minutes at RT. Cells were washed with 3 ml of PBS-1% BSA and collected by centrifugation at 2,000 rpm for 3 minutes. Next, 100 µl of PBS-1% BSA, 100 µL Permeabilization medium B and 5% of mouse serum were added into the cell pellet and incubated for 15 minutes at RT. Cells were washed with 3 ml of 1X saponin-based permeabilization buffer and collected by centrifugation at 2,000 rpm for 3 minutes. Finally, 100 µl of 1X saponin-based permeabilization buffer and 20 µl of anti-OCT3/4 (BD Bioscience) were added into the cell pellet and incubated for 30 minutes at RT. Cells were washed with 3 ml of PBS-1% BSA and collected by centrifugation at 2,000 rpm for 3 minutes. The cell pellet was resuspended in 500 µl of PBS-1% BSA. Cells were analyzed on a Gallios Flow Cytometer (Beckman Coulter, Inc) using the Kaluza® Flow Analysis Software (Beckman Coulter, Inc).

➢ Cell cycle analysis

For cell cycle analyses we used the Click-iT® EdU AlexaFluor647® Flow Cytometry Assay kit (Invitrogen). Human ES cells and EBs were plated in a 6-well plate and 2 µl of 10 mM EdU solution was added to each well for 45 minutes at 37°C. Cells/EBs were trypsinized, washed and resuspended in 100 µl of PBS-1% BSA. Then, 100 µl of Click-iT fixative solution was added into the cell suspension and incubated at RT for 15 minutes. Cells were washed with 3 ml of PBS-1% BSA and collected by centrifugation at 2,000 rpm for 3 minutes. Next, 100 µl of 1X saponin-based permeabilization buffer were added into the cell pellet and incubated for 30 minutes at RT. Cells were washed with 3 ml of 1X saponin-based permeabilization buffer and collected by centrifugation at 2,000 rpm for 3 minutes. Finally, we added 500 µl of Click-iT reaction cocktail (1X Click-iT EdU buffer additive, $CuS0_4$, and Alexa Fluor 647) and incubated for 30 minutes at RT. Cells were washed with 3 ml of 1X saponin-based permeabilization buffer and collected by centrifugation at 2,000 rpm for 3 minutes. The cell pellet was resuspended in 500 µl of saponin 1X buffer. DNA content was assessed by staining cells with the addition of DAPI solution (0.1M Tris Base pH 7.4, 150 mM NaCl, 1 mM $CaCl_2$, 0.5 mM $MgCl_2$, 0.2% BSA, 0.1% NP40, and 10 mg/ml DAPI) O/N at 4°C. Cells were analyzed on a MoFlo cell sorter (DakoCytomation) using the Kaluza® Flow Analysis Software (Beckman Coulter, Inc).

6.2.3.5. Viral production and titration

➢ Production

The production of viral particles was performed in 293T cells. The day before transfection $3.65x10^6$ cells have been plated on a 10 cm plate. Plasmids encoding the viral components were co-transfected in 293T cells: 3µg pCMV-VSVG (envelope), 8 µg psPAX2 (packaging) and 10 µg of transfer vector (for knockdown experiments we used the vector pLVTHM and for overexpression experiments the vector pWPI). Culture media was collected each 24 hours during 2 days after transfection. The virus-containing media was then filtered using a 0.22 µm filter and concentrated by ultracentrifugation during 2 hours at 19,000 g at 22°C. The viral pellet was then

resuspended in a suitable volume of PBS in order to obtain about 300-1000 fold concentrated virus.

- **Titration**

In order to obtain the best efficiency of infection with the lowest cell toxicity is necessary to determine the amount of virus obtained during the production. The day before the infection 100,000 HeLa cells were plated in each well of a 24-well plate. Infection of HeLa cells was performed by the addition of serial dilutions of virus.

The viral particle carries a GFP, encoded by the pLVTHM vectors, that can be used as a marker for infection. After 24 hours of infection HeLa cells were fixed and analyzed by FACS analysis to determine the number of GFP positive cells.

The titration of the virus, expressed as transduction units/ml (TU/ml), reflects the number of infective viral particle in a volume unit and is determined according to the following formula:

$$\text{TU/ml} = \frac{[\text{nº of HeLa cells transduced } (10^5)] \text{ x } [\text{\% of GFP positive cells}]}{\text{dilution factor}}$$

To calculate the volume of viral preparation necessary to infect a specific cell type we used the following formula:

$$\text{viral volume} = \frac{[\text{MOI}] \text{ x } [\text{nº of infected cells}]}{\text{TU/ml}}$$

6.2.3.6. Generation of stable cell lines

- **ES[4] cell lines**

ES[4] lines knockdown of SMYD2

To generate ES[4] cell lines knocked down for SMYD2 (ES[4]-shSMYD2) and control (ES[4]-shControl) we infected ES[4] cells with lentiviral particles of

pLVTHM-shSMYD2 and pLVTHM-shControl, respectively. The infection was performed using 200,000 cells suspended in 100µl of conditioned HES media for 1 hour at 37°C at and MOI of 10-15. After infection, cells were plated in a 6-well plate and later expanded to a 10 cm plate. When confluent, cells were trypsinized and resuspended in 1 ml of pre-warmed conditioned HES media and GFP positive cells were sorted using a MoFlo cell sorter (DakoCytomation) and plated back in HES conditioned media.

ES[4] lines overexpressing FLAG-SMYD2

Vectors pTP6 and pTP6-SMYD2 were linearized with PvuI, which cuts the vector on the ampicillin resistance, and 1 µg of linearized vector was transfected into ES[4] using Fugene6. After 48 hours cells were selected with 2µg/ml puromycin. Several resistant clones were selected and expanded to generate clonal ES[4] cell lines ES[4]-GFP and ES[4]-SMYD2.

- **293T cell lines**

Knockdown of SMYD2 in 293T

To generate a stable knockdown of SMYD2 in 293T cells we used the different pLVTHM constructs containing specific short-hairpins against SMYD2 (293T-shSMYD2, -shSM1, -shSM2, -shSM3, -shSM4 and -shSM5) and a short-hairpin control (293T-shControl). The infection was performed on a 50% confluent 6 cm plate at a MOI of 3. Subsequently, 293T cells were expanded to a 10 cm plate, trypsinized and resuspended in 1 ml of pre-warmed DMEM supplemented media. GFP positive cells were sorted by a MoFlo cell sorter (DakoCytomation) and plated back in DMEM supplemented media.

Overexpression of SMYD2 in 293T

To establish a stable 293T cell line overexpressing SMYD2 (293T-SMYD2) we used the pWPI lentiviral vector. A control 293T cell line was established by using a pWPI

empty vector (293T-GFP). Infection and selection of GFP positive cells were performed exactly as described above.

6.2.3.7. *In vitro* differentiation of human ES cells and treatments

➢ Embryoid body (EB) differentiation

The EB differentiation protocol was used to differentiate human ES cells into a mix of the three embryonic germ layers endoderm, mesoderm and ectoderm.

A 70-80% confluent 10 cm plate of human ES cells was trypsinized in order to obtain a single cell suspension, and 30,000-50,000 cells were plated into each well of a v-bottom low attachment 96-well plate. The 96-well plate was centrifuged at 950 g during 10 minutes and placed in the incubator for the next 24-48 hours to allow the EB formation. EBs were collected and transferred to low-attachment bacterial plate. The media was switched to conditioned HES media without bFGF and placed in the incubator for one more day. Finally EBs were placed into a gelatine coated dish in the presence of EB media (Knockout DMEM, 20% FBS, 100 U/ml penicillin, 100 μg/ml streptomycin and 2 mM Glutamax) to differentiate into the three germ layers. Media was changed every two days.

➢ Monolayer differentiation

ES[4] cells were grown to 60-70% confluency and HES conditioned media was changed to EB media. Media was changed every two days.

➢ Neural differentiation

The dual SMAD inhibition protocol [221] was used to differentiate human ES cells into neural precursors (NP) cells. A confluent 6-well plate of human ES cells was disaggregated to a single cell suspension using Accutase (Lab Clinics SA). Cells were filtered to remove clumps, washed twice in mTeSR1 media (Stemcell Technologies), and counted. The cell pellet was resuspended in mTeSR1 media containing 10 μM Rock Inhibitor (RI) (Santa Cruz Biotechnology) and 200,000 cells were plated per well

of a 6-well plate. Media was changed daily until reaching 70-80% confluency. At this point (day 0) the media was changed to KSR media (Knockout DMEM, 15% KSR, 2-mercaptoethanol, 100 U/ml penicillin, 100 μg/ml streptomycin and 2 mM Glutamax) supplemented with 10 μM SB431542 (Tocris) and 200 ng/ml Noggin (R&D Systems).. On day 4 media was changed to 3:1 ratio of KSR:N2 media (DMEM/F12, 100 U/ml penicillin, 100 μg/ml streptomycin and 2 mM Glutamax and 1X N-2 supplement (Invitrogene)) supplemented with 10 μM SB431542 and 200 ng/ml Noggin. On day 6, media was changed to 1:1 ratio of KSR:N2 media supplemented with 10 μM SB431542 and 200 ng/ml Noggin. On day 8, media was changed to 1:3 ratio of KSR:N2 media supplemented with 10 μM SB431542 and 200 ng/ml Noggin, until day 10 of differentiation.

- **BMP2 and Activin A treatment**

ES[4] cells were grown in HES contidioned media to 60-70% confluency and 200 ng/ml of BMP2 (Abnova) or 100 ng/ml of Activin A (Abnova) were added to the media. Media was replaced every day for 4 days.

6.2.3.8. Luciferase assay

About 60,000 293T cells/well were plated in a 12-well plate following a density. The day after, the media was changed 1 hour before transfection. 293T cells were transfected with 0.05 μg pmirGLO constructs, 0.25 ng pRL-CMV (*Renilla* luciferase) and 0.35 μg of pBS-SK+ to reach a total DNA concentration optimal for transfection. After 48 hours cells were lysed in Passive Lysis Buffer (Promega) for 15 minutes at RT. Luciferase assays were performed using the Dual-Luciferase® Reporter Assay System (Promega), following the manufacturer instructions. Luciferase activity of transfected 293T cells was analyzed using a Luminometer (Berthold Lumat LB 9501) and normalized to *Renilla* activity.

6.2.4. Zebrafish experiments

6.2.4.1. Zebrafish culture conditions

Zebrafish wild-type AB strain were maintained at 28°C in a recirculating aquaculture system equipped with carbon filtration, ultraviolet light sterilizers and biofiltration (Aquatic Habitats). Fertilized eggs obtained from mating pairs of adult zebrafish were cultured in Embryo Medium composed of 1 ml of Stock 1 (8 g NaCl, 0.4 g KCl in 100 ml H_2O), 0.1 ml of Stock 2 (0.358 g Na_2HPO_4, 0.6 g KH_2PO_4 in 100 ml H_2O), 1 ml of Stock 4 (0.72 g $CaCl_2$ in 50 ml H_2O), 1 ml of Stock 5 (1.23 g $MgSO_4$ x $7H_2O$ in 50 ml H_2O) and 1 ml of Stock 6 (0.35 g $NaHCO_3$ in 10 ml H_2O) in 95.9 ml of H_2O containing streptomycin and penicillin.

6.2.4.2. Zebrafish embryos injection

Antisense morpholino-oligonucleotides (MO) were supplied by Gene Tools LCC (Philomath, OR). The sequence of *smyd2a* splice-blocking MO (smyd2a-MO) was 5' TTATAAGGAGCGCTGACCTGGTAAA 3', which was designed to hybridize at the splice site of exon1-intron1 junction. The sequence of a standard control MO (Control-MO) was 5' CCTCTTACCTCAGTTACAATTTATA 3', and was used as an injection control.

FLAG-tagged SMYD2 cDNA was cloned into pCS2 vector for mRNA synthesis using the mMESSAGE mMACHINE kit (Ambion, AMI344, Life Technologies, USA) following the manufacturer's instructions.

MOs and mRNA were diluted in Danieau buffer and were injected at 1-2 cell stage of fertilized wild-type zebrafish eggs using a microinjector. Rescue experiments were performed by co-injecting 30 pg of synthesized SMYD2 mRNA with 0,35 mM smyd2a-MO and with 0,35 mM Control-MO as a control.

6.2.4.3. *In situ* hybridization (ISH)

Antisense RNA probes were synthesized using the DIG RNA Labeling Kit SP6/T7 (Roche). Embryos were collected at different stages (6hpf, 24hpf and 48hpf) and fixed with 4% PFA O/N at 4°C. The next day embryos were dehydrated with an increasing methanol-containing solution. On day 1, after intensive washes with decreasing concentrations of methanol-containing buffer, embryos were hybridized with the probe in a 50% formamide buffer O/N at 70°C. On day 2, after several washes with decreasing concentrations of formamide buffer, embryos were incubated with an anti-digoxygenin antibody conjugated to alkaline phosphatase (Roche Molecular Biochemicals) O/N at 4°C. On day 3, after several additional washes, embryos were treated with an alkaline solution containing nitro blue tetrazolium (NBT) and 5-Bromo-4-chloro-3-indolyl phosphate (BCIP) at RT in the dark. Once color is developed, the reaction is stopped at 4°C. Finally, embryos were fixed with 4% PFA O/N at RT.

6.2.4.3. RNA extraction

For each RNA extraction 20-30 zebrafish embryos were collected. Total RNA extraction and quantification of mRNA levels was performed as described in the real time PCR analysis section.

7. BIBLIOGRAPHY

1. Edmondson DG, Roth SY (1996) Chromatin and transcription. FASEB J 10: 1173–1182.
2. Felsenfeld G, Groudine M (2003) Controlling the double helix. Nature 421: 448–453. doi:10.1038/nature01411.
3. Filion GJ, van Bemmel JG, Braunschweig U, Talhout W, Kind J, et al. (2010) Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell 143: 212–224. doi:10.1016/j.cell.2010.09.009.
4. Turner BM (2002) Cellular memory and the histone code. Cell 111: 285–291.
5. Banaszynski LA, Allis CD, Lewis PW (2010) Histone Variants in Metazoan Development. Developmental Cell 19: 662–674. doi:10.1016/j.devcel.2010.10.014.
6. Kornberg RD, Lorch Y (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell 98: 285–294.
7. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389: 251–260. doi:10.1038/38444.
8. Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. J Mol Biol 319: 1097–1113. doi:10.1016/S0022-2836(02)00386-8.
9. Arents G, Burlingame RW, Wang BC, Love WE, Moudrianakis EN (1991) The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. Proc Natl Acad Sci USA 88: 10148–10152.
10. Wolffe AP, Guschin D (2000) Review: Chromatin Structural Features and Targets That Regulate Transcription. Journal of Structural Biology 129: 102–122. doi:10.1006/jsbi.2000.4217.
11. Khorasanizadeh S (2004) The nucleosome: from genomic organization to genomic regulation. Cell 116: 259–272.
12. Vitolo JM, Thiriet C, Hayes JJ (2000) The H3-H4 N-terminal tail domains are the primary mediators of transcription factor IIIA access to 5S DNA within a nucleosome. Molecular and Cellular Biology 20: 2167–2175.
13. Marmorstein R (2001) Protein modules that manipulate histone tails for chromatin regulation. Nat Rev Mol Cell Biol 2: 422–432. doi:10.1038/35073047.
14. Woodcock CL, Dimitrov S (2001) Higher-order structure of chromatin and chromosomes. Current Opinion in Genetics & Development 11: 130–135.
15. Chen T, Li E (2006) Establishment and maintenance of DNA methylation patterns in mammals. Curr Top Microbiol Immunol 301: 179–201.
16. Bird A (2002) DNA methylation patterns and epigenetic memory. Genes &

Development 16: 6–21. doi:10.1101/gad.947102.

17. Buryanov YI, Shevchuk TV (2005) DNA methyltransferases and structural-functional specificity of eukaryotic DNA modification. Biochemistry Mosc 70: 730–742.

18. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, et al. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324: 930–935. doi:10.1126/science.1170116.

19. Ito S, Shen L, Dai Q, Wu SC, Collins LB, et al. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science 333: 1300–1303. doi:10.1126/science.1210597.

20. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, et al. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature 466: 1129–1133. doi:10.1038/nature09303.

21. Wu H, D'Alessio AC, Ito S, Wang Z, Cui K, et al. (2011) Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes & Development 25: 679–684. doi:10.1101/gad.2036011.

22. Yamaguchi S, Hong K, Liu R, Inoue A, Shen L, et al. (2013) Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. Cell Res 23: 329–339. doi:10.1038/cr.2013.22.

23. Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69: 915–926.

24. Monk M (1995) Epigenetic programming of differential gene expression in development and evolution. Dev Genet 17: 188–197. doi:10.1002/dvg.1020170303.

25. Yildirim O, Li R, Hung J-H, Chen PB, Dong X, et al. (2011) Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell 147: 1498–1510. doi:10.1016/j.cell.2011.11.054.

26. Li E (2002) Chromatin modification and epigenetic reprogramming in mammalian development. Nat Rev Genet 3: 662–673. doi:10.1038/nrg887.

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

28. McCabe MT, Brandes JC, Vertino PM (2009) Cancer DNA methylation: molecular mechanisms and clinical implications. Clin Cancer Res 15: 3927–3937. doi:10.1158/1078-0432.CCR-08-2784.

29. Kouzarides T (2007) Chromatin Modifications and Their Function. Cell 128: 693–705. doi:10.1016/j.cell.2007.02.005.

30. Jenuwein T, Allis CD (2001) Translating the histone code. Science 293: 1074–1080. doi:10.1126/science.1063127.

31. Strahl BD, Allis CD (2000) The language of covalent histone modifications. Nature 403: 41–45. doi:10.1038/47412.

32. Kimura A, Matsubara K, Horikoshi M (2005) A decade of histone acetylation:

marking eukaryotic chromosomes with specific codes. J Biochem 138: 647–662. doi:10.1093/jb/mvi184.

33. Umlauf D, Goto Y, Feil R (2004) Site-specific analysis of histone methylation and acetylation. Methods Mol Biol 287: 99–120. doi:10.1385/1-59259-828-5:099.

34. Verdone L, Caserta M, Di Mauro E (2005) Role of histone acetylation in the control of gene expression. Biochem Cell Biol 83: 344–353. doi:10.1139/o05-041.

35. Verdone L, Agricola E, Caserta M, Di Mauro E (2006) Histone acetylation in gene regulation. Brief Funct Genomic Proteomic 5: 209–221. doi:10.1093/bfgp/ell028.

36. de Ruijter AJM, van Gennip AH, Caron HN, Kemp S, van Kuilenburg ABP (2003) Histone deacetylases (HDACs): characterization of the classical HDAC family. Biochem J 370: 737–749. doi:10.1042/BJ20021321.

37. Agalioti T, Chen G, Thanos D (2002) Deciphering the transcriptional histone acetylation code for a human gene. Cell 111: 381–392.

38. Yang X-J (2004) The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. Nucleic Acids Research 32: 959–976. doi:10.1093/nar/gkh252.

39. Yang X-J, Seto E (2008) Lysine acetylation: codified crosstalk with other posttranslational modifications. Molecular Cell 31: 449–461. doi:10.1016/j.molcel.2008.07.002.

40. Gregoretti IV, Lee Y-M, Goodson HV (2004) Molecular evolution of the histone deacetylase family: functional implications of phylogenetic analysis. J Mol Biol 338: 17–31. doi:10.1016/j.jmb.2004.02.006.

41. Taunton J, Hassig CA, Schreiber SL (1996) A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science 272: 408–411.

42. Weidle UH, Grossmann A (2000) Inhibition of histone deacetylases: a new strategy to target epigenetic modifications for anticancer treatment. Anticancer Res 20: 1471–1485.

43. Knoepfler PS, Eisenman RN (1999) Sin meets NuRD and other tails of repression. Cell 99: 447–450.

44. Blander G, Guarente L (2004) The Sir2 family of protein deacetylases. Annu Rev Biochem 73: 417–435. doi:10.1146/annurev.biochem.73.011303.073651.

45. Nusinzon I, Horvath CM (2005) Histone deacetylases as transcriptional activators? Role reversal in inducible gene regulation. Sci STKE 2005: re11. doi:10.1126/stke.2962005re11.

46. Ropero S, Esteller M (2007) The role of histone deacetylases (HDACs) in human cancer. Mol Oncol 1: 19–25. doi:10.1016/j.molonc.2007.01.001.

47. Zhang Y, Reinberg D (2001) Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes & Development 15: 2343–2360. doi:10.1101/gad.927301.

48. Lachner M, Jenuwein T (2002) The many faces of histone lysine methylation.

Curr Opin Cell Biol 14: 286–298.

49. Kouzarides T (2002) Histone methylation in transcriptional control. Current Opinion in Genetics & Development 12: 198–209.

50. Di Lorenzo A, Bedford MT (2011) Histone arginine methylation. FEBS LETTERS 585: 2024–2031. doi:10.1016/j.febslet.2010.11.010.

51. Hyllus D, Stein C, Schnabel K, Schiltz E, Imhof A, et al. (2007) PRMT6-mediated methylation of R2 in histone H3 antagonizes H3 K4 trimethylation. Genes & Development 21: 3369–3380. doi:10.1101/gad.447007.

52. Okada Y, Feng Q, Lin Y, Jiang Q, Li Y, et al. (2005) hDOT1L links histone methylation to leukemogenesis. Cell 121: 167–178. doi:10.1016/j.cell.2005.02.020.

53. Tschiersch B, Hofmann A, Krauss V, Dorn R, Korge G, et al. (1994) The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. EMBO J 13: 3822–3831.

54. Stassen MJ, Bailey D, Nelson S, Chinwalla V, Harte PJ (1995) The Drosophila trithorax proteins contain a novel variant of the nuclear receptor type DNA binding domain and an ancient conserved motif found in other chromosomal proteins. Mech Dev 52: 209–223.

55. Sims RJ, Nishioka K, Reinberg D (2003) Histone lysine methylation: a signature for chromatin function. Trends Genet 19: 629–639. doi:10.1016/j.tig.2003.09.007.

56. Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, et al. (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature 406: 593–599. doi:10.1038/35020506.

57. Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, et al. (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature 410: 120–124. doi:10.1038/35065138.

58. Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, et al. (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science 298: 1039–1043. doi:10.1126/science.1076997.

59. Allis CD (2007) Epigenetics. CSHL Press. p.

60. Chou R-H, Yu Y-L, Hung M-C (2011) The roles of EZH2 in cell lineage commitment. Am J Transl Res 3: 243–250.

61. Sauvageau M, Sauvageau G (2008) Polycomb group genes: keeping stem cell activity in balance. PLoS Biol 6: e113. doi:10.1371/journal.pbio.0060113.

62. Leeb M, Pasini D, Novatchkova M, Jaritz M, Helin K, et al. (2010) Polycomb complexes act redundantly to repress genomic repeats and genes. Genes & Development 24: 265–276. doi:10.1101/gad.544410.

63. Tavares L, Dimitrova E, Oxley D, Webster J, Poot R, et al. (2012) RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. Cell 148: 664–678. doi:10.1016/j.cell.2011.12.029.

64. Farcas AM, Blackledge NP, Sudbery I, Long HK, McGouran JF, et al. (2012)

KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. Elife 1: e00205. doi:10.7554/eLife.00205.

65. He J, Shen L, Wan M, Taranova O, Wu H, et al. (2013) Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. Nat Cell Biol 15: 373–384. doi:10.1038/ncb2702.

66. Wu X, Johansen JV, Helin K (2013) Fbxl10/Kdm2b Recruits Polycomb Repressive Complex 1 to CpG Islands and Regulates H2A Ubiquitylation. Molecular Cell. doi:10.1016/j.molcel.2013.01.016.

67. van Ingen H, van Schaik FMA, Wienk H, Ballering J, Rehmann H, et al. (2008) Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. Structure/Folding and Design 16: 1245–1256. doi:10.1016/j.str.2008.04.015.

68. Guenther MG, Jenner RG, Chevalier B, Nakamura T, Croce CM, et al. (2005) Global and Hox-specific roles for the MLL1 methyltransferase. Proc Natl Acad Sci USA 102: 8603–8608. doi:10.1073/pnas.0503072102.

69. Orlando V (2003) Polycomb, epigenomes, and control of cell identity. Cell 112: 599–606.

70. Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, et al. (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell 123: 581–592. doi:10.1016/j.cell.2005.10.023.

71. Martin DGE, Grimes DE, Baetz K, Howe L (2006) Methylation of histone H3 mediates the association of the NuA3 histone acetyltransferase with chromatin. Molecular and Cellular Biology 26: 3018–3028. doi:10.1128/MCB.26.8.3018-3028.2006.

72. Barrand S, Andersen IS, Collas P (2010) Promoter-exon relationship of H3 lysine 9, 27, 36 and 79 methylation on pluripotency-associated genes. Biochem Biophys Res Commun 401: 611–617. doi:10.1016/j.bbrc.2010.09.116.

73. Hu M, Sun XJ, Zhang YL, Kuang Y, Hu CQ, et al. (2010) Histone H3 lysine 36 methyltransferase Hypb/Setd2 is required for embryonic vascular remodeling. Proceedings of the National Academy of Sciences 107: 2956–2961. doi:10.1073/pnas.0915033107.

74. Wagner EJ, Carpenter PB (2012) Understanding the language of Lys36 methylation at histone H3. Nat Rev Mol Cell Biol 13: 115–126. doi:10.1038/nrm3274.

75. Congdon LM, Houston SI, Veerappan CS, Spektor TM, Rice JC (2010) PR-Set7-mediated monomethylation of histone H4 lysine 20 at specific genomic regions induces transcriptional repression. J Cell Biochem 110: 609–619. doi:10.1002/jcb.22570.

76. Lee J, Zhou P (2010) SETting the Clock for Histone H4 Monomethylation. Molecular Cell 40: 345–346. doi:10.1016/j.molcel.2010.10.033.

77. Pei H, Zhang L, Luo K, Qin Y, Chesi M, et al. (2011) MMSET regulates histone H4K20 methylation and 53BP1 accumulation at DNA damage sites. Nature 470: 124–128. doi:10.1038/nature09658.

78. Kapoor-Vazirani P, Kagey JD, Vertino PM (2011) SUV420H2-Mediated H4K20 Trimethylation Enforces RNA Polymerase II Promoter-Proximal Pausing by Blocking hMOF-Dependent H4K16 Acetylation. Molecular and Cellular Biology 31: 1594–1609. doi:10.1128/MCB.00524-10.

79. Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, et al. (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell 119: 941–953. doi:10.1016/j.cell.2004.12.012.

80. Forneris F, Binda C, Vanoni MA, Mattevi A, Battaglioli E (2005) Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process. FEBS LETTERS 579: 2203–2207. doi:10.1016/j.febslet.2005.03.015.

81. Tsukada Y-I, Zhang Y (2006) Purification of histone demethylases from HeLa cells. Methods 40: 318–326. doi:10.1016/j.ymeth.2006.06.024.

82. Yamane K, Toumazou C, Tsukada Y-I, Erdjument-Bromage H, Tempst P, et al. (2006) JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. Cell 125: 483–495. doi:10.1016/j.cell.2006.03.027.

83. Whetstine JR, Nottke A, Lan F, Huarte M, Smolikov S, et al. (2006) Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. Cell 125: 467–481. doi:10.1016/j.cell.2006.03.028.

84. Christensen J, Agger K, Cloos PAC, Pasini D, Rose S, et al. (2007) RBP2 belongs to a family of demethylases, specific for tri-and dimethylated lysine 4 on histone 3. Cell 128: 1063–1076. doi:10.1016/j.cell.2007.02.003.

85. Smith ER, Lee MG, Winter B, Droz NM, Eissenberg JC, et al. (2008) Drosophila UTX is a histone H3 Lys27 demethylase that colocalizes with the elongating form of RNA polymerase II. Molecular and Cellular Biology 28: 1041–1046. doi:10.1128/MCB.01504-07.

86. Burgold T, Spreafico F, De Santa F, Totaro MG, Prosperini E, et al. (2008) The histone H3 lysine 27-specific demethylase Jmjd3 is required for neural commitment. PLoS ONE 3: e3034. doi:10.1371/journal.pone.0003034.

87. Liu W, Tanasa B, Tyurina OV, Zhou TY, Gassmann R, et al. (2010) PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. Nature 466: 508–512. doi:10.1038/nature09272.

88. Chang B, Chen Y, Zhao Y, Bruick RK (2007) JMJD6 is a histone arginine demethylase. Science 318: 444–447. doi:10.1126/science.1145801.

89. Pedersen MT, Helin K (2010) Histone demethylases in development and disease. Trends Cell Biol 20: 662–671. doi:10.1016/j.tcb.2010.08.011.

90. Spitz F, Furlong EEM (2012) Transcription factors: from enhancer binding to developmental control. Nat Rev Genet 13: 613–626. doi:10.1038/nrg3207.

91. Luger K (2006) Dynamic nucleosomes. Chromosome Res 14: 5–16. doi:10.1007/s10577-005-1026-1.

92. Rando OJ, Ahmad K (2007) Rules and regulation in the primary structure of chromatin. Curr Opin Cell Biol 19: 250–256. doi:10.1016/j.ceb.2007.04.006.

93. Becker PB, Hörz W (2002) ATP-dependent nucleosome remodeling. Annu Rev Biochem 71: 247–273. doi:10.1146/annurev.biochem.71.110601.135400.

94. Vignali M, Hassan AH, Neely KE, Workman JL (2000) ATP-dependent chromatin-remodeling complexes. Molecular and Cellular Biology 20: 1899–1910.

95. Längst G, Becker PB (2001) Nucleosome mobilization and positioning by ISWI-containing chromatin-remodeling factors. J Cell Sci 114: 2561–2568.

96. Bouazoune K, Mitterweger A, Längst G, Imhof A, Akhtar A, et al. (2002) The dMi-2 chromodomains are DNA binding modules important for ATP-dependent nucleosome mobilization. EMBO J 21: 2430–2440. doi:10.1093/emboj/21.10.2430.

97. Fry CJ, Peterson CL (2002) Transcription. Unlocking the gates to gene expression. Science 295: 1847–1848. doi:10.1126/science.1070260.

98. Haynes SR, Dollard C, Winston F, Beck S, Trowsdale J, et al. (1992) The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins. Nucleic Acids Research 20: 2603.

99. Marmorstein R, Berger SL (2001) Structure and function of bromodomains in chromatin-regulating complexes. Gene 272: 1–9.

100. Jacobs SA, Khorasanizadeh S (2002) Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science 295: 2080–2083. doi:10.1126/science.1069473.

101. Paro R, Hogness DS (1991) The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. Proc Natl Acad Sci USA 88: 263–267.

102. Clapier CR, Cairns BR (2009) The biology of chromatin remodeling complexes. Annu Rev Biochem 78: 273–304. doi:10.1146/annurev.biochem.77.062706.153223.

103. Flanagan JF, Mi L-Z, Chruszcz M, Cymborowski M, Clines KL, et al. (2005) Double chromodomains cooperate to recognize the methylated histone H3 tail. Nature 438: 1181–1185. doi:10.1038/nature04290.

104. Liu Y, Chen W, Gaudet J, Cheney MD, Roudaia L, et al. (2007) Structural basis for recognition of SMRT/N-CoR by the MYND domain and its contribution to AML1/ETO's activity. Cancer Cell 11: 483–497. doi:10.1016/j.ccr.2007.04.010.

105. Liu H-P, Chung P-J, Liang C-L, Chang Y-S (2011) The MYND domain-containing protein BRAM1 inhibits lymphotoxin beta receptor-mediated signaling through affecting receptor oligomerization. Cell Signal 23: 80–88. doi:10.1016/j.cellsig.2010.08.006.

106. Gottlieb PD, Pierce SA, Sims RJ, Yamagishi H, Weihe EK, et al. (2002) Bop encodes a muscle-restricted protein containing MYND and SET domains and is essential for cardiac differentiation and morphogenesis. Nat Genet 31: 25–32. doi:10.1038/ng866.

107. Li D, Niu Z, Yu W, Qian Y, Wang Q, et al. (2009) SMYD1, the myogenic activator, is a direct target of serum response factor and myogenin. Nucleic Acids Research 37: 7059–7071. doi:10.1093/nar/gkp773.

108. Tan X, Rotllant J, Li H, De Deyne P, DeDeyne P, et al. (2006) SmyD1, a

histone methyltransferase, is required for myofibril organization and muscle contraction in zebrafish embryos. Proc Natl Acad Sci USA 103: 2713–2718. doi:10.1073/pnas.0509503103.

109. Hamamoto R, Furukawa Y, Morita M, Iimura Y, Silva FP, et al. (2004) SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. Nat Cell Biol 6: 731–740. doi:10.1038/ncb1151.

110. Fujii T, Tsunesumi S-I, Yamaguchi K, Watanabe S, Furukawa Y (2011) Smyd3 is required for the development of cardiac and skeletal muscle in zebrafish. PLoS ONE 6: e23491. doi:10.1371/journal.pone.0023491.

111. Hu L, Zhu YT, Qi C, Zhu Y-J (2009) Identification of Smyd4 as a potential tumor suppressor gene involved in breast cancer development. Cancer Res 69: 4067–4072. doi:10.1158/0008-5472.CAN-08-4097.

112. Thompson EC, Travers AA (2008) A Drosophila Smyd4 Homologue Is a Muscle-Specific Transcriptional Modulator Involved in Development. PLoS ONE 3: e3008. doi:10.1371/journal.pone.0003008.t001.

113. Abu-Farha M, Lanouette S, Elisma F, Tremblay V, Butson J, et al. (2011) Proteomic analyses of the SMYD family interactomes identify HSP90 as a novel target for SMYD2. Journal of Molecular Cell Biology 3: 301–308. doi:10.1093/jmcb/mjr025.

114. Xu S, Zhong C, Zhang T, Ding J (2011) Structure of human lysine methyltransferase Smyd2 reveals insights into the substrate divergence in Smyd proteins. Journal of Molecular Cell Biology 3: 293–300. doi:10.1093/jmcb/mjr015.

115. Wang L, Li L, Zhang H, Luo X, Dai J, et al. (2011) Structure of Human SMYD2 Protein Reveals the Basis of p53 Tumor Suppressor Methylation. Journal of Biological Chemistry 286: 38725–38737. doi:10.1074/jbc.M111.262410.

116. Ferguson AD, Larsen NA, Howard T, Pollard H, Green I, et al. (2011) Structural Basis of Substrate Methylation and Inhibition of SMYD2. Structure/Folding and Design: 1–12. doi:10.1016/j.str.2011.06.011.

117. Sirinupong N, Brunzelle J, Ye J, Pirzada A, Nico L, et al. (2010) Crystal structure of cardiac-specific histone methyltransferase SmyD1 reveals unusual active site architecture. Journal of Biological Chemistry 285: 40635–40644. doi:10.1074/jbc.M110.168187.

118. Sirinupong N, Brunzelle J, Doko E, Yang Z (2011) Structural insights into the autoinhibition and posttranslational activation of histone methyltransferase SmyD3. J Mol Biol 406: 149–159. doi:10.1016/j.jmb.2010.12.014.

119. Brown M, Sims R, Gottlieb P, Tucker P (2006) Molecular Cancer | Full text | Identification and characterization of Smyd2: a split SET/MYND domain-containing histone H3 lysine 36-specific methyltransferase that interacts with the Sin3 histone deacetylase complex. Mol Cancer.

120. Keogh M-C, Kurdistani SK, Morris SA, Ahn SH, Podolny V, et al. (2005) Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell 123: 593–605. doi:10.1016/j.cell.2005.10.025.

121. Joshi AA, Struhl K (2005) Eaf3 chromodomain interaction with methylated H3-

K36 links histone deacetylation to Pol II elongation. Molecular Cell 20: 971–978. doi:10.1016/j.molcel.2005.11.021.

122. Abu-Farha M, Lambert J-P, Al-Madhoun AS, Elisma F, Skerjanc IS, et al. (2008) The tale of two domains: proteomics and genomics analysis of SMYD2, a new histone methyltransferase. Mol Cell Proteomics 7: 560–572. doi:10.1074/mcp.M700271-MCP200.

123. Wu J, Cheung T, Grande C, Ferguson AD, Zhu X, et al. (2011) Biochemical Characterization of Human SET and MYND Domain-Containing Protein 2 Methyltransferase. Biochemistry 50: 6488–6497. doi:10.1021/bi200725p.

124. Huang J, Perez-Burgos L, Placek BJ, Sengupta R, Richter M, et al. (2006) Repression of p53 activity by Smyd2-mediated methylation. Nature 444: 629–632. doi:10.1038/nature05287.

125. Scoumanne A, Chen X (2008) Protein methylation: a new mechanism of p53 tumor suppressor regulation. Histol Histopathol 23: 1143–1149.

126. Saddic LA, West LE, Aslanian A, Yates JR, Rubin SM, et al. (2010) Methylation of the Retinoblastoma Tumor Suppressor by SMYD2. Journal of Biological Chemistry 285: 37733–37740. doi:10.1074/jbc.M110.137612.

127. Jiang Y, Sirinupong N, Brunzelle J, Yang Z (2011) Crystal Structures of Histone and p53 Methyltransferase SmyD2 Reveal a Conformational Flexibility of the Autoinhibitory C-Terminal Domain. PLoS ONE 6: e21640. doi:10.1371/journal.pone.0021640.t001.

128. Cho H-S, Hayami S, Toyokawa G, Maejima K, Yamane Y, et al. (2012) RB1 methylation by SMYD2 enhances cell cycle progression through an increase of RB1 phosphorylation. Neoplasia 14: 476–486.

129. Diehl F, Brown MA, van Amerongen MJ, Novoyatleva T, Wietelmann A, et al. (2010) Cardiac Deletion of Smyd2 Is Dispensable for Mouse Heart Development. PLoS ONE 5: e9748. doi:10.1371/journal.pone.0009748.g006.

130. Komatsu S, Imoto I, Tsuda H, Kozaki K-I, Muramatsu T, et al. (2009) Overexpression of SMYD2 relates to tumor cell proliferation and malignant outcome of esophageal squamous cell carcinoma. Carcinogenesis 30: 1139–1146. doi:10.1093/carcin/bgp116.

131. Zuber J, Rappaport AR, Luo W, Wang E, Chen C, et al. (2011) An integrated approach to dissecting oncogene addiction implicates a Myb-coordinated self-renewal program as essential for leukemia maintenance. Genes & Development 25: 1628–1640. doi:10.1101/gad.17269211.

132. Donlin LT, Andresen C, Just S, Rudensky E, Pappas CT, et al. (2012) Smyd2 controls cytoplasmic lysine methylation of Hsp90 and myofilament organization. Genes & Development 26: 114–119. doi:10.1101/gad.177758.111.

133. Voelkel T, Andresen C, Unger A, Just S, Rottbauer W, et al. (2012) Lysine methyltransferase Smyd2 regulates Hsp90-mediated protection of the sarcomeric titin springs and cardiac function. Biochim Biophys Acta. doi:10.1016/j.bbamcr.2012.09.012.

134. Sun X-J, Xu P-F, Zhou T, Hu M, Fu C-T, et al. (2008) Genome-Wide Survey and Developmental Expression Mapping of Zebrafish SET Domain-Containing

Genes. PLoS ONE 3: e1499. doi:10.1371/journal.pone.0001499.t002.

135. Evans MJ, Kaufman MH (1981) Establishment in culture of pluripotential cells from mouse embryos. Nature 292: 154–156.

136. Strelchenko N, Verlinsky O, Kukharenko V, Verlinsky Y (2004) Morula-derived human embryonic stem cells. Reprod Biomed Online 9: 623–629.

137. Strelchenko N, Verlinsky Y (2006) Embryonic stem cells from morula. Meth Enzymol 418: 93–108. doi:10.1016/S0076-6879(06)18006-4.

138. Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126: 663–676. doi:10.1016/j.cell.2006.07.024.

139. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, et al. (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122: 947–956. doi:10.1016/j.cell.2005.08.020.

140. Chew J-L, Loh Y-H, Zhang W, Chen X, Tam W-L, et al. (2005) Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. Molecular and Cellular Biology 25: 6031–6046. doi:10.1128/MCB.25.14.6031-6046.2005.

141. Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, et al. (1998) Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell 95: 379–391.

142. Yuan H, Corbi N, Basilico C, Dailey L (1995) Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. Genes & Development 9: 2635–2645.

143. Chambers I, Colby D, Robertson M, Nichols J, Lee S, et al. (2003) Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell 113: 643–655.

144. Chambers I, Smith A (2004) Self-renewal of teratocarcinoma and embryonic stem cells. Oncogene 23: 7150–7160. doi:10.1038/sj.onc.1207930.

145. Niwa H (2007) How is pluripotency determined and maintained? Development 134: 635–646. doi:10.1242/dev.02787.

146. Rodda DJ, Chew J-L, Lim L-H, Loh Y-H, Wang B, et al. (2005) Transcriptional regulation of nanog by OCT4 and SOX2. J Biol Chem 280: 24731–24737. doi:10.1074/jbc.M502573200.

147. Liang J, Wan M, Zhang Y, Gu P, Xin H, et al. (2008) Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. Nat Cell Biol 10: 731–739. doi:10.1038/ncb1736.

148. Azuara V, Perry P, Sauer S, Spivakov M, Jørgensen HF, et al. (2006) Chromatin signatures of pluripotent cell lines. Nat Cell Biol 8: 532–538. doi:10.1038/ncb1403.

149. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, et al. (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125: 315–326. doi:10.1016/j.cell.2006.02.041.

150. Boyer LA, Mathur D, Jaenisch R (2006) Molecular control of pluripotency.

Current Opinion in Genetics & Development 16: 455–462. doi:10.1016/j.gde.2006.08.009.

151. Spivakov M, Fisher AG (2007) Epigenetic signatures of stem-cell identity. Nat Rev Genet 8: 263–271. doi:10.1038/nrg2046.

152. Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, et al. (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441: 349–353. doi:10.1038/nature04733.

153. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, et al. (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. Cell 125: 301–313. doi:10.1016/j.cell.2006.02.043.

154. Pasini D, Bracken AP, Hansen JB, Capillo M, Helin K (2007) The polycomb group protein Suz12 is required for embryonic stem cell differentiation. Molecular and Cellular Biology 27: 3769–3779. doi:10.1128/MCB.01432-06.

155. Vastenhouw NL, Schier AF (2012) Bivalent histone modifications in early embryogenesis. Curr Opin Cell Biol 24: 374–386. doi:10.1016/j.ceb.2012.03.009.

156. Stock JK, Giadrossi S, Casanova M, Brookes E, Vidal M, et al. (2007) Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat Cell Biol 9: 1428–1435. doi:10.1038/ncb1663.

157. Jiang H, Shukla A, Wang X, Chen W-Y, Bernstein BE, et al. (2011) Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. Cell 144: 513–525. doi:10.1016/j.cell.2011.01.020.

158. Adamo A, Sesé B, Boue S, Castaño J, Paramonov I, et al. (2011) LSD1 regulates the balance between self-renewal and differentiation in human embryonic stem cells. Nat Cell Biol 13: 652–660. doi:10.1038/ncb2246.

159. Ding J, Xu H, Faiola F, Ma'ayan A, Wang J (2012) Oct4 links multiple epigenetic pathways to the pluripotency network. Cell Res 22: 155–167. doi:10.1038/cr.2011.179.

160. Landeira D, Sauer S, Poot R, Dvorkina M, Mazzarella L, et al. (2010) Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. Nat Cell Biol 12: 618–624. doi:10.1038/ncb2065.

161. Pasini D, Cloos PAC, Walfridsson J, Olsson L, Bukowski J-P, et al. (2010) JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature 464: 306–310. doi:10.1038/nature08788.

162. Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, et al. (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature 464: 1082–1086. doi:10.1038/nature08924.

163. Vastenhouw NL, Zhang Y, Woods IG, Imam F, Regev A, et al. (2010) Chromatin signature of embryonic pluripotency is established during genome activation. Nature 464: 922–926. doi:10.1038/nature08866.

164. Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, et al. (2008) Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet 4: e1000242. doi:10.1371/journal.pgen.1000242.

165. Agger K, Cloos PAC, Christensen J, Pasini D, Rose S, et al. (2007) UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature 449: 731–734. doi:10.1038/nature06145.

166. Lan F, Bayliss PE, Rinn JL, Whetstine JR, Wang JK, et al. (2007) A histone H3 lysine 27 demethylase regulates animal posterior development. Nature 449: 689–694. doi:10.1038/nature06192.

167. Ezhkova E, Pasolli HA, Parker JS, Stokes N, Su I-H, et al. (2009) Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. Cell 136: 1122–1135. doi:10.1016/j.cell.2008.12.043.

168. Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, et al. (2008) Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Molecular Cell 30: 755–766. doi:10.1016/j.molcel.2008.05.007.

169. Oktaba K, Gutiérrez L, Gagneur J, Girardot C, Sengupta AK, et al. (2008) Dynamic regulation by polycomb group protein complexes controls pattern formation and the cell cycle in Drosophila. Developmental Cell 15: 877–889. doi:10.1016/j.devcel.2008.10.005.

170. Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, et al. (2010) Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell 6: 479–491. doi:10.1016/j.stem.2010.03.018.

171. Li J-Y, Pu M-T, Hirasawa R, Li B-Z, Huang Y-N, et al. (2007) Synergistic function of DNA methyltransferases Dnmt3a and Dnmt3b in the methylation of Oct4 and Nanog. Molecular and Cellular Biology 27: 8748–8759. doi:10.1128/MCB.01380-07.

172. Santoriello C, Zon LI (2012) Hooked! Modeling human disease in zebrafish. J Clin Invest 122: 2337–2343. doi:10.1172/JCI60434.

173. Woods IG, Kelly PD, Chu F, Ngo-Hazelett P, Yan YL, et al. (2000) A comparative map of the zebrafish genome. Genome Research 10: 1903–1914.

174. Terriente J, Pujades C (2013) Use of zebrafish embryos for small molecule screening related to cancer. Dev Dyn 242: 97–107. doi:10.1002/dvdy.23912.

175. Chan J, Mably JD (2011) Dissection of cardiovascular development and disease pathways in zebrafish. Prog Mol Biol Transl Sci 100: 111–153. doi:10.1016/B978-0-12-384878-9.00004-2.

176. Bassett DI, Currie PD (2003) The zebrafish as a model for muscular dystrophy and congenital myopathy. Human Molecular Genetics 12 Spec No 2: R265–70. doi:10.1093/hmg/ddg279.

177. Newman M, Verdile G, Martins RN, Lardelli M (2011) Zebrafish as a tool in Alzheimer's disease research. Biochim Biophys Acta 1812: 346–352. doi:10.1016/j.bbadis.2010.09.012.

178. Liu S, Leach SD (2011) Zebrafish models for cancer. Annu Rev Pathol 6: 71–93. doi:10.1146/annurev-pathol-011110-130330.

179. Driever W, Stemple D, Schier A, Solnica-Krezel L (1994) Zebrafish: genetic tools for studying vertebrate development. Trends Genet 10: 152–159.

180. Summerton J, Weller D (1997) Morpholino antisense oligomers: design,

preparation, and properties. Antisense Nucleic Acid Drug Dev 7: 187–195.

181. Nasevicius A, Ekker SC (2000) Effective targeted gene “knockdown” in zebrafish. Nat Genet 26: 216–220. doi:10.1038/79951.

182. Wienholds E, van Eeden F, Kosters M, Mudde J, Plasterk RHA, et al. (2003) Efficient target-selected mutagenesis in zebrafish. Genome Research 13: 2700–2707. doi:10.1101/gr.1725103.

183. Zhu C, Smith T, McNulty J, Rayla AL, Lakshmanan A, et al. (2011) Evaluation and application of modularly assembled zinc-finger nucleases in zebrafish. Development 138: 4555–4564. doi:10.1242/dev.066779.

184. Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, et al. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Research 39: e82. doi:10.1093/nar/gkr218.

185. Suster ML, Kikuta H, Urasaki A, Asakawa K, Kawakami K (2009) Transgenesis in zebrafish with the tol2 transposon system. Methods Mol Biol 561: 41–63. doi:10.1007/978-1-60327-019-9_3.

186. Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF (1995) Stages of embryonic development of the zebrafish. Dev Dyn 203: 253–310. doi:10.1002/aja.1002030302.

187. Webb SE, Miller AL (2007) Ca2+ signalling and early embryonic patterning during zebrafish development. Clin Exp Pharmacol Physiol 34: 897–904. doi:10.1111/j.1440-1681.2007.04709.x.

188. Schier AF, Talbot WS (2005) Molecular genetics of axis formation in zebrafish. Annu Rev Genet 39: 561–613. doi:10.1146/annurev.genet.37.110801.143752.

189. Kudoh T, Concha ML, Houart C, Dawid IB, Wilson SW (2004) Combinatorial Fgf and Bmp signalling patterns the gastrula ectoderm into prospective neural and epidermal domains. Development 131: 3581–3592. doi:10.1242/dev.01227.

190. Kimelman D (2006) Mesoderm induction: from caps to chips. Nat Rev Genet 7: 360–372. doi:10.1038/nrg1837.

191. Wardle FC, Papaioannou VE (2008) Teasing out T-box targets in early mesoderm. Current Opinion in Genetics & Development 18: 418–425. doi:10.1016/j.gde.2008.07.017.

192. Martin BL, Kimelman D (2010) Brachyury establishes the embryonic mesodermal progenitor niche. Genes & Development 24: 2778–2783. doi:10.1101/gad.1962910.

193. Warga RM, Nüsslein-Volhard C (1999) Origin and development of the zebrafish endoderm. Development 126: 827–838.

194. Poulain M, Fürthauer M, Thisse B, Thisse C, Lepage T (2006) Zebrafish endoderm formation is regulated by combinatorial Nodal, FGF and BMP signalling. Development 133: 2189–2200. doi:10.1242/dev.02387.

195. Feng X-H, Derynck R (2005) Specificity and versatility in tgf-beta signaling through Smads. Annu Rev Cell Dev Biol 21: 659–693. doi:10.1146/annurev.cellbio.21.022404.142018.

196. Feldman B, Gates MA, Egan ES, Dougan ST, Rennebeck G, et al. (1998) Zebrafish organizer development and germ-layer formation require nodal-related signals. Nature 395: 181–185. doi:10.1038/26013.

197. Kunwar PS, Zimmerman S, Bennett JT, Chen Y, Whitman M, et al. (2003) Mixer/Bon and FoxH1/Sur have overlapping and divergent roles in Nodal signaling and mesendoderm induction. Development 130: 5589–5599. doi:10.1242/dev.00803.

198. Liu D, Kang JS, Derynck R (2004) TGF-beta-activated Smad3 represses MEF2-dependent transcription in myogenic differentiation. EMBO J 23: 1557–1566. doi:10.1038/sj.emboj.7600179.

199. Bubnoff von A, Cho KW (2001) Intracellular BMP signaling regulation in vertebrates: pathway or network? Developmental Biology 239: 1–14. doi:10.1006/dbio.2001.0388.

200. Fisher S, Halpern ME (1999) Patterning the zebrafish axial skeleton requires early chordin function. Nat Genet 23: 442–446. doi:10.1038/70557.

201. Bauer H, Meier A, Hild M, Stachel S, Economides A, et al. (1998) Follistatin and noggin are excluded from the zebrafish organizer. Developmental Biology 204: 488–507. doi:10.1006/dbio.1998.9003.

202. Dal-Pra S, Fürthauer M, Van-Celst J, Thisse B, Thisse C (2006) Noggin1 and Follistatin-like2 function redundantly to Chordin to antagonize BMP activity. Developmental Biology 298: 514–526. doi:10.1016/j.ydbio.2006.07.002.

203. Nikaido M, Tada M, Saji T, Ueno N (1997) Conservation of BMP signaling in zebrafish mesoderm patterning. Mech Dev 61: 75–88.

204. Agathon A, Thisse C, Thisse B (2003) The molecular nature of the zebrafish tail organizer. Nature 424: 448–452. doi:10.1038/nature01822.

205. Mullins MC, Hammerschmidt M, Kane DA, Odenthal J, Brand M, et al. (1996) Genes establishing dorsoventral pattern formation in the zebrafish embryo: the ventral specifying genes. Development 123: 81–93.

206. Ober EA, Schulte-Merker S (1999) Signals from the yolk cell induce mesoderm, neuroectoderm, the trunk organizer, and the notochord in zebrafish. Developmental Biology 215: 167–181. doi:10.1006/dbio.1999.9455.

207. Pyati UJ, Webb AE, Kimelman D (2005) Transgenic zebrafish reveal stage-specific roles for Bmp signaling in ventral and posterior mesoderm development. Development 132: 2333–2343. doi:10.1242/dev.01806.

208. Szeto DP, Kimelman D (2006) The regulation of mesodermal progenitor cell commitment to somitogenesis subdivides the zebrafish body musculature into distinct domains. Genes & Development 20: 1923–1932. doi:10.1101/gad.1435306.

209. Kelly C, Chin AJ, Leatherman JL, Kozlowski DJ, Weinberg ES (2000) Maternally controlled (beta)-catenin-mediated signaling is required for organizer formation in the zebrafish. Development 127: 3899–3911.

210. Nasevicius A, Hyatt T, Kim H, Guttman J, Walsh E, et al. (1998) Evidence for a frizzled-mediated wnt pathway required for zebrafish dorsal mesoderm formation. Development 125: 4283–4292.

211. Heisenberg CP, Houart C, Take-Uchi M, Rauch GJ, Young N, et al. (2001) A mutation in the Gsk3-binding domain of zebrafish Masterblind/Axin1 leads to a fate transformation of telencephalon and eyes to diencephalon. Genes & Development 15: 1427–1434. doi:10.1101/gad.194301.

212. Cao Y, Zhao J, Sun Z, Zhao Z, Postlethwait J, et al. (2004) fgf17b, a novel member of Fgf family, helps patterning zebrafish embryos. Developmental Biology 271: 130–143. doi:10.1016/j.ydbio.2004.03.032.

213. Fürthauer M, Van-Celst J, Thisse C, Thisse B (2004) Fgf signalling controls the dorsoventral patterning of the zebrafish embryo. Development 131: 2853–2864. doi:10.1242/dev.01156.

214. Lipchina I, Elkabetz Y, Hafner M, Sheridan R, Mihailovic A, et al. (2011) Genome-wide identification of microRNA targets in human ES cells reveals a role for miR-302 in modulating BMP response. Genes & Development 25: 2173–2186. doi:10.1101/gad.17221311.

215. Rosa A, Spagnoli FM, Brivanlou AH (2009) The miR-430/427/302 family controls mesendodermal fate specification via species-specific target selection. Developmental Cell 16: 517–527. doi:10.1016/j.devcel.2009.02.007.

216. Ren J, Jin P, Wang E, Marincola FM, Stroncek DF (2009) MicroRNA and gene expression patterns in the differentiation of human embryonic stem cells. J Transl Med 7: 20. doi:10.1186/1479-5876-7-20.

217. Kuo C-H, Deng JH, Deng Q, Ying S-Y (2012) A novel role of miR-302/367 in reprogramming. Biochem Biophys Res Commun 417: 11–16. doi:10.1016/j.bbrc.2011.11.058.

218. Lee MR, Prasain N, Chae H-D, Kim Y-J, Mantel C, et al. (2012) Epigenetic Regulation of Nanog by MiR-302 Cluster-MBD2 Completes Induced Pluripotent Stem Cell Reprogramming. STEM CELLS 31: 666–681. doi:10.1002/stem.1302.

219. Xia X, Zhang Y, Zieth CR, Zhang S-C (2007) Transgenes delivered by lentiviral vector are suppressed in human embryonic stem cells in a promoter-dependent manner. Stem Cells Dev 16: 167–176. doi:10.1089/scd.

220. Pratt T, Sharp L, Nichols J, Price DJ, Mason JO (2000) Embryonic stem cells and transgenic mice ubiquitously expressing a tau-tagged green fluorescent protein. Developmental Biology 228: 19–28. doi:10.1006/dbio.2000.9935.

221. Chambers SM, Fasano CA, Papapetrou EP, Tomishima M, Sadelain M, et al. (2009) Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat Biotechnol 27: 275–280. doi:10.1038/nbt.1529.

222. Morcos PA (2007) Achieving targeted and quantifiable alteration of mRNA splicing with Morpholino oligos. Biochem Biophys Res Commun 358: 521–527. doi:10.1016/j.bbrc.2007.04.172.

223. Halpern ME, Ho RK, Walker C, Kimmel CB (1993) Induction of muscle pioneers and floor plate is distinguished by the zebrafish no tail mutation. Cell 75: 99–111.

224. Schulte-Merker S, van Eeden FJ, Halpern ME, Kimmel CB, Nüsslein-Volhard C (1994) no tail (ntl) is the zebrafish homologue of the mouse T (Brachyury)

gene. Development 120: 1009–1015.

225. Odenthal J, Haffter P, Vogelsang E, Brand M, van Eeden FJ, et al. (1996) Mutations affecting the formation of the notochord in the zebrafish, Danio rerio. Development 123: 103–115.

226. Weinberg ES, Allende ML, Kelly CS, Abdelhamid A, Murakami T, et al. (1996) Developmental regulation of zebrafish MyoD in wild-type, no tail and spadetail embryos. Development 122: 271–280.

227. Kodjabachian L, Dawid IB, Toyama R (1999) Gastrulation in zebrafish: what mutants teach us. Developmental Biology 213: 231–245. doi:10.1006/dbio.1999.9392.

228. Frontelo P, Leader JE, Yoo N, Potocki AC, Crawford M, et al. (2004) Suv39h histone methyltransferases interact with Smads and cooperate in BMP-induced repression. Oncogene 23: 5242–5251. doi:10.1038/sj.onc.1207660.

229. Phan D, Rasmussen TL, Nakagawa O, McAnally J, Gottlieb PD, et al. (2005) BOP, a regulator of right ventricular heart development, is a direct transcriptional target of MEF2C in the developing heart. Development 132: 2669–2678. doi:10.1242/dev.01849.

230. Just S, Meder B, Berger IM, Etard C, Trano N, et al. (2011) The myosin-interacting protein SMYD1 is essential for sarcomere organization. J Cell Sci 124: 3127–3136. doi:10.1242/jcs.084772.

231. Bibikova M, Laurent LC, Ren B, Loring JF, Fan J-B (2008) Unraveling Epigenetic Regulation in Embryonic Stem Cells. Cell Stem Cell 2: 123–134. doi:10.1016/j.stem.2008.01.005.

232. Akkers RC, van Heeringen SJ, Jacobi UG, Janssen-Megens EM, Françoijs K-J, et al. (2009) A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in Xenopus embryos. Developmental Cell 17: 425–434. doi:10.1016/j.devcel.2009.08.005.

233. Lindeman LC, Andersen IS, Reiner AH, Li N, Aanes H, et al. (2011) Prepatterning of developmental gene expression by modified histones before zygotic genome activation. Developmental Cell 21: 993–1004. doi:10.1016/j.devcel.2011.10.008.

234. Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G (2007) Genome regulation by polycomb and trithorax proteins. Cell 128: 735–745. doi:10.1016/j.cell.2007.02.009.

235. Li B, Carey M, Workman JL (2007) The role of chromatin during transcription. Cell 128: 707–719. doi:10.1016/j.cell.2007.01.015.

236. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, et al. (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129: 1311–1323. doi:10.1016/j.cell.2007.05.022.

237. Song JS, Johnson WE, Zhu X, Zhang X, Li W, et al. (2007) Model-based analysis of two-color arrays (MA2C). Genome Biol 8: R178. doi:10.1186/gb-2007-8-8-r178.

238. Vermeulen M, Mulder KW, Denissov S, Pijnappel WWMP, van Schaik FMA, et al. (2007) Selective anchoring of TFIID to nucleosomes by trimethylation of

histone H3 lysine 4. Cell 131: 58–69. doi:10.1016/j.cell.2007.08.016.

239. Lauberth SM, Nakayama T, Wu X, Ferris AL, Tang Z, et al. (2013) H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. Cell 152: 1021–1036. doi:10.1016/j.cell.2013.01.052.

240. Ang Y-S, Tsai S-Y, Lee D-F, Monk J, Su J, et al. (2011) Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. Cell 145: 183–197. doi:10.1016/j.cell.2011.03.003.

241. Foreman KW, Brown M, Park F, Emtage S, Harriss J, et al. (2011) Structural and Functional Profiling of the Human Histone Methyltransferase SMYD3. PLoS ONE 6: e22290. doi:10.1371/journal.pone.0022290.t001.

242. Terme J-M, Sesé B, Millán-Ariño L, Mayor R, Belmonte JCI, et al. (2011) Histone H1 variants are differentially expressed and incorporated into chromatin during differentiation and reprogramming to pluripotency. Journal of Biological Chemistry 286: 35347–35357. doi:10.1074/jbc.M111.281923.

243. Sims RJ, Weihe EK, Zhu L, O'Malley S, Harriss JV, et al. (2002) m-Bop, a repressor protein essential for cardiogenesis, interacts with skNAC, a heart- and muscle-specific transcription factor. J Biol Chem 277: 26524–26529. doi:10.1074/jbc.M204121200.

244. Ruden DM, Xiao L, Garfinkel MD, Lu X (2005) Hsp90 and environmental impacts on epigenetic states: a model for the trans-generational effects of diethylstibesterol on uterine development and cancer. Human Molecular Genetics 14 Spec No 1: R149–55. doi:10.1093/hmg/ddi103.

245. Takenaga M, Fukumoto M, Hori Y (2007) Regulated Nodal signaling promotes differentiation of the definitive endoderm and mesoderm from ES cells. J Cell Sci 120: 2078–2090. doi:10.1242/jcs.004127.

246. D'Amour KA, Agulnick AD, Eliazer S, Kelly OG, Kroon E, et al. (2005) Efficient differentiation of human embryonic stem cells to definitive endoderm. Nat Biotechnol 23: 1534–1541. doi:10.1038/nbt1163.

247. Xu X, Browning VL, Odorico JS (2011) Activin, BMP and FGF pathways cooperate to promote endoderm and pancreatic lineage cell differentiation from human embryonic stem cells. Mech Dev 128: 412–427. doi:10.1016/j.mod.2011.08.001.

248. Jain AK, Allton K, Iacovino M, Mahen E, Milczarek RJ, et al. (2012) p53 regulates cell cycle and microRNAs to promote differentiation of human embryonic stem cells. PLoS Biol 10: e1001268. doi:10.1371/journal.pbio.1001268.

249. Taylor JS, Van de Peer Y, Braasch I, Meyer A (2001) Comparative genomics provides evidence for an ancient genome duplication event in fish. Philos Trans R Soc Lond, B, Biol Sci 356: 1661–1679. doi:10.1098/rstb.2001.0975.

250. Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y (2003) Genome duplication, a trait shared by 22000 species of ray-finned fish. Genome Research 13: 382–390. doi:10.1101/gr.640303.

251. Dosch R, Wagner DS, Mintzer KA, Runke G, Wiemelt AP, et al. (2004) Maternal control of vertebrate development before the midblastula transition: mutants from the zebrafish I. Developmental Cell 6: 771–780.

doi:10.1016/j.devcel.2004.05.002.

252. Wagner DS, Dosch R, Mintzer KA, Wiemelt AP, Mullins MC (2004) Maternal control of development at the midblastula transition and beyond: mutants from the zebrafish II. Developmental Cell 6: 781–790. doi:10.1016/j.devcel.2004.04.001.

253. Yang Y, Thorpe C (2011) BMP and non-canonical Wnt signaling are required for inhibition of secondary tail formation in zebrafish. Development 138: 2601–2611. doi:10.1242/dev.058404.

254. Chocron S, Verhoeven MC, Rentzsch F, Hammerschmidt M, Bakkers J (2007) Zebrafish Bmp4 regulates left-right asymmetry at two distinct developmental time points. Developmental Biology 305: 577–588. doi:10.1016/j.ydbio.2007.03.001.

255. Lenhart KF, Lin S-Y, Titus TA, Postlethwait JH, Burdine RD (2011) Two additional midline barriers function with midline lefty1 expression to maintain asymmetric Nodal signaling during left-right axis specification in zebrafish. Development 138: 4405–4410. doi:10.1242/dev.071092.

256. Dick A, Hild M, Bauer H, Imai Y, Maifeld H, et al. (2000) Essential role of Bmp7 (snailhouse) and its prodomain in dorsoventral patterning of the zebrafish embryo. Development 127: 343–354.

257. Kishimoto Y, Lee KH, Zon L, Hammerschmidt M, Schulte-Merker S (1997) The molecular nature of zebrafish swirl: BMP2 function is essential during early dorsoventral patterning. Development 124: 4457–4466.

258. Lekven AC, Thorpe CJ, Waxman JS, Moon RT (2001) Zebrafish wnt8 encodes two wnt8 proteins on a bicistronic transcript and is required for mesoderm and neurectoderm patterning. Developmental Cell 1: 103–114.

259. Yao S, Qian M, Deng S, Xie L, Yang H, et al. (2010) Kzp controls canonical Wnt8 signaling to modulate dorsoventral patterning during zebrafish gastrulation. Journal of Biological Chemistry 285: 42086–42096. doi:10.1074/jbc.M110.161554.

260. Dee CT, Gibson A, Rengifo A, Sun S-K, Patient RK, et al. (2007) A change in response to Bmp signalling precedes ectodermal fate choice. Int J Dev Biol 51: 79–84. doi:10.1387/ijdb.062204cd.

261. Orkin SH, Zon LI (1997) Genetics of erythropoiesis: induced mutations in mice and zebrafish. Annu Rev Genet 31: 33–60. doi:10.1146/annurev.genet.31.1.33.

262. Chang LL, Kessler DS (2010) Foxd3 is an essential Nodal-dependent regulator of zebrafish dorsal mesoderm development. Developmental Biology 342: 39–50. doi:10.1016/j.ydbio.2010.03.017.

263. Latinkic BV, Smith JC (1999) Goosecoid and mix.1 repress Brachyury expression and are required for head formation in Xenopus. Development 126: 1769–1779.

264. Reiter JF, Kikuchi Y, Stainier DY (2001) Multiple roles for Gata5 in zebrafish endoderm formation. Development 128: 125–135.

265. Stainier DYR (2002) A glimpse into the molecular entrails of endoderm formation. Genes & Development 16: 893–907. doi:10.1101/gad.974902.

266. Szeto DP (2004) Combinatorial gene regulation by Bmp and Wnt in zebrafish posterior mesoderm formation. Development 131: 3751–3760. doi:10.1242/dev.01236.

267. Harvey SA, Tümpel S, Dubrulle J, Schier AF, Smith JC (2010) no tail integrates two modes of mesoderm induction. Development 137: 1127–1135. doi:10.1242/dev.046318.

268. Shinya M, Eschbach C, Clark M, Lehrach H, Furutani-Seiki M (2000) Zebrafish Dkk1, induced by the pre-MBT Wnt signaling, is secreted from the prechordal plate and patterns the anterior neural plate. Mech Dev 98: 3–17.

269. Pera MF, Andrade J, Houssami S, Reubinoff B, Trounson A, et al. (2004) Regulation of human embryonic stem cell differentiation by BMP-2 and its antagonist noggin. J Cell Sci 117: 1269–1280. doi:10.1242/jcs.00970.

270. Li Z, Chen Y-G (2013) Functions of BMP signaling in embryonic stem cell fate determination. Exp Cell Res 319: 113–119. doi:10.1016/j.yexcr.2012.09.016.

271. Ying QL, Nichols J, Chambers I, Smith A (2003) BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell 115: 281–292.

272. Ying QL, Stavridis M, Griffiths D, Li M, Smith A (2003) Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. doi:10.1038/nbt780.

273. Zeng X, Cai J, Chen J, Luo Y, You Z-B, et al. (2004) Dopaminergic differentiation of human embryonic stem cells. STEM CELLS 22: 925–940. doi:10.1634/stemcells.22-6-925.

274. Xiao L, Yuan X, Sharkis SJ (2006) Activin A maintains self-renewal and regulates fibroblast growth factor, Wnt, and bone morphogenic protein pathways in human embryonic stem cells. STEM CELLS 24: 1476–1486. doi:10.1634/stemcells.2005-0299.

275. Wakabayashi Y, Tamiya T, Takada I, Fukaya T, Sugiyama Y, et al. (2011) Histone 3 lysine 9 (H3K9) methyltransferase recruitment to the interleukin-2 (IL-2) promoter is a mechanism of suppression of IL-2 transcription by the transforming growth factor-β-Smad pathway. Journal of Biological Chemistry 286: 35456–35465. doi:10.1074/jbc.M111.236794.

276. Chen J, Liu H, Liu J, Qi J, Wei B, et al. (2013) H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. Nat Genet 45: 34–42. doi:10.1038/ng.2491.

277. Pan G, Tian S, Nie J, Yang C, Ruotti V, et al. (2007) Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell Stem Cell 1: 299–312. doi:10.1016/j.stem.2007.08.003.

278. Zhao X-D, Han X, Chew J-L, Liu J, Chiu K-P, et al. (2007) Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. Cell Stem Cell 1: 286–298. doi:10.1016/j.stem.2007.08.004.

279. Fei T, Xia K, Li Z, Zhou B, Zhu S, et al. (2010) Genome-wide mapping of SMAD target genes reveals the role of BMP signaling in embryonic stem cell fate determination. Genome Research 20: 36–44. doi:10.1101/gr.092114.109.

280. Rodríguez-Pizà I, Richaud-Patin Y, Vassena R, González F, Barrero MJ, et al. (2010) Reprogramming of human fibroblasts to induced pluripotent stem cells under xeno-free conditions. STEM CELLS 28: 36–44. doi:10.1002/stem.248.

281. Aasen T, Raya A, Barrero MJ, Garreta E, Consiglio A, et al. (2008) Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes. Nat Biotechnol 26: 1276–1284. doi:10.1038/nbt.1503.

282. Raya A, Rodríguez-Pizà I, Arán B, Consiglio A, Barri PN, et al. (2008) Generation of cardiomyocytes from new human embryonic stem cell lines derived from poor-quality blastocysts. Cold Spring Harb Symp Quant Biol 73: 127–135. doi:10.1101/sqb.2008.73.038.

ANNEX I. LIST OF ABBREVIATIONS

5mC	5-methylcytosine
5hmC	5-hidroximethylcytosine
ATP	Adenosine triphosphate
bp	Base pair
bFGF	Basic fibroblast growth factor
BSA	Bovine serum albumin
BMP	Bone morphogenetic protein
ChIP	Chromatin immunoprecipitation
CTD	Carboxy-terminal domain
DMSO	Dimethylsulfoxide
DNMT	DNA methyltransferasa
DTT	Dithiothreitol
EB	Embryoid bodies
EDTA	Ethylenediaminetetraacetic acid
ES (cells)	Embryonic stem cells
GFP	Green fluorescent protein
GST	Glutathione S-transferase
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HDM	Histone demethylase
HFF	Human foreskin fibroblasts
HK	Human keratinocytes
HKDM	Histone lysine demethylase
HKMT	Histone lysine methyltransferase
HMT	Histone methyltransferase
hpf	Hours post-fertilization
ICM	Inner cell mas
iPS (cells)	Induced pluripotent stem cells
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kD	kiloDalton
KiPS (cells)	Keratinocytes induced pluripotent stem cells

LB	Luria Bertani culture media
MBD	Methyl-CpG binding domain
MBT	Midblastula transition
MEF	Mouse embryonic fibroblasts
MLL	Mixed-linage leukemia
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PRC	Polycomb-repressive complex
PRMT	Protein arginine methyltransferase
SDS	Sodium dodecyl sulfate
TAE	Tris-acetate-EDTA
TBS	Tris-buffered saline
TGF-β	Transforming growth factor beta
TE	Tris-EDTA
TEMED	Tetramethylethylenediamine
TPR	Tetratricopeptide repeat
wt	Wild Type

ANNEX II. PUBLICATIONS

Paper 1: Adamo A, Sesé B, Boue S, Castaño J, Paramonov I, et al. (2011) **LSD1 regulates the balance between self-renewal and differentiation in human embryonic stem cells.** Nat Cell Biol 13: 652–660. doi:10.1038/ncb2246. http://www.nature.com/ncb/journal/v13/n6/full/ncb2246.html

Paper 2: Terme J-M, Sesé B, Millán-Ariño L, Mayor R, Belmonte JCI, et al. (2011) **Histone H1 variants are differentially expressed and incorporated into chromatin during differentiation and reprogramming to pluripotency.** Journal of Biological Chemistry 286: 35347–35357. doi:10.1074/jbc.M111.281923. http://www.jbc.org/content/286/41/35347.long

Paper 3: Barrero MJ, Sesé B, Kuebler B, Bilic J, Boue S, et al. (2013) **Macrohistone Variants Preserve Cell Identity by Preventing the Gain of H3K4me2 during Reprogramming to Pluripotency.** Cell Rep. doi:10.1016/j.celrep.2013.02.029. http://www.cell.com/cell-reports/retrieve/pii/S2211124713001058?_returnURL=http://linkinghub.elsevier.com/retrieve/pii/S2211124713001058?showall=true

Paper 4: Barrero MJ, Sesé B, Martí M, Izpisúa Belmonte JC (2013) **Macro Histone Variants are Critical for the Differentiation of Human Pluripotent Cells.** Journal of Biological Chemistry. doi:10.1074/jbc.M113.466144. http://www.jbc.org/content/early/2013/04/17/jbc.M113.466144.long

Paper 5: Sesé B, Barrero MJ, Fabregat MC, Sander V, Izpisúa Belmonte JC (2013) **SMYD2 is Induced during Differentiation and Participates in Early Development.** International Journal of Developmental Biology. (*In press*)

SMYD2 is induced during differentiation and participates in early development

Borja Sesé[1], Maria J. Barrero[1], Maria Carme Fabregat[1], Veronika Sander[1] and Juan Carlos Izpisua Belmonte[1,2]

[1]Center for Regenerative Medicine in Barcelona, Dr.Aiguader 88, 7th floor. Barcelona, E-08003
[2]Salk Institute for Biological Studies, 10010 N. Torrey Pines Rd., La Jolla, CA 92037

Running tittle: Role of SMYD2 in differentiation
Key words: SMYD2, differentiation, development, stem cells, methyltransferase

Abstract

Histone modifying enzymes play critical roles in differentiation and development. In this study we describe that SMYD2 (SET and MYND domain containing protein 2), a histone lysine methyltransferase, is induced during human ES cells differentiation and it is preferentially expressed in somatic cells versus pluripotent cells. Knockdown of SMYD2 in human ES cells promotes the induction of endodermal markers during differentiation, while overexpression has opposite effects. *In vivo* experiments in zebrafish reveal that knockdown of *smyd2a* (a homologue gene of human SMYD2) causes developmental delays and aberrant tail formation, which is coincident with low expression of ntl and over induction Nodal-related genes during gastrulation. Taken together, these findings suggest that SMYD2 plays a critical role at early stages during development and in human ES cells differentiation.

Introduction

Embryonic stem (ES) cells are characterized by their ability to self-renew indefinitely *in vitro* maintaining their undifferentiated state, and the capacity to give rise to any cell type in the body (Thomson *et al.*, 1998). The molecular mechanisms underlying ES cells identity and their potential for differentiation are still poorly understood (Boyer *et al.*, 2005, Niwa, 2007). During the last years, post-translational covalent modifications of histone proteins have emerged as a crucial epigenetic event to regulate the pluripotent state of ES cells and to establish cell fate decisions (Bibikova *et al.*, 2008, Goldberg *et al.*, 2007, Meissner, 2010). Histone modifications help to define chromatin structure and can be associated with active marks, offering accessible DNA for transcription, and repressive marks, where DNA is more packaged and inaccessible (Cheung and Lau, 2005, Kouzarides, 2007). One particular property of ES cells in many developmental genes present both active H3K4 and repressive H3K27 methylation on their regulatory regions, the so-called "bivalent domains" (Azuara *et al.*, 2006, Bernstein *et al.*, 2006). This bivalency keeps these genes silenced in ES cells, but "poised" to become activated by losing the repressive mark, or kept silenced by removal of active marks during differentiation (Spivakov and Fisher, 2007).

In order to select candidates of histone modifying enzymes involved in pluripotency and differentiation of human ES cells, we identified a group of histone lysine demethylases and methyltransferases differentially expressed between undifferentiated and differentiated human ES cells (data not shown). Among them, SMYD2 emerged as a potential candidate due to its barely expression in undifferentiated human ES cells and later up-regulation after 30 days of differentiation towards the three germ layers. The SMYD (SET and MYND domain) protein family presents methyltransferase activity provided by its SET domain, which is split into two segments by its MYND domain, responsible for protein-protein interactions (Brown *et al.*, 2006). In human, there are five members in the SMYD protein family (SMYD1-5) and have been shown to participate regulating gene transcription and cell proliferation. SMYD1 is a heart and muscle specific histone methyltransferase involved in cardiomyocyte and myogenic differentiation (Gottlieb *et al.*, 2002, Li *et al.*, 2009). The lack of Smyd1 in mice development results in embryonic death due to cardiac defects (Gottlieb *et al.*, 2002). Knockdown of *smyd1a/b* in zebrafish causes skeletal and cardiac muscle defects and presents a disrupted expression of myofibril organization (Tan *et al.*, 2006), SMYD3 has been mainly related with cancer cell proliferation (Hamamoto *et al.*, 2004). Several findings indicate that

endogenous expression of SMYD3 is present at very high levels in hepatocellular, colon and breast carcinoma, and silencing through siRNAs have an inhibitory effect in cell growth (Hamamoto *et al.*, 2004, Hamamoto *et al.*, 2006). Similarly to *smyd1a/b*, *smyd3* plays an important role in cardiac and skeletal muscle development in zebrafish (Fujii *et al.*, 2011). In the other hand, SMYD4 is significantly reduced in tumor cells and its re-expression dramatically decreases cancer cell growth (Hu *et al.*, 2009). Also, *Drosophila* SMYD4 homologue has been involved in muscle development (Thompson and Travers, 2008). Little is known so far about SMYD5. Unlike the rest of family members, SMYD5 does not present a C-terminal tetratricopeptide repeat (TPR) domain (Abu-Farha *et al.*, 2011).

SMYD2 was first described as a histone lysine methyltransferase mainly expressed in heart and brain tissue, with specific catalytic activity for H3K36 dimethylation, a mark associated with actively transcribed genes (Brown *et al.*, 2006). Moreover, SMYD2 was associated to interact with HDAC1 and the Sin3A repression complex (Brown *et al.*, 2006). In yeast, there is evidence of a link between H3K36 methylation and the recruitment of a repressive Rpd3 (the prototypical yeast HDAC) complex (Carrozza *et al.*, 2005, Joshi and Struhl, 2005, Keogh *et al.*, 2005) . However, SMYD2 was also described to specifically monomethylate H3K4 in the presence of HSP90α *in vitro*, with no activity for H3K36 (Abu-Farha *et al.*, 2008). Despite observing a weak activity of H3K36 methylation in the absence of HSP90α, *in vivo* experiments suggested that H3K4 is the predominant site of methylation for SMYD2 (Abu-Farha *et al.*, 2008). Regarding histone methylation, recent studies determined that SMYD2 also methylates histones H2B and H4 more efficiently than H3 *in vitro* (Wu *et al.*, 2011). Further, several non-histone proteins have been identified as substrates for SMYD2 methylation such as p53 and retinoblastoma (Rb) (Cho *et al.*, 2012, Huang *et al.*, 2006, Jiang *et al.*, 2011, Saddic *et al.*, 2010, Scoumanne and Chen, 2008). Monomethylation of p53 at K370 reduces its binding ability to promoter target genes like p21 and mdm2, resulting in a decreased expression of these genes (Huang *et al.*, 2006). Unexpectedly, adult hearts of Smyd2 conditional knockout mice showed no changes in p21 and mdm2 expression levels, and had no global effect in H3K36 or H3K4 methylation (Diehl *et al.*, 2010). Additionally, Smyd2 was found dispensable for proper heart development in mouse (Diehl *et al.*, 2010). Rb protein can be methylated by SMYD2 at K860 and facilitates its interaction with the methyl-binding protein L3MBTL1 (Saddic *et al.*, 2010). More recently, SMYD2 was also found to methylate Rb at K810, which increases phosphorylation of Rb protein, and promotes cell cycle progression (Cho *et al.*, 2012). In accordance with the effects observed on p53 and Rb tumour suppressors, a wide variety of human cancer showed high levels of SMYD2 expression (Cho *et al.*, 2012, Komatsu *et al.*, 2009). It was also reported that SMYD2 is involved in maintaining self-renewal activity of MLL-AF9-induced acute myeloid leukaemia (Zuber *et al.*, 2011). Another novel non-histone substrate for SMYD2, cytoplasmic HSP90, is methylated at K209 and K615 (Abu-Farha *et al.*, 2011). In mouse, Smyd2 methylates HSP90 to form a complex with the sarcomeric protein titin to protect myocyte organization (Donlin *et al.*, 2012, Voelkel *et al.*, 2012). Here we reported the characterization of SMYD2 during human ES cell differentiation. Our data suggest that SMYD2 plays an important role during early differentiation events.

Results

SMYD2 expression is strongly induced during human ES cells differentiation

We first set up to analyze the potential differential expression of the SMYD family members in pluripotent and somatic cells by qPCR. mRNA levels of all five SMYD family members were measured in human embryonic stem (ES) cells (ES[2] and ES]4]), induced pluripotent (iPS) cells derived from human keratinocytes (KiPS) and human fibroblasts (HF) and keratinocytes (HEK) (Fig. 1a). Among all family members, SMYD2 clearly showed higher mRNA levels in somatic cells compared to pluripotent cells. We next tested the expression profile of the SMYD family members during the differentiation of human ES cells in the form of embryoid bodies (Fig. 1b). As expected, SMYD2 showed the most remarkable induction among the SMYD family members, which is coincident with the dowregulation of the expression of pluripotency-related genes (Fig. 1c). Taken together, our results show that SMYD2 is preferentially expressed in human somatic cells and induced during the differentiation of human pluripotent cells.

SMYD2 promoter presents active chromatin marks in human ES cells

Since SMYD2 is expressed at low levels in human ES cells but rapidly induced during differentiation we considered the potential presence of bivalent domains on its promoter in pluripotent cells. We performed ChIP assay to detect the presence of histone marks related with

transcriptional activation (H3K4me) and repression (H3K27me) in ES[4] (Fig. 2a) and ES[2] (Fig. 2b). As a control, we tested the presence of histone marks for the pluripotency gene *OCT4*, and two well described genes containing bivalent domains in human ES cells, *SOX17* and *FOXA2*. Despite being transcribed at low levels, *SMYD2* showed remarkable levels of H3K4me2/3 at its promoter. Levels of H3K27me3 showed variability between lines, being very low in ES[4] but significant in ES[2], while compared with the well known bivalent genes *SOX17* and *FOXA2*. The presence of H3K4me2/3 only at the *SMYD2* promoter in ES[4] might suggests the existence of post-transcriptional mechanisms responsible for blocking its expression in human ES cells.

The knockdown of SMYD2 promotes induction of endodermal genes during human ES cells differentiation

To test if SMYD2 is plays a role in the differentiation of human ES cells we performed loss-of-function experiments. We generated lentiviral-transduced stable ES[4] cell lines expressing a shRNA against *SMYD2* (shSMYD2) (Huang *et al.*, 2006) and a random non target shRNA (shControl). In self-renewing cells, the knockdown of SMYD2 did not show any morphological differences compared to the control, neither differences regarding the expression of selected pluripotency and differentiation genes (data not shown). However, during *in vitro* differentiation the SMYD2 knockdown line showed higher levels of endodermal genes (*HNF4*, *FOXA2*, *SOX17*) compared to the shControl line (Fig. 3a), but not significant differences in the induction of ectordermal (*PAX6*) or mesodermal (*BRACHYURY*) genes. The pluripotency-related genes *OCT4*, *NANOG* and *SOX2* were similarly downregulated in the shSMYD2 and shControl lines. Our results suggest that SMYD2 might act as a negative regulator of endodermal differentiation.

The SMYD2 overexpression impairs proper differentiation of human ES cells

To further confirm the involvement of SMYD2 in differentiation we generated stable ES[4] cell lines stably overexpressing SMYD2 (SMYD2) or GFP as a control (GFP). The overexpression of SMYD2 in self-renewing cells did not cause significant morphological differences neither differences in the expression of selected pluripotency and differentiation genes compared to the control (Fig. 3b). Contrary to the knockdown, cells overexpressing SMYD2 showed a reduced induction of endodermal and mesodermal genes compared to the control line during differentiation (Fig. 3b). However, the ectodermal gene *PAX6* was more induced in SMYD2 overexpressing cells than control cells. No differences were detected between the SMYD2 and GFP overexpressing cell lines regarding the silencing of pluripotency genes during differentiation. These results confirm that SMYD2 acts as a repressor of endodermal genes and has no effect in the downregulation of pluripotency genes during human ES differentiation.

The knockdown of *smyd2a* in zebrafish results in tail formation defects

To assess the effect of SMYD2 at early stages of development we performed knockdown experiments in zebrafish as *in vivo* model. The zebrafish presents two homologous genes of SMYD2: *smyd2a* and *smyd2b*. We first quantified the mRNA levels of both genes during zebrafish development from 0.2 to 48 hours post-fertilization (hpf) (Fig. 4a). Interestingly, *smyd2a* was maternally expressed, rapidly degraded after fertilization and induced again during gastrulation (5 to 10 hpf). *smyd2b* expression levels remained low until 10 hpf and was dramatically induced after gastrulation. The induction of *smyd2a* during gastrulation suggests a potential role in germ layer specification. To confirm our hypothesis we designed a splice-blocking morpholino (smyd2a-MO) at the exon 1-intron 1 junction, to block proper splicing of the zygotic transcripts without affecting the maternal mRNA (Fig. 4b, upper panel). Zebrafish embryos were injected at one-cell stage with smyd2a-MO and a standard-control morpholino (Control-MO). As expected, the mature mRNA of *smyd2a* was undetectable in embryos injected with smyd2a-MO at 24 hpf compared to embryos injected with Control-MO, while *smyd2b* expression was unaffected by the injection of smyd2a-MO (Fig. 4b, lower panel). Injected embryos were examined morphologically up to 6 days after injection (Fig. 4c). At 5 hpf, smyd2a-MO injected embryos could not be morphologically distinguished from the controls. At 24 hpf, we could detect some *smyd2a* morphant embryos with a tail defect (mild), and some others with a strong delay in development (severe). At 48 hpf, we could observe several degrees of tail defects, including a morphant phenotype with a complete absence of the tail (very severe). Later on, at 6 days after injection, all three *smyd2a* morphant phenotypes were distinguishable. To further confirm the specificity of smyd2a-MO effect, we performed a rescue experiment by co-injecting with *in vitro*-

transcribed human *SMYD2* mRNA (smyd2a-MO + *SMYD2* mRNA) and evaluated the percentage embryos with different phenotypes at 24h post-injection (Fig. 4d). Table 1 shows that co-injection of smyd2a-MO with human *SMYD2* mRNA rescues both mortality (from 57% to 36%) and morphology (from 25% mild and 60% severe phenotype to 11% mild and 18% severe phenotype). Co-injection of human *SMYD2* mRNA and Control-MO had no effects on mortality or phenotype compared to the injection of Control-MO alone. Therefore, the effects of *smyd2a* knockdown in zebra fish development appear specific.

smyd2a knockdown in zebrafish affects the expression of genes involved in the Nodal pathway

To gain insights into the potential pathways affected by the *smyd2a* knockdown we investigated the expression pattern of the T-box gene *no tail* (*ntl*), a zebrafish Brachyury ortholog required for the formation of mesoderm and essential for tail development (Halpern *et al.*, 1993, Schulte-Merker *et al.*, 1994) , by whole-mount *in situ* hybridization (ISH) on embryos at 6 hpf (Fig. 5a). As expected, all control embryos showed *ntl* expression at the margin cells as gastrulation proceeds. In contrast most smyd2a-MO injected embryos showed altered pattern of *ntl* expression characterized by either a reduction (35% of embryos) or complete absence of *ntl* expression (53% of embryos). Moreover, the ratio of the three different *ntl* expression patters in *smyd2a* morphants is similar to the percentages of different phenotypes at 24 hpf (Fig. 4d, Table 1). To further identify specific transcripts affected by *smyd2a* knockdown, we analyzed the expression profile of a set of genes involved in gastrulation from 4 hpf to 10 hpf by qPCR (Fig. 5b). Nodal-related genes (*chd*, *gsc*, *bon* and *cas*) were up-regulated in the smyd2a-MO compared to the Control-MO. In addition to their strong induction, *gsc* and *bon* presented also a different profile of expression as they were still present at high levels even at 8 hpf, while the control-MO injected embryos had already reduced their expression levels. *bmp2a* and *β-catenin-1* are unaffected by *smyd2a* knockdown, suggesting that genes of the Bmp2 and Wnt pathways remain unaffected. These results indicate that *smyd2a* knockdown causes an induction of Nodal-target genes *chd*, *gsc*, *bon* and *cas* during gastrulation events in zebrafish.

Discussion

In this work we show that SMYD2 is involved in human ES cells differentiation. SMYD2 is expressed at high levels in somatic cells compared with pluripotent cells, whereas the rest of family members are not so differentially expressed. During human ES cells differentiation, SMYD2 is also the most induced family member. SMYD1 and 3, which share the highest degree of sequence homology with SMYD2 (Abu-Farha *et al.*, 2011), are also induced but not so remarkably. SMYD3 present a progressive induction but only around 3-fold change compared with undifferentiated cells. Even at low levels, SMYD1 was induced during differentiation reaching its peak of expression at day 8. This indicates that SMYD1 might be involved also in early development, consistent with previous studies involving Smyd1 in mouse developing heart as a direct target of MEF2c (Phan *et al.*, 2005).

Given the low levels of expression of SMYD2 in pluripotent cells, we hypothesized the presence of bivalent domains on its promoter to keep it repressed. Unexpectedly, we found variable levels of H3K27me3 between cell lines. In ES[4] the *SMYD2* promoter was marked with H3K4 methylation marks only, suggesting that SMYD2 expression could be blocked by an alternative mechanism. Recently, Lipchina *et al.* (Lipchina *et al.*, 2011) identified SMYD2, together with other 145 genes, as a high-confidence target of the miR-302/357 cluster in human ES cells. miR-302/357 is expressed in undifferentiated conditions and is downregulated upon neuronal differentiation, releasing the expression of its target genes (Lipchina *et al.*, 2011). Thus, a possible miR-302/357 regulation of SMYD2 might explain the low levels of expression in pluripotent cells.

The knockdown of SMYD2 promoted the induction of endodermal genes during human ES cells differentiation, whereas overexpression had opposite effects. We further investigated the effect of SMYD2 depletion *in vivo* using the zebrafish as a model. Our results show that both *smyd2a* and *smyd2b* were induced during development, and surprisingly, *smyd2a* was found to be maternally expressed in zebrafish embryos. Knockdown of *smyd2a* causes a delay in zebrafish development and tail deformities in more than 80% of surviving embryos at 24 hpf, whereas rescue experiments reduced the morphant phenotype to 30%. The fact that we were able to rescue *smyd2a* morphant phenotype by co-injecting human *SMYD2* mRNA suggests a very well conserved function of SMYD2 among species. Intriguingly, it was recently reported that *smyd2a* knockdown generates severe skeletal and cardiac muscle defects in zebrafish (Donlin *et al.*, 2012, Voelkel *et al.*, 2012) . Regarding the cardiac phenotype, we did not appreciate any abnormalities, even in the very severe morphant

phenotype at 6 dpf, consistent with mice conditional-knockout experiments where Smyd2 was dispensable for heart development (Diehl *et al.*, 2010).

The *smyd2a* knockdown phenotype that we observe is characteristic of embryonic dorsalization and resembles the phenotype of *ntl* mutant embryos (Odenthal *et al.*, 1996). Accordingly, the observed dorsalization effects were consistent with reduced or absent levels of *ntl* transcript at 6 hpf. On the contrary, knockdown of SMYD2 in human ES cell differentiation did not affect the expression of the *ntl* ortholog, *BRACHYURY* (Fig. 3a). This discrepancy might be caused by the differences between the two model systems or to the lower efficiency of SMYD2 knockdown in ES cells. We also observed increased expression levels of Nodal-related genes, such as *Bon* and *cas,* described to promote the formation of endoderm in zebrafish (Reiter *et al.*, 2001, Stainier, 2002). Also, in *Xenopus, goosecoid* and *Mix.1*, the homologous genes of *gsc* and *bon* respectively, act together to promote endodermal differentiation and suppress expression of mesodermal *Xbra*, the *ntl* homologous gene (Latinkic and Smith, 1999). Eventually, high levels of *gsc* and *bon* at 6 and 8 hpf in smyd2a-MO embryos might have repressive effects over mesoderm induction. Overall, the zebrafish phenotype seems consistent with the induction of endodermal genes in the knockdown of SMYD2 during human ES cell differentiation.

Since, SMYD2 has been previously involved in transcriptional activation through monomethylation of H3K4 we speculate that it might participate in the induction of repressors of the endodermal fate. Alternatively, it might act as a transcriptional repressor itself through its association with HDACI and Sin3a repression complexes (Brown *et al.*, 2006). However, methylation of non histone targets might also account for these effects. For example, recent findings indicate that p53 knockdown causes a delay in differentiation of human ES cells, and ectopic expression of p53R175H, a mutated inactive form of p53, failed to induce differentiation (Jain *et al.*, 2012).

In summary, our study shows that SMYD2 is expressed at low levels in pluripotent cells but is strongly induced during differentiation. Knockdown of SMYD2 induced expression of endodermal genes, whereas overexpression leads to a blockade of differentiation by impairing induction of most of differentiation genes. Moreover, *in vivo* experiments in the zebrafish showed that *smyd2a* is involved in mesoderm formation and has a critical role from very early stages during development. In conclusion, our work suggests that SMYD2 plays an early decisive role in embryonic differentiation.

Acknowledgments

B.S. was a recipient of a FPU predoctoral fellowship from MICINN. M.J.B was partially supported by the Ramón y Cajal program and MICINN grants RYC-2007-01510 and SAF2009-08588. This work was supported by grants from the G. Harold and Leila Y. Mathers Charitable Foundation, The Leona M. and Harry B. Helmsley Charitable Trust, TERCEL-ISCIII-MINECO, CIBER and Fundacion Cellex to J.C.I.B.

Materials and Methods

Cell culture and differentiation

Human embryonic stem cell lines ES[4] and ES[2] (Raya *et al.*, 2008), were grown on matrigel-coated plates in irradiated mouse embryonic fibroblasts conditioned HES media (Knock out DMEM, 20% Knock out serum replacement, nonessential amino acids, 2mM L-glutamine and 50μM β -mercaptoethanol) and supplemented with 10ng/ml FGF.

For *in vitro* differentiation, cells were trypsinized into a single cell suspension and resuspended in MEF's conditioned HES media. Embryoid body (EB) formation was induced by seeding 100,000 cells in each well of 96-well v-bottom, low attachment plates and centrifuging the plates at 950g for 5 min to aggregate the cells. After 3 days the embryoid bodies were transferred to 0.1% gelatin-coated dishes and cultured in differentiation medium (Knock out DMEM, 20% fetal bovine serum, nonessential amino acids, 2mM L-glutamine and 50μM β -mercaptoethanol) up to 15 days.

RNA extraction and qPCR

RNA was extracted using TRIZOL (Invitrogen) and cDNA synthesis was performed using the Cloned AMV First-Strand cDNA Synthesis Kit (Invitrogen). Quantification of mRNA levels was carried out by real time PCR using SYBER Green. qPCR levels were normalized to GAPDH in human cells and to 18S in zebrafish.

ChIP assays

Chromatin immunoprecipitation (ChIP) assays were performed according to Adamo *et al.* (Adamo *et al.*, 2011). The antibodies used were anti- H3K4me2 (07030 from Millipore), H3K4me3 (07-473 from Millipore), and H3K27me3 (07-449 from Millipore). Purified chromatin was quantified using qPCR. The sequences of the oligonucleotydes are available on request.

Lentiviral production and infection
Previously described short hairpin against *SMYD2* (Huang *et al.*, 2006) and a non target short hairpin were cloned into the lentiviral pLVTHM vector (Addgene plasmid 12247). Virus production was performed as described (Wiznerowicz and Trono, 2003). After infection, GFP positive cells were selected by FACS sorting.

Vector construction and transfection
Flag tagged SMYD2 cDNA was cloned into vector pTP6 (Pratt *et al.*, 2000). ES[4] were transfected with linearized pTP6-SMYD2 and pTP6 empty vector and clones were selected with 2ug/ml puromycin and pooled. SMYD2 expression was confirmed by qPCR and western blot.

Zebrafish microinjection of morpholino and SMYD2 mRNA
Antisense morpholino-oligonucleotides (MO) were supplied by Gene Tools LCC (Philomath, OR). The sequence of *smyd2a* splice-blocking MO (smyd2a-MO) is 5' TTATAAGGAGCGCTGACCTGGTAAA 3' and was designed to block *smyd2a* proper splicing by binding to the splice site located between exon 1 and intron 1. The sequence of the standard control MO (Control-MO) was 5' CCTCTTACCTCAGTTACAATTTATA 3' and was used as an injection control. MOs were diluted in Danieau buffer to a final concentration of 0,35mM and were injected at 1-2 cell stage of fertilized wild-type zebrafish (AB strain) eggs using microinjector. Flag tagged SMYD2 cDNA was cloned into pCS2 vector for mRNA synthesis using the mMESSAGE mMACHINE kit (Ambion, AMI344, Life Technologies, USA) following the manufacturer's instructions. Rescue experiments were performed by co-injecting 30 pg of synthesized *SMYD2* mRNA with 0.35mM of smyd2a-MO or Control-MO.

Whole mount in situ hybridization
Antisense ntl RNA probe was synthesized using the DIG RNA Labeling Kit SP6/T7 (Roche). Embryos were collected and fixed at 6 hpf. Whole mount *in situ* hybridization was performed as previously described (Jopling *et al.*, 2010).

References

ABU-FARHA, M., LAMBERT, J.P., AL-MADHOUN, A.S., ELISMA, F., SKERJANC, I.S. and FIGEYS, D. (2008). The tale of two domains: proteomics and genomics analysis of SMYD2, a new histone methyltransferase. *Mol Cell Proteomics* 7: 560-572.

ABU-FARHA, M., LANOUETTE, S., ELISMA, F., TREMBLAY, V., BUTSON, J., FIGEYS, D. and COUTURE, J.F. (2011). Proteomic analyses of the SMYD family interactomes identify HSP90 as a novel target for SMYD2. *J Mol Cell Biol* 3: 301-308.

ADAMO, A., SESE, B., BOUE, S., CASTANO, J., PARAMONOV, I., BARRERO, M.J. and IZPISUA BELMONTE, J.C. (2011). LSD1 regulates the balance between self-renewal and differentiation in human embryonic stem cells. *Nat Cell Biol* 13: 652-659.

AZUARA, V., PERRY, P., SAUER, S., SPIVAKOV, M., JORGENSEN, H.F., JOHN, R.M., GOUTI, M., CASANOVA, M., WARNES, G., MERKENSCHLAGER, M. *et al.* (2006). Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* 8: 532-538.

BERNSTEIN, B.E., MIKKELSEN, T.S., XIE, X., KAMAL, M., HUEBERT, D.J., CUFF, J., FRY, B., MEISSNER, A., WERNIG, M., PLATH, K. *et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125: 315-326.

BIBIKOVA, M., LAURENT, L.C., REN, B., LORING, J.F. and FAN, J.B. (2008). Unraveling epigenetic regulation in embryonic stem cells. *Cell Stem Cell* 2: 123-134.

BOYER, L.A., LEE, T.I., COLE, M.F., JOHNSTONE, S.E., LEVINE, S.S., ZUCKER, J.P., GUENTHER, M.G., KUMAR, R.M., MURRAY, H.L., JENNER, R.G. *et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122: 947-956.

BROWN, M.A., SIMS, R.J., 3RD, GOTTLIEB, P.D. and TUCKER, P.W. (2006). Identification and characterization of Smyd2: a split SET/MYND domain-containing histone H3 lysine 36-specific methyltransferase that interacts with the Sin3 histone deacetylase complex. *Mol Cancer* 5: 26.

CARROZZA, M.J., LI, B., FLORENS, L., SUGANUMA, T., SWANSON, S.K., LEE, K.K., SHIA, W.J., ANDERSON, S., YATES, J., WASHBURN, M.P. *et al.* (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123: 581-592.

CHEUNG, P. and LAU, P. (2005). Epigenetic regulation by histone methylation and histone variants. *Mol Endocrinol* 19: 563-573.

CHO, H.S., HAYAMI, S., TOYOKAWA, G., MAEJIMA, K., YAMANE, Y., SUZUKI, T., DOHMAE, N., KOGURE, M., KANG, D., NEAL, D.E. *et al.* (2012). RB1 methylation by SMYD2 enhances cell cycle progression through an

increase of RB1 phosphorylation. *Neoplasia* 14: 476-486.
DIEHL, F., BROWN, M.A., VAN AMERONGEN, M.J., NOVOYATLEVA, T., WIETELMANN, A., HARRISS, J., FERRAZZI, F., BOTTGER, T., HARVEY, R.P., TUCKER, P.W. *et al.* (2010). Cardiac deletion of Smyd2 is dispensable for mouse heart development. *PLoS One* 5: e9748.
DONLIN, L.T., ANDRESEN, C., JUST, S., RUDENSKY, E., PAPPAS, C.T., KRUGER, M., JACOBS, E.Y., UNGER, A., ZIESENISS, A., DOBENECKER, M.W. *et al.* (2012). Smyd2 controls cytoplasmic lysine methylation of Hsp90 and myofilament organization. *Genes Dev* 26: 114-119.
FUJII, T., TSUNESUMI, S., YAMAGUCHI, K., WATANABE, S. and FURUKAWA, Y. (2011). Smyd3 is required for the development of cardiac and skeletal muscle in zebrafish. *PLoS One* 6: e23491.
GOLDBERG, A.D., ALLIS, C.D. and BERNSTEIN, E. (2007). Epigenetics: a landscape takes shape. *Cell* 128: 635-638.
GOTTLIEB, P.D., PIERCE, S.A., SIMS, R.J., YAMAGISHI, H., WEIHE, E.K., HARRISS, J.V., MAIKA, S.D., KUZIEL, W.A., KING, H.L., OLSON, E.N. *et al.* (2002). Bop encodes a muscle-restricted protein containing MYND and SET domains and is essential for cardiac differentiation and morphogenesis. *Nat Genet* 31: 25-32.
HALPERN, M.E., HO, R.K., WALKER, C. and KIMMEL, C.B. (1993). Induction of muscle pioneers and floor plate is distinguished by the zebrafish no tail mutation. *Cell* 75: 99-111.
HAMAMOTO, R., FURUKAWA, Y., MORITA, M., IIMURA, Y., SILVA, F.P., LI, M., YAGYU, R. and NAKAMURA, Y. (2004). SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. *Nat Cell Biol* 6: 731-740.
HAMAMOTO, R., SILVA, F.P., TSUGE, M., NISHIDATE, T., KATAGIRI, T., NAKAMURA, Y. and FURUKAWA, Y. (2006). Enhanced SMYD3 expression is essential for the growth of breast cancer cells. *Cancer Sci* 97: 113-118.
HU, L., ZHU, Y.T., QI, C. and ZHU, Y.J. (2009). Identification of Smyd4 as a potential tumor suppressor gene involved in breast cancer development. *Cancer Res* 69: 4067-4072.
HUANG, J., PEREZ-BURGOS, L., PLACEK, B.J., SENGUPTA, R., RICHTER, M., DORSEY, J.A., KUBICEK, S., OPRAVIL, S., JENUWEIN, T. and BERGER, S.L. (2006). Repression of p53 activity by Smyd2-mediated methylation. *Nature* 444: 629-632.
JAIN, A.K., ALLTON, K., IACOVINO, M., MAHEN, E., MILCZAREK, R.J., ZWAKA, T.P., KYBA, M. and BARTON, M.C. (2012). p53 regulates cell cycle and microRNAs to promote differentiation of human embryonic stem cells. *PLoS Biol* 10: e1001268.
JIANG, Y., SIRINUPONG, N., BRUNZELLE, J. and YANG, Z. (2011). Crystal structures of histone and p53 methyltransferase SmyD2 reveal a conformational flexibility of the autoinhibitory C-terminal domain. *PLoS One* 6: e21640.
JOPLING, C., SLEEP, E., RAYA, M., MARTI, M., RAYA, A. and IZPISUA BELMONTE, J.C. (2010). Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation. *Nature* 464: 606-609.
JOSHI, A.A. and STRUHL, K. (2005). Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Mol Cell* 20: 971-978.
KEOGH, M.C., KURDISTANI, S.K., MORRIS, S.A., AHN, S.H., PODOLNY, V., COLLINS, S.R., SCHULDINER, M., CHIN, K., PUNNA, T., THOMPSON, N.J. *et al.* (2005). Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* 123: 593-605.
KOMATSU, S., IMOTO, I., TSUDA, H., KOZAKI, K.I., MURAMATSU, T., SHIMADA, Y., AIKO, S., YOSHIZUMI, Y., ICHIKAWA, D., OTSUJI, E. *et al.* (2009). Overexpression of SMYD2 relates to tumor cell proliferation and malignant outcome of esophageal squamous cell carcinoma. *Carcinogenesis* 30: 1139-1146.
KOUZARIDES, T. (2007). Chromatin modifications and their function. *Cell* 128: 693-705.
LATINKIC, B.V. and SMITH, J.C. (1999). Goosecoid and mix.1 repress Brachyury expression and are required for head formation in Xenopus. *Development* 126: 1769-1779.
LI, D., NIU, Z., YU, W., QIAN, Y., WANG, Q., LI, Q., YI, Z., LUO, J., WU, X., WANG, Y. *et al.* (2009). SMYD1, the myogenic activator, is a direct target of serum response factor and myogenin. *Nucleic Acids Res* 37: 7059-7071.
LIPCHINA, I., ELKABETZ, Y., HAFNER, M., SHERIDAN, R., MIHAILOVIC, A., TUSCHL, T., SANDER, C., STUDER, L. and BETEL, D. (2011). Genome-wide identification of microRNA targets in human ES cells reveals a role for miR-302 in modulating BMP response. *Genes Dev* 25: 2173-2186.
MEISSNER, A. (2010). Epigenetic modifications in pluripotent and differentiated cells. *Nat Biotechnol* 28: 1079-1088.
NIWA, H. (2007). How is pluripotency determined and maintained? *Development* 134: 635-646.

ODENTHAL, J., HAFFTER, P., VOGELSANG, E., BRAND, M., VAN EEDEN, F.J., FURUTANI-SEIKI, M., GRANATO, M., HAMMERSCHMIDT, M., HEISENBERG, C.P., JIANG, Y.J. *et al.* (1996). Mutations affecting the formation of the notochord in the zebrafish, Danio rerio. *Development* 123: 103-115.
PHAN, D., RASMUSSEN, T.L., NAKAGAWA, O., MCANALLY, J., GOTTLIEB, P.D., TUCKER, P.W., RICHARDSON, J.A., BASSEL-DUBY, R. and OLSON, E.N. (2005). BOP, a regulator of right ventricular heart development, is a direct transcriptional target of MEF2C in the developing heart. *Development* 132: 2669-2678.
PRATT, T., SHARP, L., NICHOLS, J., PRICE, D.J. and MASON, J.O. (2000). Embryonic stem cells and transgenic mice ubiquitously expressing a tau-tagged green fluorescent protein. *Dev Biol* 228: 19-28.
RAYA, A., RODRIGUEZ-PIZA, I., ARAN, B., CONSIGLIO, A., BARRI, P.N., VEIGA, A. and IZPISUA BELMONTE, J.C. (2008). Generation of cardiomyocytes from new human embryonic stem cell lines derived from poor-quality blastocysts. *Cold Spring Harb Symp Quant Biol* 73: 127-135.
REITER, J.F., KIKUCHI, Y. and STAINIER, D.Y. (2001). Multiple roles for Gata5 in zebrafish endoderm formation. *Development* 128: 125-135.
SADDIC, L.A., WEST, L.E., ASLANIAN, A., YATES, J.R., 3RD, RUBIN, S.M., GOZANI, O. and SAGE, J. (2010). Methylation of the retinoblastoma tumor suppressor by SMYD2. *J Biol Chem* 285: 37733-37740.
SCHULTE-MERKER, S., VAN EEDEN, F.J., HALPERN, M.E., KIMMEL, C.B. and NUSSLEIN-VOLHARD, C. (1994). no tail (ntl) is the zebrafish homologue of the mouse T (Brachyury) gene. *Development* 120: 1009-1015.
SCOUMANNE, A. and CHEN, X. (2008). Protein methylation: a new mechanism of p53 tumor suppressor regulation. *Histol Histopathol* 23: 1143-1149.
SPIVAKOV, M. and FISHER, A.G. (2007). Epigenetic signatures of stem-cell identity. *Nat Rev Genet* 8: 263-271.
STAINIER, D.Y. (2002). A glimpse into the molecular entrails of endoderm formation. *Genes Dev* 16: 893-907.
TAN, X., ROTLLANT, J., LI, H., DE DEYNE, P. and DU, S.J. (2006). SmyD1, a histone methyltransferase, is required for myofibril organization and muscle contraction in zebrafish embryos. *Proc Natl Acad Sci U S A* 103: 2713-2718.
THOMPSON, E.C. and TRAVERS, A.A. (2008). A Drosophila Smyd4 homologue is a muscle-specific transcriptional modulator involved in development. *PLoS One* 3: e3008.
THOMSON, J.A., ITSKOVITZ-ELDOR, J., SHAPIRO, S.S., WAKNITZ, M.A., SWIERGIEL, J.J., MARSHALL, V.S. and JONES, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. *Science* 282: 1145-1147.
VOELKEL, T., ANDRESEN, C., UNGER, A., JUST, S., ROTTBAUER, W. and LINKE, W.A. (2012). Lysine methyltransferase Smyd2 regulates Hsp90-mediated protection of the sarcomeric titin springs and cardiac function. *Biochim Biophys Acta*.
WIZNEROWICZ, M. and TRONO, D. (2003). Conditional suppression of cellular genes: lentivirus vector-mediated drug-inducible RNA interference. *J Virol* 77: 8957-8961.
WU, J., CHEUNG, T., GRANDE, C., FERGUSON, A.D., ZHU, X., THERIAULT, K., CODE, E., BIRR, C., KEEN, N. and CHEN, H. (2011). Biochemical characterization of human SET and MYND domain-containing protein 2 methyltransferase. *Biochemistry* 50: 6488-6497.
ZUBER, J., RAPPAPORT, A.R., LUO, W., WANG, E., CHEN, C., VASEVA, A.V., SHI, J., WEISSMUELLER, S., FELLMANN, C., TAYLOR, M.J. *et al.* (2011). An integrated approach to dissecting oncogene addiction implicates a Myb-coordinated self-renewal program as essential for leukemia maintenance. *Genes Dev* 25: 1628-1640.

Figure Legends

Figure 1. Expression of SMYD family members during human ES cells differentiation. (a) mRNA levels of *SMYD1-5* family members in human embryonic stem cells (ES[4] and ES[2]), induced pluripotent stem cells from human keratinocytes (KiPS4F1 and KiPS4F8), human fibroblasts (HF) and human keratinocytes (HEK). **(b)** mRNA levels of *SMYD1-5* in ES[4] and ES[2] undifferentiated (d0) and at days 4, 8 and 15 of differentiation (**c**) mRNA levels of pluripotency-related genes during .ES[4] and ES[2] differentiation. Three independent differentiations were performed and one representative experiment is shown. Levels were determined by qPCR and normalized to *GAPDH*. Mean and standard deviation of triplicates are shown.

Figure 2. Analysis of histone methylation marks in the *SMYD2* promoter. ChIP assays using antibodies against H3K4me2, H3K4me3 and H3K27me3 in ES[4] and ES[2]. The presence of the indicated genes regulatory regions in the immunoprecipitated chromatin was analyzed by qPCR Values are represented as percentage of input. Mean and standard deviation corresponding to three independent experiments are shown.

Figure 3. Gain-and-loss of function of *SMYD2* during human ES cell differentiation. (a) mRNA levels of the indicated genes during the *in vitro* differentiation of SMYD2 knockdown (shSMYD2) and control (shControl) ES[4] cell lines. **(b)** mRNA levels of the indicated genes during the *in vitro* differentiation of SMYD2 overexpressing (SMYD2) and control (GFP) ES[4] cell lines. mRNA levels were measured by qPCR at the indicated days of differentiation and normalized to *GAPDH*. One representative experiment out of three independent experiments is shown. Mean and standard deviation from triplicates are shown.

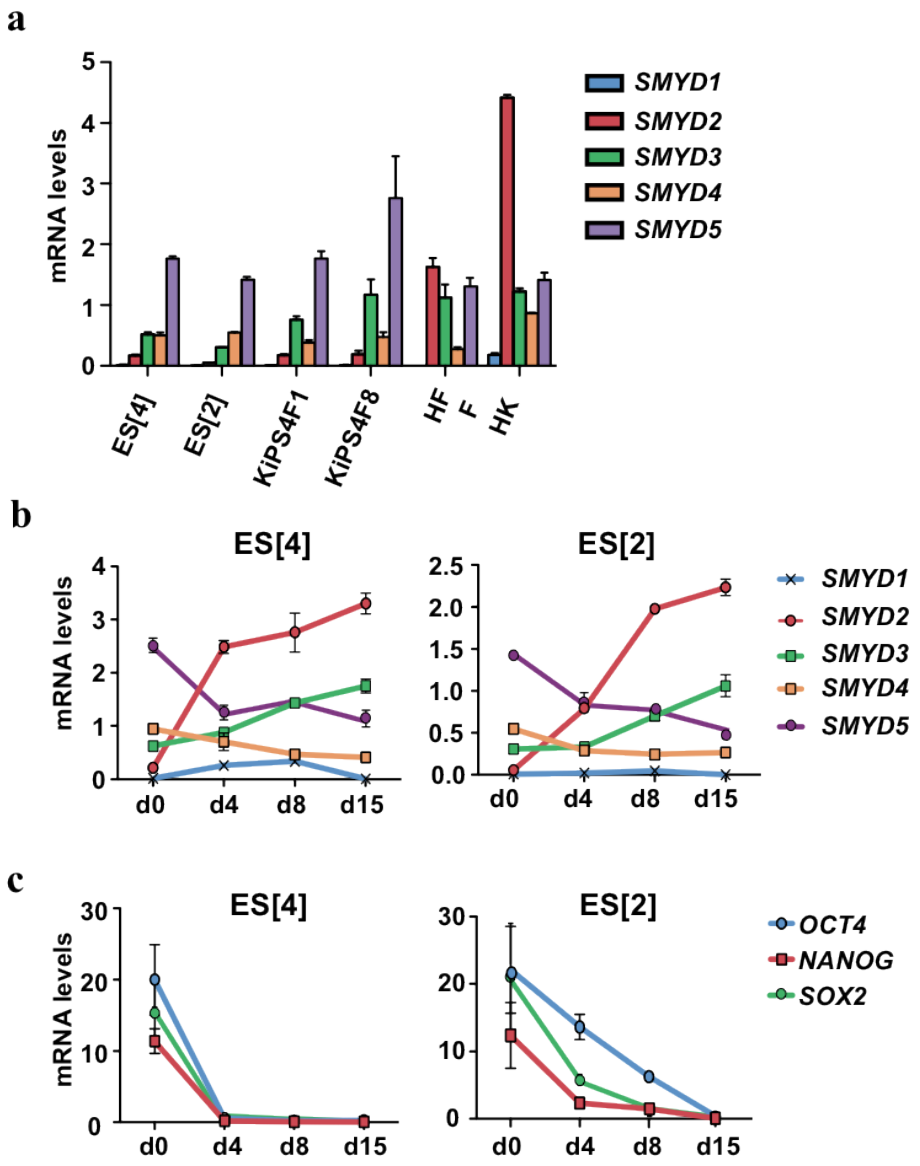
Figure 4. *smyd2a* knockdown in zebrafish development. (a) *smyd2a* and *smyd2b* mRNA levels at different hours post fertilization (hpf). Levels were determined by qPCR and normalized to 18S. Two independent experiments were carried out and one representative experiment is shown. Mean and standard deviation from two independent quantifications are shown. **(b)** The upper panel shows a schematic representation of the splice-blocking morpholino smyd2a-MO at the exon1-intron 1 junction. Specific primers for *smyd2a* were designed at each side of exon 1 and intron 1 flanking a region of 410 bp in wild type mRNA, and 4473 bp in unspliced mRNA. Lower panel shows the absence of mature *smyd2a* mRNA at 24 hpf in smyd2a-MO compared to Control-MO and non-injected embryos. Expression of *smyd2b* was unaffected by the smyd2a-MO injection. *β-actin 2* was used as control. **(c)** Different phenotypes at 5, 24, 48 hpf and 6 dpf of embryos injected with Control-MO and smyd2a-MO. **(d)** Percentage of phenotypes of morpholino injection (Control-MO and smyd2a-MO) and rescue experiment (smyd2a-MO + *SMYD2* mRNA and Control-MO + *SMYD2* mRNA) at 24 hpf. Means from three independent experiments are shown.

Figure 5. Effects of *smyd2a* knockdown on gene expression during zebrafish gastrulation. (a) *In situ* hybridization of ntl in Control-MO and smyd2a-MO injected embryos at 6 hpf. Embryos are shown in lateral views (left), and animal pole views (right). **(b)** mRNA levels of indicated genes in Control-MO and smyd2a-MO embryos at 4, 6, 8, and 10 hpf. 50 embryos were collected at each point. Mean and standard deviation of triplicate quantifications are shown.

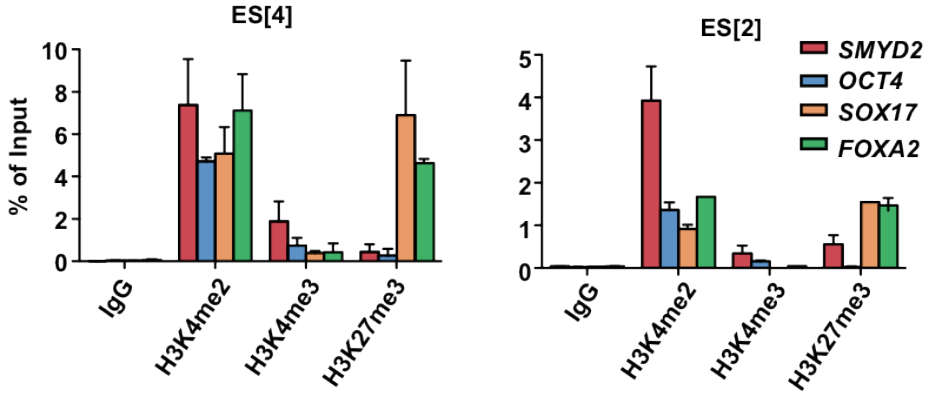
	n.	**Dead (%)**	**No. Surviving**	**wild type (%)**	**mild (%)**	**severe (%)**
Control-MO	339	9 ± 3	307	94 ± 3	6 ± 3	0 ± 0
smyd2a-MO	545	57 ± 3	238	15 ± 7	25 ± 5	60 ± 12
smyd2a-MO + *SMYD2* RNA	665	36 ± 9	436	71 ± 16	11 ± 4	18 ± 16
Control-MO + *SMYD2* RNA	512	7 ± 2	475	97 ± 3	3 ± 2	0 ± 0

Table 1: Survival rates and phenotypes of morpholino injections (Control-MO and smyd2a-MO) and rescue experiment (smyd2a-MO + *SMYD2* RNA and Control-MO + *SMYD2* RNA) at 24 hpf. Mean and standard deviation from three independent experiments are shown.

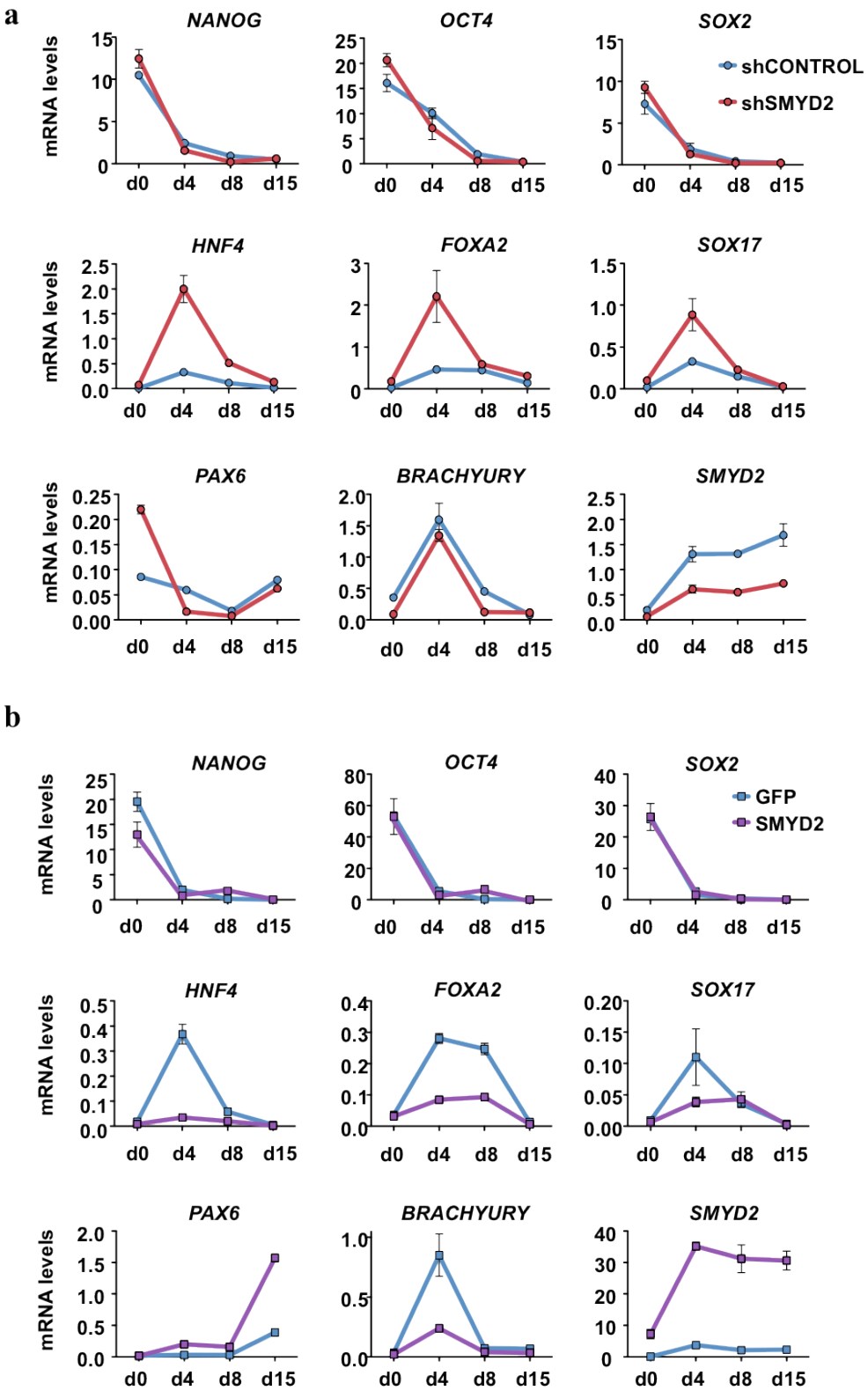
Sesé_Fig 1


a
mRNA levels
SMYD1
SMYD2
SMYD3
SMYD4
SMYD5
ES[4]
ES[2]
KiPS4F1
KiPS4F8
HF
F
HK
b
ES[4]
ES[2]
mRNA levels
d0
d4
d8
d15
c
OCT4
NANOG
SOX2

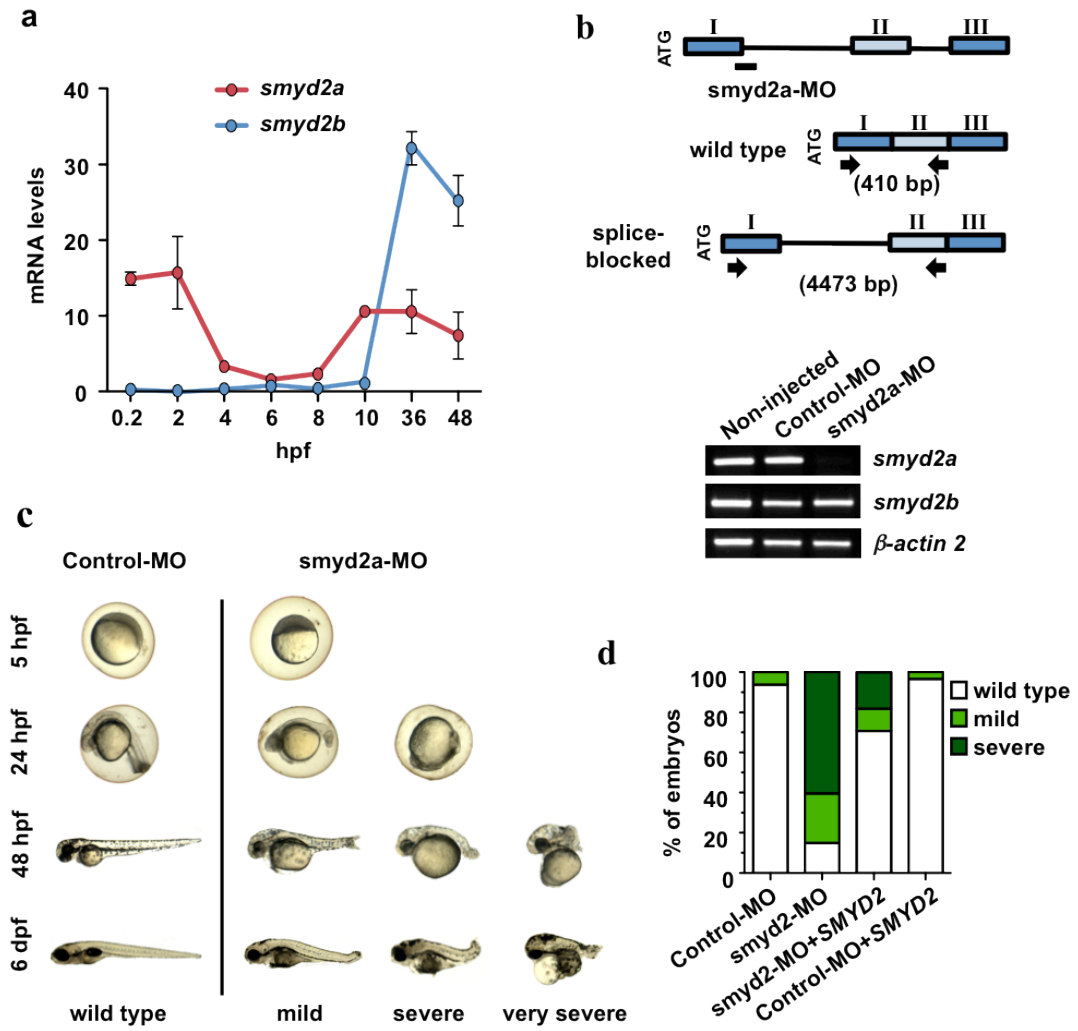
Sesé_Fig 2

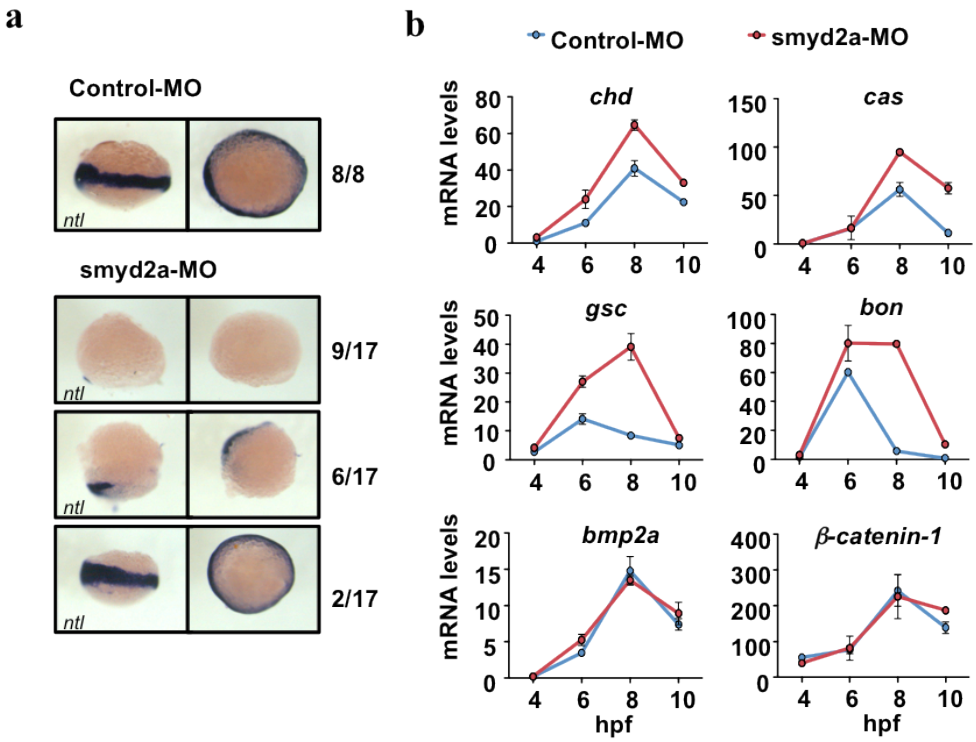

ES[4]
% of Input
10
8
6
4
2
0
IgG
H3K4me2
H3K4me3
H3K27me3
ES[2]
5
4
3
2
1
0
IgG
H3K4me2
H3K4me3
H3K27me3
SMYD2
OCT4
SOX17
FOXA2

Sesé_Fig 3


a
NANOG
OCT4
SOX2
shCONTROL
shSMYD2
HNF4
FOXA2
SOX17
PAX6
BRACHYURY
SMYD2
mRNA levels
d0 d4 d8 d15
b
NANOG
OCT4
SOX2
GFP
SMYD2
HNF4
FOXA2
SOX17
PAX6
BRACHYURY
SMYD2
mRNA levels
d0 d4 d8 d15

Sesé_Fig 4


a
mRNA levels
smyd2a
smyd2b
hpf
b
ATG
smyd2a-MO
wild type
(410 bp)
splice-blocked
(4473 bp)
Non-injected
Control-MO
smyd2a-MO
smyd2a
smyd2b
β-actin 2
c
Control-MO
smyd2a-MO
5 hpf
24 hpf
48 hpf
6 dpf
wild type
mild
severe
very severe
d
% of embryos
wild type
mild
severe
Control-MO
smyd2-MO
smyd2-MO+SMYD2
Control-MO+SMYD2


a
Control-MO
ntl
8/8
smyd2a-MO
ntl
9/17
ntl
6/17
ntl
2/17
b
Control-MO
smyd2a-MO
chd
cas
gsc
bon
bmp2a
β-catenin-1
mRNA levels
hpf