TECHNISCHE UNIVERSITÄT DRESDEN

The role of Histone H3 Lysine 4 trimethylation in zebrafish embryonic development

Dissertation

zur Erlangung des akademischen Grades

Doctor rerum naturalium
(Dr. rer. nat.)

vorgelegt

der Fakultät Mathematik und Naturwissenschaften
der Technischen Universität Dresden

von

Maximilian Krause
(Matrikel-Nr 3851535)

geboren am 08.06.1987 in
Ostseebad Kühlungsborn, Deutschland

dresden international phd program

Diese Dissertation wurde im Zeitraum vom 1. Juni 2012 bis zum 30. August 2016 unter der Betreuung von Dr. Nadine Vastenhouw am Max-Planck-Institut für molekulare Zellbiologie und Genetik angefertigt. Die Arbeit wurde durch ein Stipendium der Dresden International Graduate School for Biomedicine and Bioengineering (DIGS-BB) finanziert.

This dissertation was prepared at the Max-Planck-Institute of Molecular Cell Biology and Genetics, Dresden, between 1 June 2012 and 30 August 2016, under the supervision of Dr. Nadine Vastenhouw. The work was financially supported by a fellowship of the Dresden International Graduate School for Biomedicine and Bioengineering (DIGS-BB).

DIGS-BB

CBG Max Planck Institute of Molecular Cell Biology and Genetics

Gutachter: Prof. Dr. Francis Stewart

Prof. Dr. Karla Neugebauer

Eingereicht am: _____

Verteidigt am: _____

L'important dans la vie ce n'est point le triomphe mais le combat;
l'essentiel ce n'est pas d'avoir vaincu mais de s'être bien battu.

The important thing in life is not victory but the challenge; it is not to have vanquished but to have fought well.

Pierre de Frédy, Baron de Coubertin,
Founder of the International Olympic Committee

ACKNOWLEDGEMENTS

First and foremost I would like to thank my daughter Ida Linnea. Ida, thanks for making every day of my life with you a happy day. And moreover, thank you for making my role as a father so easy!
My wholehearted thanks and appreciation go to my partner Anika for all her scientific and personal support throughout these last years. Thanks for always being there for me, and for making our family life so enjoyable. Big thanks go to both of our families. Thanks for all the support in minor and major crisis, as well as for the cheerful time we have together. Thanks to all the friends I got to know here in Dresden. Thank you for making this time, and Dresden as a place, a special one. Thanks for distraction, support, advice and shared joy throughout that time!

I would like to thank Nadine, thanks for your constant time, support and advice during the last four years. Thanks for sharing the enthusiasm about epigenetics with me, guiding my way into this topic and the necessary technical details. Thanks for the chance to work on this project and to grow as a scientist.
Thank you very much to my TAC members Karla Neugebauer, Attila Toth and Francis Stewart. Thank you very much for scientific guidance, critical assessment of my work, but also personal words during my studies.
Thank you very much Carine for being a great colleague from start to end. You were truly a role model of a hard-working and successful PhD student to me. I will carry these memories along.
Thank you very much Sabrina, Katja and Maxi for technical support during the project, especially for the challenging task of germline transplantation. Without your constant support the success of this thesis would be smaller.
Thank you very much to present and past Vastenhouw lab, for not only being colleagues, but also friends and advisors at any time during my PhD works. You all contributed to a great working athmosphere that made the days in the lab much more enjoyable. Furthermore, thanks to all of you for critically assessing this thesis work.
Thank you very much Holger Brandl and Pranav Kulkarni, for guiding me into bioin-

formatics analyses. Without your expertise and patience I would not have been able to fulfill my studies.

A big thanks goes also to the MPI-directors, to the PhD office (namely Birgit Knepper) and to present and past Student Representatives. Thank you all for making the Dresden International PhD Program and the MPI-CBG such a special place, that combines outstanding research with a great working spirit and excellent education! I appreciated very much to be a part of it.

Last but not least my wholehearted thanks to all MPI Services and facilities, especially the fish facility around Evelyn Lehmann and the NGS-facility, namely Andreas Dahl, for all the technical assistance. These scientific groundworks not only made this work easier and better, but more importantly they made this work feasible. Thank you all for the countless efforts that went into this project.

THESIS SUMMARY

English summary

Cells within multicellular organisms share the same genetic information, yet their shape and function can differ dramatically. This diversity of form and function is established by differential use of the genetic information. Early embryonic development describes the processes that lead to a fully differentiated embryo starting from a single fertilized cell - the zygote. Interestingly, in metazoan species this early development is governed by maternally provided factors (nutrients, RNA, protein), while the zygotic genome is transcriptionally inactive. Only at a specific developmental stage, the zygotic genome becomes transcriptionally active, and zygotic transcripts drive further embryonic development. This major change is called zygotic genome activation (ZGA). While major regulators of activation of early zygotic genes could be identified recently, the molecular mechanisms that contribute to robust global genome activation during embryonic development is not fully understood.

In this study, I investigated whether the establishment of histone H3 lysine 4 trimethylation (H3K4me3) is involved in zebrafish zygotic transcription activation and early embryonic development. H3K4me3 is a chromatin modification that is implicated in transcription regulation. H3K4me3 has been shown to be enriched at Transcription Start Sites (TSS) of genes prior to their activation, and is postulated facilitate transcription activation of developmentally important genes. To interfere with H3K4me3 establishment, I generated histone methyltransferase mutants. I further inhibited H3K4me3 establishment by introduction of histones with lysine 4-to-methionine (K4-to-M) substitution, which act as dominant-negative inhibitors of H3K4me3 establishment. Upon H3K4me3 reduction, I studied the resulting effect on early transcription activation. I found that H3K4me3 is not involved in transcription activation during early zebrafish embryogenesis. Finally I analyzed possible cues in DNA sequence and chromatin environment that might favor early H3K4me3 establishment.

These studies show that H3K4me3 is established during ZGA, yet it is not involved in transcription activation during early zebrafish development. Establishment of H3K4me3 might be a consequence of histone methyltransferase recruitment during a permissive chromatin state, and might be targeted to CpG-rich promoter elements that are enriched for the histone variant H2A.z.

German summary

Jede Zelle eines multizellulären Organismus enthält dieselbe Erbinformation, und doch können Form und Funktion von Zellen untereinander sehr unterschiedlich sein. Diese Diversität wird durch unterschiedliches Auslesen - Transkribieren - der Erbinformation erreicht. Embryogenese beschreibt den Prozess, der aus einer einzelnen Zelle - der Zygote - einen multizellulären Embryo entstehen lässt. Interessanterweise laufen frühe Stadien der Embryogenese ohne Transkription der embryonalen Erbinformation ab, sondern werden durch maternal bereitgestellte Faktoren ermöglicht. Erst nach einer spezies-spezifischen Entwicklungsphase wird das Erbgut der Zygote aktiv transkribiert und ermöglicht die weitere Embryonalentwicklung. Obwohl bereits wichtige Regulatoren dieser globalen Genomaktivierung identifiziert werden konnten, sind viele molekulare Mechanismen, die zur Aktivierung des zygotischen Genoms beitragen, bisher unbekannt.

In der hier vorliegenden Doktorarbeit habe ich die Rolle von Histon H3 Lysin 4 Trimethylierung (H3K4me3) während der frühen Embryogenese des Zebrafischs untersucht. H3K4me3 ist eine Chromatinmodifikation, die mit aktiver Transkription in Verbindung gebracht wird. H3K4me3 ist an Transkriptions-Start-Stellen von aktiv ausgelesenen Genen angereichert und es wird vermutet, dass diese Modifikation das Binden von Transkriptionsfaktoren und der Transkriptionsmaschinerie erleichtert.

Während meiner Arbeit habe ich durch Mutation verschiedener Histon-Methyltransferasen beziehungsweise die Überexpression eines dominant-negativen Histonsubstrats versucht, die Etablierung von H3K4me3 in frühen Entwicklungsstadien des Zebrafischs zu verhindern. Anschliessend habe untersucht, welchen Effekt H3K4me3-Reduktion auf Tranksriptionsaktivität entsprechender Gene hat. Allerdings konnte ich keinen Zusammenhang zwischen H3K4me3-Reduktion und Transkriptionsaktivität beobachten. Um herauszufinden, weshalb H3K4me3 dennoch während früher Embryonalstadien etabliert wird, habe ich nachfolgend untersucht, ob möglicherweise bestimmte DNA-Sequenzen oder Chromatin-Modifikationen zur Etablierung von H3K4me3 wÃhrend der Embryogenese des Zebrafischs beitragen.

Aus der hier vorliegenden Arbeit lässt sich schlussfolgern, dass H3K4me3 in Tranksriptionsaktivierung während früher Embryonalstadien des Zebrafischs nicht involviert ist. Möglicherweise wird H3K4me3 in diesen Stadien in einer permissiven Chromatinumgebung etabliert, bevorzugt an Promotoren mit starker H2A.z-Anreicherung und CpG-reichen DNA-Elementen.

Table of Contents

Frontmatter II

Acknowledgements . . . VII
Thesis Summary (English) . . . IX
Thesis Summary (German) . . . X
Table of Contents . . . XIV
List of Figures . . . XVI
List of Tables . . . XVII
List of Abbreviations . . . XXIII

1 Introduction 1

1.1 Transcription regulation . . . 2
1.1.1 Promoter elements - genetic information that guides transcription initiation . . . 2
1.1.2 Enhancers - fine-tuning of transcription by distal DNA elements 3
1.1.3 CpG islands - DNA sequences that allow for epigenetic regulation 4
1.2 Chromatin . . . 4
1.2.1 Histone variants . . . 7
1.2.2 Posttranslational histone modifications . . . 7
1.2.3 Histone Lysine methylation . . . 8
1.2.4 H3K4me3 in embryonic development . . . 10
1.3 Establishment and removal of H3K4me3 . . . 10
1.3.1 Set1 homologs - Set1a and Set1b . . . 11
1.3.2 Trithorax homologs - Mll1 and Mll2 . . . 11
1.3.3 Homologs of Trithorax-related - Mll3 and Mll4 . . . 13
1.3.4 COMPASS complex proteins . . . 13
1.3.5 H3K4me3 removal. . . . 14
1.4 Transcription activation in embryos . . . 14
1.4.1 Zebrafish early embryonic development . . . 15
1.4.2 H3K4me3 during early zebrafish development . . . 17
1.5 Thesis aim . . . 17

2 Materials and Methods 19
2.1 Materials . . . 19
2.2 Methods . . . 36
2.2.1 Zebrafish husbandry and care . . . 36
2.2.2 Generation of zebrafish knock-out lines by TALEN mutagenesis 36
2.2.3 Generation of plasmids for mRNA production . . . 38
2.2.4 Microinjection . . . 39
2.2.5 Germline transplantation . . . 39
2.2.6 Western Blot Assays . . . 40
2.2.7 RNA extraction and quantification assays . . . 41
2.2.8 Chromatin immunoprecipitation (ChIP) . . . 43
2.3 Bioinformatics Analyses . . . 46
2.3.1 Quality control, alignment and peak calling . . . 46
2.3.2 Lambda normalization . . . 46
2.3.3 Differential ChIP enrichment analysis . . . 47
2.3.4 Data integration . . . 47
2.3.5 Gene classification . . . 48

3 Results I: H3K4me3 interference by Histone methyltransferase mutation 49
3.1 Generation and phenotypic description of histone methyl-transferase mutants . . . 49
3.1.1 HMT TALEN mutagenesis workflow . . . 49
3.1.2 *Ash2l* TALEN mutation does not result in a larval or adult phenotype . . . 52
3.1.3 *Mll2* mutation results in increased larval mortality, while adult fish are healthy and fertile . . . 54
3.1.4 *Mll1* mutation results in increased larval mortality and a severe adult phenotype . . . 56
3.2 HMT mutations do not affect global H3K4me3 levels in early zebrafish embryos . . . 60
3.3 *Mll1* mutation results in local H3K4me3 reduction of a small subset of genes . . . 62
3.4 Early embryonic transcription is not altered in *mll1* maternal-zygotic mutants . . . 67

3.5 Conclusion 70

4 Results II: H3K4me3 interference by introduction of HMT inhibitors 71
4.1 Establishing a Western Blot assay to monitor H3K4me3 reduction . . . 71
4.2 Overexpression of H3K4-specific histone demethylases does not result in global H3K4me3 reduction 73
4.3 Global reduction of H3K4me3 could not be achieved by small-molecule inhibition of HMT activity 75
4.4 Overexpression of K4-specific methylation-defective H3 results in global H3K4me3 reduction 76
4.4.1 Overexpression of H3K4-to-E constructs does not affect global H3K4me3 establishment 76
4.4.2 H3K4-to-M constructs act as dominant-negative substrate for H3K4me3 establishment 77
4.5 H3K4me3 levels at gene promoters are reduced upon introduction of methylation-defective Histone H3 79
4.6 Early transcription activation is not altered upon K4M overexpression . 88
4.7 Conclusion 92

5 Results III: Promoters rich in CpG and H2A.z gain H3K4me3 early 93
5.1 H3K4me3 levels increase over developmental time at all gene classes . 93
5.2 H3K4me3 is gained at CpG-rich elements 98
5.3 H2A.z marks overlaps with H3K4me3 at promoters of non-transcribed genes 100
5.4 High CpG density and H2A.z enrichment are predictive for H3K4me3 establishment 101
5.5 Maternally provided genes are enriched for H2A.z and CpG content . . 103
5.6 Conclusion 104

6 Discussion 105
6.1 Neither Mll1 nor Mll2 are the main histone methyltransferase for H3K4me3 establishment in early zebrafish development 106
6.2 H3K4me3 reduction does not affect transcription initiation during genome activation 107
6.3 The timing of H3K4me3 establishment might be determined by a permissive chromatin state 109

6.4 H3K4me3 potentially gains importance during later developmental stages 111
6.5 CpG-content and H2A.z enrichment might be predictive for H3K4me3 establishment during genome activation 112
6.6 Conclusion 115

Appendix 117

Bibliography 139

Authorship Declaration 159

LIST OF FIGURES

1.1 Chromatin organization and promoter structure in higher eukaryotes . . 6
1.2 Posttranslational histone modifications 8
1.3 The COMPASS complexes 12
1.4 Zebrafish embryonic development 16

3.1 Description of TALEN mutagenesis 51
3.2 *Ash2l* TALEN mutations 53
3.3 *Mll2* TALEN mutations 55
3.4 *Mll1* mutations 57
3.5 *Mll1* mutation results in a severe adult phenotype 59
3.6 HMT mutation does not result in reduction of global H3K4me3 levels . 61
3.7 H3K4me3 is enriched at promoters in wildtype and *mll1* mutant embryos 64
3.8 Fraction of promoters with changes in H3K4me3 levels 66
3.9 Change of H3K4me3 levels in *mll1* mutants does not affect transcription 68
3.10 Transcription of early activated genes is not changed upon *mll1* mutation 69

4.1 Western Blot assay to monitor H3K4me3 reduction 72
4.2 Global levels of H3K4me3 are not affected by KDM5 overexpression or injection of N-terminal histone tails 74
4.3 H3K4M, but not H3K4E introduction leads to global H3K4me3 reduction 78
4.4 Lambda DNA spike-in allows for correction of overamplification during Sequencing 80
4.5 H3K4me3 is reduced at promoter regions upon K4M introduction . . . 83
4.6 K4M overexpression reduces H3K4me3 levels reproducibly at TSS . . . 84
4.7 K4M overexpression reduces H3K4me3 levels reproducibly at TSS . . . 86
4.8 The majority of promoters shows reduced H3K4me3 levels 87
4.9 K4M overexpression does not result in transcriptional change of zygotic genes 88
4.10 Transcriptional change does not correlate with change in promoter H3K4me3 levels 90

4.11 Local H3K4me3 reduction does not result in transcriptional change of target genes 91

5.1 Maternal genes are overrepresented in H3K4me3 enrichment 95
5.2 H3K4me3 is enriched at CpG-rich promoters 99
5.3 H2A.z is enriched at promoter regions and overlaps with H3K4me3 enrichment 101
5.4 H2A.z and CpG density predict H3K4me3 establishment 102
5.5 H2A.z and CpG density are overrepresented in maternal genes 103

6.1 Model for the role of H3K4me3 in embryonic development 111

LIST OF TABLES

3.1 Expression of COMPASS components during zebrafish embryogenesis . 50

4.1 Lambda Phage library size reflects differences in sample DNA content . 82

LIST OF ABBREVIATIONS

3D	Three-dimensional
ash2l	**A**bsent, **s**mall or **h**omeotic transformation 2-**l**ike
ATAC-Seq	**A**ssay for **t**ransposase-**a**ccessible **c**hromatin followed by high-throughput **seq**uencing
BRE	TFII **B** **R**ecognition **E**lement
BSA	Bull serum albumine
C-terminal	Carboxy-terminal - end of a peptide chain or protein
cDNA	complementary DNA, DNA synthesized from mRNA
CHD1	**Ch**romo-**d**omain helicase DNA binding protein 1
ChIP	Chromatin immunoprecipitation
ChIP-Seq	Chromatin immunprecipitation followed by deep sequencing
COMPASS	**Com**plex of **P**roteins **as**sociated with **S**et1
CpG	Cytosine-Guanine dinucleotides
CpG islands	DNA sequences enriched in Cytosine-Guanine dinucleotides
DNA	Deoxy-ribonucleic acid
DPE	**D**ownstream **P**romoter **E**lement
dpf	Days post fertilization
Dpy-30	Dumpy(dpy)-morphology gene 30
ECL	Enhanced chemiluminescence
ENU	*N*-ethyl-*N*-nitrosourea

FC	Fold change
FLAG-tag	Flag peptide sequence, protein tag (DYKDDDDK)
FPKM	Fragments per kilobase per million (mapped reads)
g	Gram *or* gravity as standard acceleration force
GFP	Green-fluorescent protein
H3	histone H3
H3HA	mRNA construct encoding zebrafish H3.1 sequence tagged with HA sequence
H3K27me3	H3 lysine 27 trimethylation
H3K4me1	H3 lysine 4 monomethylation
H3K4me2	H3 lysine 4 dimethylation
H3K4me3	H3 lysine 4 trimethylation
H3K9me3	H3 lysine 9 trimethylation
HA-tag	Human influenza hemagglutinin peptide sequence, protein tag (YPYDVPDYA)
HMT	Histone methyltransferase
HP1	**H**eterochromatin **P**rotein 1
hpf	Hours post fertilization
Inr	Initiator Sequence
IP	Immunoprecipitation
Jarid1	**J**umonji **A**T-**R**ich **I**nteractive **D**omain 1 - H3K4-specific histone demethylase
K4E	mRNA construct encoding zebrafish H3.1 sequence tagged with HA sequence, with Lysine-to-Glutamate (K-to-E) substitution at amino acid 4

K4M	mRNA construct encoding zebrafish H3.1 sequence tagged with HA sequence, with Lysine-to-Methionine (K-to-M) substitution at amino acid 4
kb	Kilobases, 1000 bases nucleotide sequence
KDM5	Lysine (K) demethylase family 5
Lsd1	**L**ysine-**s**pecific **d**emethylase 1 - H3K4-specific histone demethylase
mA	milli-Ampere
Mb	Megabases, 1 million bases nucleotide sequence
MBT	Mdi-Blastula transition
mll	Mixed-lineage leukemia *or* myeloid/lymphoid leukemia
MNase	Micrococcal Nuclease
mRNA	messenger RNA
MZT	Maternal zygotic transition
N-terminal	Amino-terminal - beginning of a peptide chain or protein
NCOA6	**N**uclear receptor **Co-A**ctivator 6
ng	Nanogram
nl	Nanoliter
PA1	PTIP Associated protein 1
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PRC2	**P**olycomb **R**epressive **C**omplex 2
PTIP / PAXIP1	**P**ax **t**ranscription factor **i**teracting **p**rotein / **PAX I**nteracting **P**rotein 1
PVDF	Polyvinylidene fluoride

qRT-PCR	quantitative Real-Time PCR
Rbbp5	**R**etino**b**lastoma **b**inding **p**rotein 5
RNA	Ribonucleic acid
RNA-PolII	RNA-Polymerase II
RPKM	Reads per kilobase per million (mapped reads)
rpm	Revolutions per minute
s.d.	Standard deviation
SAM	*S*-Adenosyl methionine
SDS	Sodium dodecyl sulfate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
SET domain	'**S**u(var)3-9'-'**E**nhancer-of-Zeste'-'**T**rithorax' domain
SWR1	**Sw**i/Snf2-**r**elated ATPase - Chromatin remodeler
TALEN	**T**ranscription **a**ctivator-**l**ike **e**ffector **n**uclease
TATA	TATA-Box element - A/T rich DNA sequence in promoters
TBP	**T**ATA-Box **B**inding **P**rotein
TBS	Tris-buffered saline
TF	Transcription Factor
TFII A-D	Transcription Factor II A-D - general Transcription factors that build the Preinitiation complex with RNA-Polymerase II
Trr	Trithorax-related - H3K4-specific histone methyltransferase in *Drosophila melanogaster*
Trx	Trithorax - H3K4-specific histone methyltransferase in *Drosophila melanogaster*
TSS	Transcription start site

ug	Microgram
ul	Microliter
UTR	Untranslated region (of the respective mRNA)
V	Volt
Wdr5	**WD R**epeat containing protein 5
Wdr82	**WD R**epeat containing protein 82
ZGA	Zygotic genome activation

1 INTRODUCTION

Introduction Summary

Early embryonic development of most vertebrate species is governed by maternally provided factors (nutrients, RNA, protein) while the zygotic genome is transcriptionally inactive. Only at a specific developmental stage, the zygotic genome becomes transcriptionally active, and zygotic transcripts drive further embryonic development. This major change is called zygotic genome activation (ZGA). While major players in activation of early zygotic genes could be identified recently, the molecular mechanisms that contribute to robust global genome activation during embryonic development are not fully understood. Previous studies suggest that chromatin is involved in regulating this major transcriptional change. During ZGA, multiple histone modifications involved in transcription regulation are established, raising the possibility that these are involved in zygotic genome activation.
In this study I focus on the role of Histone H3 Lysine 4 trimethylation (H3K4me3) during zebrafish embryonic development. H3K4me3 was shown to be enriched at Transcription Start Sites (TSS) of genes prior to their activation, and is postulated facilitate transcription activation of developmentally important genes. To address whether H3K4me3 is involved in zygotic genome activation, I aim to interfere with H3K4me3 establishment in zebrafish embryonic development and study the resulting effects on transcription activation.

In this introduction I will briefly describe mechanisms of transcription regulation. I will summarize our knowledge about chromatin and its influence on transcription, with emphasis on the role of histone modifications. I will then introduce the histone modification H3K4me3, and its establishment and removal. Finally I will introduce embryonic transcription activation, with a focus on the zebrafish embryonic development as a model to study embryonic transcription regulation.

1.1 Transcription regulation

In multicellular organisms, the majority of cells share the same genetic information, yet their shape and function can differ dramatically. This diversity of form and function is established by differential use of the genetic information. Transcription refers to the process of the read-out of the genetic information - the DNA - into functional transcripts - mRNAs - that are the templates for protein synthesis. Understanding regulation of transcription is fundamental to understand the biological complexity that is the basis of multicellular life. In the following section I will briefly introduce different aspects of transcription regulation.

1.1.1 Promoter elements - genetic information that guides transcription initiation

Transcription is initiated at gene promoters. The main role of the promoter sequence is to recruit and position the basal transcription machinery, which allows for transcription of the DNA into mRNA by RNA-Polymerase II (Myer and Young, 1998; Tolunay et al., 1984). The core promoter is the minimal DNA region that recruits the basal transcription machinery. Four DNA sequences contribute to core promoters: the TATA box, the B-recognition element, the Initiator element (Inr) and the Downstream Promoter Element (DPE) (Figure 1.1). These sequences specify the binding of the general transcription factors (TF) TBP and TFII A-D to the gene transcription start site (TSS), that together help to unwind the DNA and recruit RNA-Polymerase II (Kutyavin et al., 2000). While all of these sequences are assigned to important regulatory functions, not all promoters contain all elements.

The TATA-box is an AT-rich sequence positioned 25-35 basepairs (bp) upstream of the TSS (Lifton et al., 1978). It was the first promoter element to be identified. While it was long thought to be an essential element for transcription initiation, recent studies have shown that only 32% of mammalian promoters include this promoter element (Suzuki et al., 2001). The TATA-box recruits the TATA binding protein (TBP) to the core promoter.

The Inr element is located at or immediate adjacent to the TSS. It is recognized by TBP-associated transcription factors of the TFIID complex (Reinberg et al., 1987). This element is especially important for the activation of genes that do not harbor TATA-Box motifs (Yang et al., 2007). It specifies accurate initiation and strength of core promoters

(Hahn et al., 1985; Nagawa and Fink, 1985).
The Downstream Promoter Element is centered 30bp downstream of the TSS, and is involved in recruitment of transcription factors of the TFIID family (Burke and Kadonaga, 1997). It is conserved from *Drosophila* to humans and is typically - but not exclusively - found in TATA-less promoters (Kutach and Kadonaga, 2000).
The B-recognition element is a GC-rich DNA sequence located 35-45 bp upstream of the TSS (Lagrange et al., 1998). It is the only core promoter element that is not bound by TFIID-like transcription factors, but by TFIIB. TFIIB cooperates and stabilizes TFIID complex binding, especially in the absence of TATA-box sequences (Lagrange et al., 1998; Tsai and Sigler, 2000).
Next to core promoter sequences that allow for the recruitment of the basal transcription machinery, the general promoter can further contain recognition sequences for multiple other transcription factors, that allow lineage-specific and environment-specific regulation of transcription by integration of multiple signals into the transcriptional output of RNA-Polymerase II.

1.1.2 Enhancers - fine-tuning of transcription by distal DNA elements

While promoters are regulatory DNA sequences in the direct vicinity of the TSS, other more distant DNA sequences can fine-tune transcriptional activity of respective genes. These so-called enhancers (of transcription) are short DNA sequences (50-1500bp) that harbor transcription factor binding sites. Enhancers are *in cis*-acting sequences that influence transcription of genes in their local surrounding, however their regulatory function can extend up to 1 Mb distance (Blackwood and Kadonaga, 1998). By three-dimensional reorganization of the DNA, transcription factors bound to enhancers are brought in close vicinity to the basal transcription machinery, and influence its activity (Li and Noll, 1994; Merli et al., 1996). Enhancer sequences can be found upstream and downstream of gene bodies, but also within introns of the respective gene. Furthermore, a gene can have multiple enhancers, which allow integration of multiple stimuli. By combining binding sites for different transcription factors, enhancers can fine-tune expression of specific genes and thus allow for a higher flexibility in transcription regulation.

1.1.3 CpG islands - DNA sequences that allow for epigenetic regulation

The genetic information of individual cells (genotype) alone is not sufficient to describe the variety of cellular shape and function (phenotype) found in multicellular organisms. This phenomenon was first described by Conrad Waddington, who proposed the existence of an 'epigenotype' above the genetic information (Waddington, 2012). Epigenetics refers to transmittable mechanisms of transcription regulation that are not directly encoded in the DNA sequence, but allow for differential use of the genetic information.

Interestingly, DNA itself can confer inheritable changes in transcription regulation without change of the underlying coding sequence. Specific bases within the DNA can be methylated. DNA methylation is restricted to cytosine in eukaryotes, where it typically occurs at CpG dinucleotides. CpG dinucleotides are underrepresented in vertebrate genomes (Lander et al., 2001), because methylated cytosine is prone to spontaneous deamination, resulting in replacement by thymine during subsequent rounds of replication. While CpG dinucleotides are generally underrepresented, they are found in high density at specific genomic regions, often in gene promoters (Takai and Jones, 2002). These so-called CpG islands are at least 200bp long and show a higher-than-expected ratio of CpG dinucleotides (Saxonov et al., 2006). CpG islands are highly conserved throughout evolution.

In general, DNA methylation is associated with repressed transcription of respective genes by preventing transcription factors from binding to their target DNA (Comb and Goodman, 1990; Prendergast et al., 1991). Furthermore, DNA methylation is a hub for recruitment of repressive transcription regulators (Jones et al., 1998; Sansom et al., 2007). However, the highly conserved CpG islands can be differentially methylated, and unmethylated CpG islands allow for transcription of their associated genes. Thus, DNA methylation is understood as a constitutive repressive modification, and CpG islands are conserved regulatory elements in the vicinity of gene promoters.

1.2 Chromatin

Another mechanism that allows epigenetic transcription regulation is the change of chromatin structure. Chromatin is the complex of DNA and DNA-associated proteins. The basic subunit of chromatin is the nucleosome, which consists of each two histone proteins H2A, H2B, H3 and H4 (Kornberg, 1974). Their globular domains form a

positively charged octameric complex around which DNA is wound in 1.65 helical turns (Luger et al., 1997). Nucleosome-compacted DNA forms a beads-on-the string structure of 11nm diameter. The DNA between two nucleosomes is of variable length, and can be bound by linker histone H1 (Allan et al., 1981; Morris, 1976). Nucleosome-nucleosome contacts allow for a three-dimensional folding that leads to compaction into the chromatin fiber (Nishino et al., 2012), which is further compacted by loop formation on a chromatin scaffold, ultimately forming the chromosome.

Chromatin exists in two main states, euchromatin and heterochromatin (Passarge, 1979). Euchromatin is a less compacted form of chromatin, and is thought to be the predominant state (Abdellah et al., 2004). Euchromatin is enriched in genes that are actively transcribed, and its decompacted state might allow for easier recruitment of regulatory factors as well as the transcription machinery.
Heterochromatin refers to a tightly compacted state of chromatin, which is mainly refractory to transcription (Volpe et al., 2002). Heterochromatin can be further subdivided in two classes: constitutive heterochromatin, which is thought to be stable within one species and constitutes long-term repression, and facultative heterochromatin, which is not consistent between cell types and can be dynamically regulated (Grewal and Jia, 2007).
Chromatin poses a physical barrier to DNA binding and read-out (Churchman and Weissman, 2011; Weber et al., 2014). Processes that change chromatin compaction are involved in transcription regulation by allowing easier access to genomic information (reviewed in Narlikar et al. 2013).

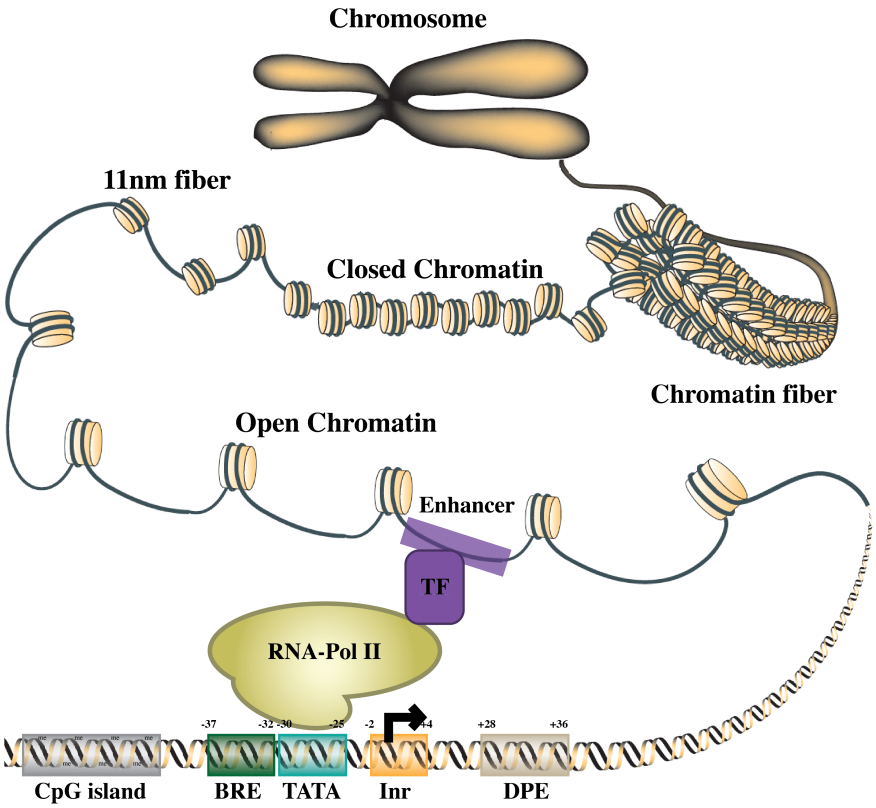


Figure 1.1: Chromatin organization and promoter structure in higher eukaryotes DNA is compacted into the small volume of the nucleus by tight association to nucleosomes. Nucleosomes and DNA form a beads-on-the-string structure, called the 11nm fiber, that is further compacted into the chromatin fiber by internucleosomal interactions. This chromatin fiber forms loops on a scaffold, that ultimately compacts chromatin into the chromosome. Chromatin compaction hinders DNA read-out, while open chromatin allows for DNA-association and gene transcription. Transcription start sites (TSS) are surrounded by regulatory sequences. The union of regulatory sequences at TSS is called promoter. The core promoter of eukaryotes can contain Initiator sequence (Inr), TATA-box (TATA), Trancription factor II B recoginition element (BRE) and the Downstream Promoter Element (DPE). Together, these regulatory sequences allow for binding of general transcription factors, and ultimately RNA-Polymerase II recruitment. Enhancers are DNA sequences that can be situated in up to one megabase distance to TSS. These can recruit further regulatory proteins, that interact with the transcription machinery by threedimensional chromatin rearrangements. CpG islands are sites of epigenetic transcription regulation by differential DNA-methylation. Adapted from Sparmann and van Lohuizen 2006.

1.2.1 Histone variants

One mechanism to regulate DNA accessibility is by nucleosome replacement. Besides canonical histones H2A, H2B, H3 and H4, replacement histones with slightly different protein composition exist (reviewed in Kamakaka and Biggins 2005). The best-studied variants are histone H3.3, a replacement for canonical H3, and H2A.z for histone H2A (Grove and Zweidler, 1984; Wu et al., 1982). Both histone variants are incorporated at specific genomic locations. H3.3 is one of the most conserved proteins in eukaryotes (Malik and Henikoff, 2003). It is incorporated into chromatin in a replication-independent manner at actively transcribed genes (Ahmad and Henikoff, 2002).
H2A.z has high similarity to canonical H2A in its globular domain, but differs in its C-terminal domain structure (Clarkson et al., 1999; Subramanian et al., 2013). H2A.z is incorporated at TSS via histone exchange by the nucleosome remodeling complex SWR1 (Kobor et al., 2004; Mizuguchi et al., 2004).
Together, H3.3 and H2A.z were found to destabilize nucleosomes by weakening intranucleosomal interactions *in vitro*, and thus increase nucleosome turnover and DNA accessibility (Jin and Felsenfeld, 2007). Furthermore, H3.3 and H2A.z were found to overlap with active histone modifications at TSS (Creyghton et al., 2008; McKittrick et al., 2004).
In conflict with these findings, nucleosomes containing two H2A.z proteins but no H3.3 have been shown to be stabilized (Park et al., 2004; Thakar et al., 2010). Additionally, H2A.z-containing nucleosomes favor a compact chromatin fiber formation by stabilizing inter-nucleosomal interactions (Park et al., 2004). H2A.z has been further shown to favor recruitment of proteins involved in transcriptional repression (Creyghton et al., 2008; Fan et al., 2004; Surface et al., 2016). Unifying these conflicting results, it has been hypothesized that H2A.z allows for a switch-like behaviour in transcription regulation (Subramanian et al., 2015). While H2A.z incorporation alone seems to favor transcriptional repression, in combination with H3.3 and possible other activating signals it favors DNA accessibility and thus facilitates transcription activation.

1.2.2 Posttranslational histone modifications

Another mechanism how chromatin compaction could be regulated is posttranslational histone modification. Histones contain flexible N-terminal tails which protrude from the globular structure. While these N-terminal tails are not essential for nucleosome formation (Ausio et al., 1989)), they are hubs for various posttranslational modifications

(reviewed in Bannister and Kouzarides 2011). These modifications can influence the local compaction of chromatin. Histone lysine acetylation for example changes the net charge of basic histones, which weakens the electrostatic interaction with the negatively charged DNA (Grunstein, 1997; Sterner and Berger, 2000).
However, not all histone modifications directly affect nucleosome-DNA interaction. The molecular mechanisms by which histone modifications regulate transcription regulation are often elusive. Their general role is often assessed by their local enrichment in either euchromatin or repressive heterochromatin, or by correlation with transcription activity of respective genes. However, functional validation of these hypotheses are scarce. Most histone modifications elicit their effects by serving as a recruitment platform for effectors of chromatin accessibility and transcription regulation (Bannister and Kouzarides, 2011). This is especially exemplified by histone lysine methylation.

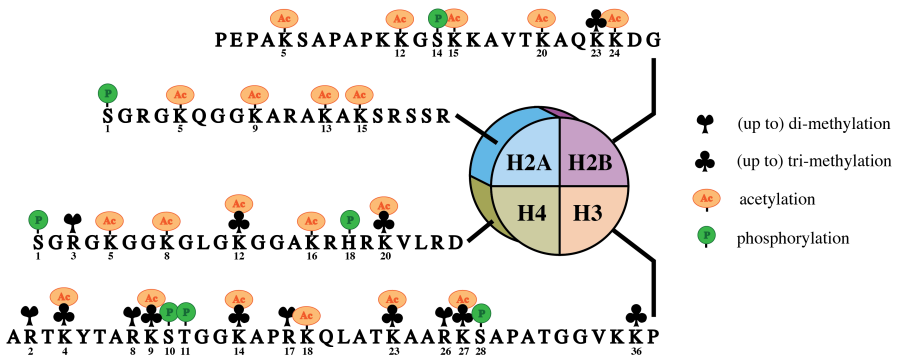


Figure 1.2: Posttranslational histone modifications. The N-terminal tails of histones protrude from the octameric protein complex. These tails are sites for various posttranslational modifications, from which the most common modifications are depicted.

1.2.3 Histone Lysine methylation

Histone methylation was among the first modifications to be identified, and is conserved in all eukaryotes (Allfrey et al., 1964; Woo and Li, 2012). Lysine residues within histones can be mono-, di- or trimethylated at the ε-amino group. Histone lysine methylation is believed to be the primary determinant of chromatin structure (K et al., 2001). In contrast to histone acetylation, transcription regulation by histone lysine methylation is multifaceted. The effect of histone lysine methylation depends on the

exact positon within histone proteins, and specific binding partners.
Trimethylation at histone H3 lysine 9 (H3K9me3), is associated with transcriptionally inactive genes (Barski et al., 2007). H3K9me3 is considered to be a hallmark of condensed, transcriptionally inactive constitutive heterochromatin. H3K9me3 has been shown to recruit heterochromatin protein 1 (HP1), which is involved in formation and maintenance of heterochromatin (Lehnertz et al., 2003). Together, H3K9me3 and HP1 are involved in establishment of DNA methylation, which further consolidates transcription repression (Lehnertz et al., 2003).
Like H3K9me3, H3K27me3 is found to be associated with inactive gene promoters (Barski et al., 2007). H3K27me3 is established by the repressive protein complex PRC2 (Margueron and Reinberg, 2011). In contrast to H3K9me3, it is mainly found in facultative heterochromatin. H3K27me3 is thought to be permissive for chromatin regulation, and can be lost at developmentally important genes during cell differentiation (Bernstein et al., 2006).
Histone trimethylation at lysine 36 (H3K36me3) is enriched in gene bodies of actively transcribed genes (Bannister et al., 2005). H3K36me3 has been shown to be the result of elongating RNA Polymerase II activity (Joshi and Struhl, 2005), and is experimentally used as marker for active transcription (Zhang et al., 2014). However, its general role in transcription regulation is repressive, as it recruits histone deacetylases (Carrozza et al., 2005) and H3K9 methylation to its chromatin locations to prevent erroneous transcription initiation within gene bodies (Bartke et al., 2010).
Methylation at lysine 4 of histone H3 is implicated in different regulatory processes depending on its methylation state. H3K4 monomethylation (H3K4me1) in the absence of di- and trimethylation is found at enhancer elements (Heintzman et al., 2007), where it is thought to recruit histone acetyltransferases (Jeong et al., 2011). Furthermore, H3K4me1 has been found to prevent recruitment of repressive chromatin modifiers to enhancer sites (Lan et al., 2007; Nady et al., 2011).
H3K4 dimethylation (H3K4me2) is predominantly found in broad domains at promoter regions of actively transcribed genes (Bernstein et al., 2005; Heintzman et al., 2007). Little is known about specific functions of H3K4me2, as proteins that show binding affinity to H3K4me2 are often also associated with H3K4me3 (Pinskaya and Morillon, 2009). H3K4me2 has been reported to specifically recruit Histone deacetylases at active promoters, thus confining histone acetylation towards the promoter region bound by H3K4me3 (Kim and Buratowski, 2009).
H3K4 trimethylation (H3K4me3) is found largely overlapping with H3K4me2, but

with higher spatial restriction towards the TSS (Barski et al., 2007). This histone modification has been found to correlate with active transcription (Barski et al., 2007; Heintzman et al., 2007). Activating Histone acetyltransferases were shown to have *in vitro* binding affinity towards H3K4me3 (Hung et al., 2009). Furthermore, H3K4me3 has been reported to facilitate the recruitment of the general transcription factor TFIID and the chromatin remodeller CHD1 (Flanagan et al., 2005; Vermeulen et al., 2007). However, while these findings suggest for a role of H3K4me3 in active transcription, its direct involvement in transcription initiation remains to be tested.

1.2.4 H3K4me3 in embryonic development

In embryonic stem cells, H3K4me3 was not only found at actively transcribed genes, but also at transcriptionally repressed promoters, where it overlap with the histone modification H3K27me3 (Bernstein et al., 2006). This specific chromatin state was termed 'bivalent', as it combines active and repressive chromatin modifications. Bivalent domains are preferentially found at genes important during cell differentiation (Bernstein et al., 2006). Further studies found that bivalently marked promoters become resolved into either repressive or active chromatin states during differentiation (Bernstein et al., 2006; Dhar et al., 2016). These cell culture experiments suggest that in pluripotent stem cells, genes are marked by H3K4me3 to allow robust activation, yet H3K27me3 restricts activation towards a later differentiation state. Indeed, further studies were able to identify bivalent promoters in developing embryos of multiple species, where bivalent chromatin modification is enriched in genes involved in developmental processes (Lindeman et al., 2011; Rugg-Gunn et al., 2010; Schertel et al., 2015; Vastenhouw et al., 2010). These results suggest that bivalent promoters are a conserved feature of development, and H3K4me3 might be essential for transcriptional activation of genes during embryonic development. However, functional validation of the role of H3K4me3 in transcription activation in developing organisms is scarce (Hödl and Basler, 2012).

1.3 Establishment and removal of H3K4me3

Histone lysine methylation is established by Set-domain containing proteins. The Set-domain is named after the three most-studied enzymes that contain this protein motif: **S**upressor-of-variegation 3-9, **E**nhancer of zeste, and **T**rithorax (Jenuwein et al., 1998).

The first enzyme that was found to have H3K4 methylation activity was the Set1 protein in yeast, where it is the only H3K4-specific histone methyltransferase (HMT) (Roguev et al., 2001). In Drosophila, two other H3K4-specific HMTs are known, Trithorax (Trx) and Trithorax-related (Trr) (Ingham, 1998; Mohan et al., 2011). These proteins were identified in Drosophila mutant screens, where their mutation induced homeotic transformations (Ingham, 1998; Lewis, 1978). In vertebrates, further diversification has led to the existence of at least six different H3K4-specific HMTs. These HMTs are classified in three subgroups based on their homology to the three Set1-homologs in *Drosophila*. They are involved in multi-component complexes, which are called COMPASS complexes (see subsection 1.3.4, reviewed in Shilatifard 2012).

1.3.1 Set1 homologs - Set1a and Set1b

In yeast, Set1 is the only H3K4-specific HMT (Roguev et al., 2001). Two copies of Set1-homologs are found in vertebrates: Set1a and Set1b (Lee and Skalnik, 2005; Lee et al., 2007). Little is known about the developmental roles of Set1 proteins. Setd1a knock-out in mice results in embryo lethality due to gastrulation failures in preimplantation stages (Bledau et al., 2014). Set1a was further shown to regulate Oct4 activity in in the inner cell mass of embryos (Fang et al., 2016b). In *Drosophila*, Set1 is thought to be the main H3K4-specific methyltransferase in development, as Set1 knock-out leads to a significant reduction of H3K4me2 and H3K4me3 levels (Hallson et al., 2012). Set1 complexes have been shown to be recruited to chromatin via active RNA-Polymerase II (Krogan et al., 2003; Ng et al., 2003), thus their function might be restricted to transcription-coupled H3K4me3 establishment.

1.3.2 Trithorax homologs - Mll1 and Mll2

In constrast to Set1, the Trithorax homologs Mll1 and Mll2 have been shown to bind chromatin independent of active transcription, and thereby could establish H3K4me3 prior to active transcription (Bach et al., 2009; Birke et al., 2002; Milne et al., 2002; Yokoyama and Cleary, 2008; Zeleznik-Le et al., 1994). Mll1 is the most-studied H3K4-specific HMT. Initially, *mll1* was identified as a gene involved in chromosome translocations that are causal for acute myeloid/lymphoid leukemia (MLL) (Ziemin-van der Poel et al., 1991). Mll1 has been shown to regulate hematopoiesis both in cell culture experiments as well as during embryonic development (Ernst et al., 2004; Mishra et al.,

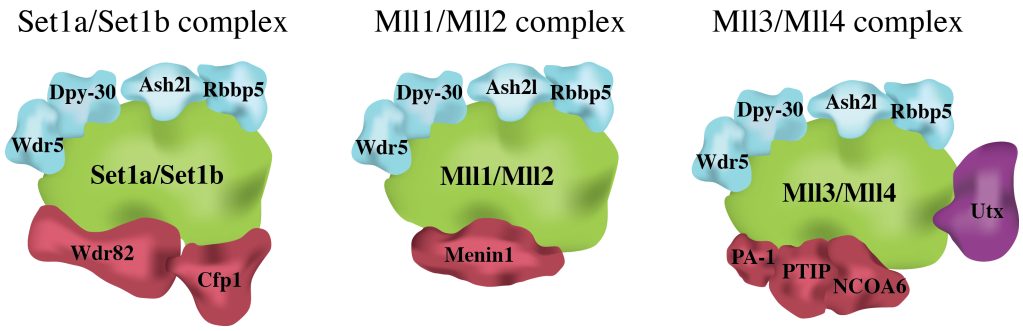


Figure 1.3: The COMPASS complexes of vertebrates. Three H3K4-specific HMT complexes (termed COMPASS) are found in vertebrates: Set1a/Set1b, Mll1/Mll2 and Mll3/Mll4 complexes. Each of the complexes contains only one HMT. The four light blue subunits (Ash2l, Wdr5, Rbbp5 and Dpy-30) are shared among all complexes and essential for full enzymatic activity. Subunits in red are complex-specific and confer target-gene specificity (Cfp1 and Wdr82 for Set1a/Set1b; Menin1 for Mll1/Mll2; PA1, PTIP and NCOA6 for Mll3/Mll4). Utx (violet) is specific to Mll3/Mll4 complex and is described as a H3K27-specific Lysine-demethylase.

2014; Yagi et al., 1998). Furthermore, genetic ablation of Mll1 in mouse has been shown to result in embryonic lethality (Yagi et al., 1998; Yu et al., 1995).
The second homolog of Trithorax, Mll2, has been found to be involved in solid tumor cell proliferation (Huntsman et al., 1999).Like Mll1, Mll2 mutation results in embryonic lethality (Glaser et al., 2006). Additionally, conditional knock-out of Mll2 in the germline leads to infertility due to interference with bulk H3K4me3 establishment (Andreu-Vieyra et al., 2010; Glaser et al., 2009).
Despite their importance in development, both Mll1 and Mll2 interference in cell culture experiments did not result in global changes of H3K4me3 levels, suggesting that both proteins are not the main H3K4-specific HMTs in differentiated cells (Guenther et al., 2005; Hu et al., 2013b; Lubitz et al., 2007; Terranova et al., 2006; Wang et al., 2009, 2012; Zhang et al., 2016). Furthermore, deletion of the enzymatic active Set-domain of Mll1 in mouse resulted in viable offspring, suggesting that the Set-domain is dispensible for embryonic development (Terranova et al., 2006). Cancer-inducing chromosome translocations did not contain the enzymatic active Set1-domain either, suggesting that its tumorigenic function cannot be explained by its role in H3K4me3 establishment (Hess, 2004; Mishra et al., 2014).
Mll1 and Mll2 share an enzymatic protein complex scaffold (Dou et al., 2006; Miller et al., 2001; Yokoyama et al., 2004) (Figure 1.3), therefore they might be redundant to each other in establishing H3K4me3. However, while loss of Mll1 seems to be

compensated for by the presence of other HMTs, Mll2 has been reported to play a unique role in establishing H3K4me3 in bivalent domains in development (Denissov et al., 2014). Thus Mll2 might be crucial for establishing chromatin signatures that are essential for embryonic development, and both Mll1 and Mll2 might be involved in establishing H3K4me3 at transcriptionally inactive promoters.

1.3.3 Homologs of Trithorax-related - Mll3 and Mll4

Mll3 and Mll4 are the vertebrate homologs of *Drospophila* Trr (Morgan and Shilatifard, 2013). Similar to Mll1 and Mll2, Mll3 and Mll4 have been reported to be involved in disease progression of multiple types of cancer (Kim et al., 2014; Li et al., 2013; O'Meara et al., 2014; Ortega-Molina et al., 2015; Wang et al., 2011; Watanabe et al., 2011). While Mll3 and Mll4 have H3K4 trimethylation activity *in vitro*, recent reports have shown that their function is restricted to H3K4 monomethylation in cell culture experiments, which is mainly found at enhancer chromatin (Herz et al., 2012; Hu et al., 2013a; Schuettengruber et al., 2009). Thus, HMT complexes containing Mll3 and Mll4 might be preferentially involved in chromatin regulation at enhancer elements (Lee et al., 2013a).

1.3.4 COMPASS complex proteins

H3K4-specific HMTs are involved in a multi-component complex, which is called COMPASS complex. Each individual COMPASS complex contains one of the HMTs described above.
Four COMPASS components have been shown to be essential for full enzymatic activity *in vivo*: Ash2l, Wdr5, Dpy-30 and Rbbp5 (Cao et al., 2010; Chen et al., 2012; Wysocka et al., 2005). These core complex subunits allow for substrate recognition and substrate presentation to the enzymatic pocket the COMPASS complex (Dou et al., 2006). Especially Ash2l has been shown to be essential for cell proliferation. In stem cells, Ash2l is involved in bulk H3K4me3 establishment (Wan et al., 2013). Furthermore, *ash2l* knock-out in mouse results in embryonic lethality (Stoller et al., 2010). Studies of Rbbp5, Wdr5 and Dpy-30 in embryonic stem cells show that knock-out of each of these components affects self renewal and maintenance of pluripotency (Ang et al., 2011; Jiang et al., 2011), supporting their essential role in differentiation processes.
Next to these core complex proteins, COMPASS complexes can contain unique pro-

tein components depending on their central histone methyltransferase. Menin1 is an essential component for complexes containing Mll1 or Mll2, and is involved in DNA sequence selectivity and protein complex assembly (Agarwal and Jothi, 2012; Matkar et al., 2013; Yokoyama et al., 2004). The same role is fulfilled by Wdr82 in COMPASS-complexes containing Set1a or Set1b (Lee and Skalnik, 2008). Furthermore, Set1-containing COMPASS complexes contain Cfp1, that has been shown to interact with hypomethylated CpG islands (Clouaire et al., 2012; Lee et al., 2007). COMPASS complexes with Mll3 and Mll4 contain the four proteins PTIP, Utx, PA1 and NCOA6, that regulate DNA recognition and modify the chromatin environment around COMPASS recruitment sites (Cho et al., 2007; Kim et al., 2014; Lee et al., 2008; Mansour et al., 2012). Together, these complex proteins allow for improved substrate recognition and DNA binding specificity of the different H3K4-specific HMTs.

1.3.5 H3K4me3 removal.

H3K4me3 is biochemically stable, and it was hypothesized that removal of H3K4 methylation relies on eviction and degradation of respective histone proteins. In subsequent *in vitro* studies, Lsd1 could be identified as an enzyme capable to remove methylation marks from histone tails without protein degradation (Shi et al., 2004). However, its enzymatic activity is restricted to removal of H3K4 mono- and di-methylation. Further studies identified the Jarid1/KDM5 protein family as H3K4-specific demethylases that are able to demethylate H3K4me3 as well (Iwase et al., 2007; Secombe and Eisenman, 2007; Seward et al., 2007). In vertebrates, four KDM5 proteins were identified. While all are able to demethylate H3K4, little is known about their isoform-specific regulation and locus-specificity. Their differential expression and regulation might present an additional mechanism for negative transcription regulation by H3K4me3 removal at promoter elements.

1.4 Transcription activation in embryos

Early embryonic development of almost all metazoans relies on maternally provided factors, while the zygotic transcriptome is transcriptionally inactive (Schier, 2007). Only after a species-specific period of development, the embryonic genome becomes permissive for transcription, and zygotic transcripts take over further embryonic development (Foe and Alberts, 1983; Kane and Kimmel, 1993; Oron and Ivanova, 2012;

Sulston et al., 1983; Wu and Gerhart, 1980). Enucleation experiments in sea urchins were the first to demonstrate that transcription from the zygotic genome is essential for further embryonic development (reviewed in (Laubichler and Davidson, 2008)). Furthermore, transcription inhibition in Xenopus eggs has shown that gastrulation relies on zygotic transcription (Newport and Kirschner, 1982). Thus, zygotic transcription activation is essential to allow for embryonic development. We use zebrafish as a model to study transcription regulation during zygotic genome activation.

1.4.1 Zebrafish early embryonic development

The zebrafish *Danio rerio* is a small fresh-water fish native to the tropical Himalayan region. It was established as a model organism for vertebrate genetics during the 1970s by George Streisinger and colleagues at the University of Oregon (Streisinger et al., 1981). Over the years it has been established as an outstanding model organism for developmental biology, for which numerous molecular biology and transgenic techniques are established (Grunwald and Eisen, 2002). Mating of zebrafish gives rise to hundreds of embryos, that develop synchronously *ex utero*. This makes zebrafish embryos amenable for manipulation and easy observation of developmental processes. Furthermore, embryos are transparent, which makes them an excellent experimental system for microscopy.

The early embryonic development in zebrafish is well characterized (Kane and Kimmel, 1993; Kimmel et al., 1995). After fertilization, the animal cap forms as a single cell by influx of cytoplasmic material from the underlying large yolk cell (Zygote period, 0-0.75 hours post fertilization, hpf). This initial stage is followed by rapid, synchronous cell divisions, which lead to an exponential increase of cell numbers while the cytoplasmic volume stays constant. After 10 cell divisions, the embryo enters the so-called midblastula transition (MBT). MBT is characterized by a slowdown and desynchronization of cell cycles, the beginning of cell motility, and the activation of the zygotic genome.

The early embryonic development occurs in the absence of zygotic transcription. Although a subset of genes is activated as early as 128-cell stage (Heyn et al., 2014), the major wave of zygotic genome activation occurs at the 1000cell stage, coinciding with MBT. Recent studies have demonstrated that specific transcription factors are essential for embryonic genome activation (Liang et al., 2008). In zebrafish, the transcription

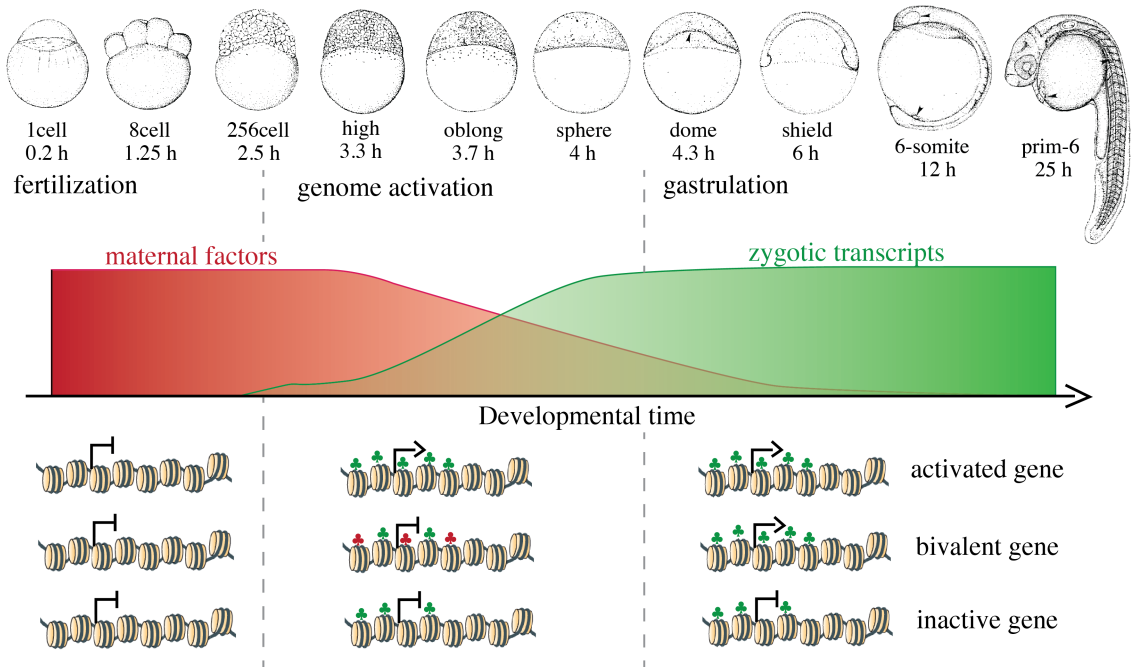


Figure 1.4: Zebrafish embryonic development. Zebrafish embryonic development starts from a fertilized egg that soon forms a single cell on top of a big yolk cell by influx of cytoplasmic material from the yolk. Until genome activation (at around 3 hours post fertilization), cells divide synchronously. Shortly after (stages high, oblong, sphere), zygotic transcription commences (green), and maternal factors (red) are gradually degraded. Genome activation is accompanied by establishment of H3K4me3 (green leaflets) and H3K27me3 (red leaflets). H3K4me3 is found on genes that are transcriptionally active, but also on inactive genes. Genes marked with both H3K4me3 and H3K27me3 (bivalent genes) are initially repressed but thought to become resolved to either active state (only H3K4me3) or repressed state (only H3K27me3) during cell differentiation that happens in the gastrulation period. Stages of embryonic development adapted from Kimmel et al. 1995.

factors Pou5f3 (Oct4 homolog), sox19b (sox2 homolog) and Nanog were found to be crucial for genome activation (Lee et al., 2013b; Leichsenring et al., 2013). However, these transcription factors are present during earlier stages of embryonic development as well. Thus it is still unclear which mechanisms keep the zygotic genome inactive during early stages, and which mechanisms allow for transcription factor binding and thus genome activation at a specific developmental stage. Unpublished observations in our laboratory suggest that histone content might be a general repressive mechanism that prevents genome activation in early stages. Histones might compete with transcription factors and other proteins for binding to DNA during early developmental stages

(Joseph et al., in preparation), thus their excess might prevent chromatin modification and transcription factor binding prior to ZGA.

1.4.2 H3K4me3 during early zebrafish development

Zygotic genome activation is accompanied by major changes in chromatin organization. These changes might be involved in genome activation, as previous experiments have shown that chromatin might be involved in transcription repression prior to ZGA (Harvey et al., 2013; Newport and Kirschner, 1982; Wiekowski et al., 1993). In zebrafish it has been shown that nucleosomes are equally distributed, yet unstructured, in earlier stages, and become well-positioned at specific promoters during ZGA (Zhang et al., 2014). Furthermore, multiple histone modifications were found to be absent from chromatin prior to genome activation, and are established coinciding with ZGA (Lindeman et al., 2011; Vastenhouw et al., 2010). In this context, H3K4me3 is of particular interest, as it might facilitate transcription activation during genome activation (Flanagan et al., 2005; Hung et al., 2009; Vermeulen et al., 2007). H3K4me3 has been shown to be established during ZGA, and was found on genes that are actively transcribed, but also genes that are transcriptionally inactive genes. So far it is not known, whether early H3K4me3 establishment is involved in transcription activation, and whether H3K4me3 is essential for zygotic genome activation and embryonic development.

1.5 Thesis aim

The aim of my thesis is to investigate the role of H3K4me3 during early embryonic development. I applied genetic ablation strategies as well as introduction of inhibitors to interfere with global H3K4me3 establishment. Upon H3K4me3 interference I aim reveal whether promoter-associated H3K4me3 establishment is essential for transcription activation in zebrafish embryos. Finally I address possible DNA sequence and chromatin that might allow for early H3K4me3 establishment at the majority of promoters during early embryonic development. Together, these experiments will shed light on the importance of H3K4me3 in transcriptional processes, and will increase our understanding of the interplay of chromatin modifications and transcription regulation in embryonic development.

2 Materials and Methods

2.1 Materials

Chemicals

Name	Supplier	Order number
NaCl	Merck Millipore	1064041000
Glycine	Merck Millipore	1042011000
NaOH Pellets	Merck Millipore	1064981000
Sodium dodecylsulfate (SDS)	Serva	20765
Tricaine / MS-222	Sigma-Aldrich	E10521-10G
UltraPure Agarose	Invitrogen	16500-500
Red Sea Salt	Read Sea Europe	R11070
Methylene Blue	AppliChem	A4084,0025
Blotting-Grade Blocker	Bio-Rad	170-6404
NaOCl	Roth	9062.3
Bovine Serum Albumine (BSA)	Sigma-Aldrich	A8022-50G
Tris Pufferan	Roth	48552
$CaCl_2$	Merck Millipore	102378500
KCl	Merck Millipore	104936500
$NaHCO_3$	Merck Millipore	1063291000
Glycerol bidistilled	VWR Chemicals	24388.295
Dithiothreitol (DTT)	Serva	2071004
Bromophenolblue	Chroma Gesellschaft Münster	4F-093
Ethylenediaminetetraacetic acid (EDTA)	Merck Millipore	1084181000

LiCl	Merck Millipore	1056791000
Deoxycholic acid (DOC)	Merck Millipore	264101-100GM
2-[4-(2-hydroxyethyl)-piperazin-1-yl]-ethanesulfonic acid (HEPES)	Roth	91053

Commercial Solutions and Reagents

Name	Supplier	Order number
cOmplete Protease Inhibitors	Roche Science	11697498001
AccuGENE PBS	Lonza	51225
Formaldehyde, 37% solution	Amresco	0493-200ML
NuPAGE Transfer Buffer (20x)	Thermo Scientific	NP00061
NuPAGE MOPS SDS Running Buffer (20x)	Thermo Scientific	NP0001
NuPAGE MES SDS Running Buffer (20x)	Thermo Scientific	NP0002
Restore WB Stripping Buffer	Thermo Scientific	21059
PageRuler Plus Prestained Protein Ladder	Thermo Scientific	26619
Lambda DNA EcoRI+HindIII Digest	New England Biolabs	N3012S
Rhodamine-Dextrane solution	Thermo Scientific	R1371
Phenol Red 0.5% solution	Sigma-Aldrich	P0290-100ML
dead end Morpholino	GeneTools LLC	*custom order*
N-terminal H3 Peptide, 30 mg	GenScript	*custom order*
Ambion NaAc, pH=5.5	Thermo Scientific	AM9740

β-Mercaptoethanol	Sigma-Aldrich	M6250-500ML
Methanol	VWR Chemicals	20847307
HCl	Fisher Scientific	J/4315/17
Tween20	Sigma-Aldrich	P1379-10ML
Triton X-100	Sigma-Aldrich	T9284-100ML
IGEPAL CA-630	Sigma-Aldrich	I3021-100ML

Consumables

Name	Supplier	Order number
1.5ml tubes	Sarstedt	72.690.001
2ml tubes	Sarstedt	72.695.500
1.5ml Low Retention tubes	Kisker Biotech	G017
ABgene 96well PCR plates	Thermo Scientific	AB0600
white 96well Multiwell 480 qPCR plate	Roche Scientific	0472969200.1
96well sealing foil	Thermo Scientific	121860
8well stripes	Axygen	321-11-051
0.1-10ul tips	Axygen	12541298
2-200ul tips	Axygen	12571288
100-1000ul tips	Axygen	12551308
1-200ul Gel tip round	Sigma-Aldrich	CLS4853-400EA
6well plates	Thermo Scientific	140675
24well plates	Thermo Scientific	142475
cell culture dishes	Greiner Bio-One	639102
Glass capillary w/ filament	World Precision Instruments	TW100F-4
Glass capillary w/o filament	World Precision Instruments	TW100-4
Glass pasteur pipettes	Sigma-Aldrich	Z310727-1PAK
LuerLok syringe 1 ml	BDScience	309628
Parafilm M	Sigma-Aldrich	P7793-1EA

Slide-A-Lyzer G2 Dialysis Cassettes	Thermo Scientific	87729
Dynabeads Protein G	Thermo Scientific	10005
NuPAGE Novex 12% Bis-Tris Protein Gels, 10-well	Thermo Scientific	NP0341BOX
NuPAGE Novex 4-12% Bis-Tris Protein Gels, 10-well	Thermo Scientific	NP0321BOX
Amersham Hybond P 0.45 PVDF membrane	GE LifeSciences	10600023
Whatman 3mm Blotting Paper	GE LifeSciences	3030-917

Enzymes

Name	Supplier	Order number
Pronase	Sigma-Aldrich	000000010165921001
Proteinase K	Sigma-Aldrich	000000003115887001
XhoI	New England Biolabs	R0146S
XbaI	New England Biolabs	R0145S
EcoRI	New England Biolabs	R0101T
SnaBI	New England Biolabs	R0130L
StuI	New England Biolabs	R0187S
PstI	New England Biolabs	R0140S
SacII	New England Biolabs	R0157S
BsrI	New England Biolabs	R0527S
Hin4L	New England Biolabs	*discontinued*
HincII	New England Biolabs	R0103S
TSP45I	New England Biolabs	R0583S
T4 DNA Ligase	New England Biolabs	M0202S
T7 Endonuclease I	New England Biolabs	M0302S
Taq DNA Polymerase	MPI-CBG	*internal*

Micrococcal Nuclease (MNase)	Sigma-Aldrich	N5386-50UN

Commercial Kits

Name	Supplier	Order number
Qubit dsDNA HS Assay	Thermo Scientific	Q32854
iScript cDNA Synthesis Kit	BIO-RAD	1708891
FastStart SYBR Green Master	Roche Science	04673484001
ABsolute qPCR SYBR Green Mix	Thermo Scientific	AB-1158
CloneJET PCR Cloning Kit	Thermo Scientific	K1231
mMESSAGE mMACHINE SP6 Transcription Kit	Thermo Scientific	AM1340
QIAprep Spin Miniprep Kit	Qiagen	27106
QIAquick PCR Purification Kit	Qiagen	28104
QIAquick Gel Extraction Kit	Qiagen	28704
MinElute PCR Purification Kit	Qiagen	28004
RNeasy Mini Kit	Qiagen	74106
Amersham ECL WB Detection Reagent	GE LifeSciences	RPN2106
Pfu DNA Polymerase Kit	Promega	M7741
Blot Developper RP X-Omat LO	Carestream	124 9259
Blot Developper X-Omat ExII	Carestream	113 5433

TruSeq ChIP Library Preparation Kit	Illumina	IP-202-1012
High Sensitivity DNA Kit	Agilent	5067-4626
RNA 6000 Nano Kit	Agilent	5067-1511
ProbeSet for Nanostring	Integrated DNA Technologies (IDT)	*custom order*
nCounter Elements TagSet-96	Nanostring Technologies	112008
Elements Master-Kit	Nanostring Technologies	110054

Laboratory Equipment

Name	Supplier	Order number
Benchtop centrifuges Mini Star	VWR Chemicals	521-2844
Vortex-Genie 2	Scientific Industries	SI-0236
Cooling centrifuge 5415R	Eppendorff	Z605212
Thermomixer C	Eppendorff	5382000015
Pipetman Classic set	Gilson	F167300
Bel-Art Pipette Pump	Fisher Scientific	13-683C
Accurpette	VWR Chemicals	612-4552
Qubit 2.0 Fluorometer	Thermo Scientific	Q32866
2100 BioAnalyzer	Agilent	G2939AA
Nanodrop 1000	Thermo Scientific	*discontinued*
LightCycler 96	Roche Science	05815916001
Mx3000P QPCR	Agilent	401512
Micropipette puller Model P-97	Sutter Instruments	P-97
Pneumatic PicoPump	World Precision Instruments	SYS-PV820
Microinjection stand	Narishige	MN-153
Stereomicroscope 2000	Carl Zeiss	*discontinued*
Ultrasound probe sonifier	Branson	W-450
DynaMag 2 Magnet	Thermo Scientific	12321D

XCell SureLock Mini-Cell	Thermo Scientific	EI0001
XCell II Blot Module	Thermo Scientific	EI9051
Odyssey SA Infrared Imaging	LI-COR	9260-11
MasterCycler Pro	Eppendorf	6321000515
Zebrafish incubator	Memmert	PP400
Photofilm Developer X-Omat 2000	Kodak/Carestream	*discontinued*
HiSeq2000 NextGen Sequencer	Illumina	*discontinued*

Plasmids

Insert	Vector	Purpose	Source
zebrafish H3.1 *zgc:173552*, C-terminal HA-tag	pCS2+	mRNA *in vitro* transcription (SP6 RNA-Polymerase + NotI restriction)	*this study, see chapter 6.6*
zebrafish H3.1 *zgc:173552*, K4-to-M substitution, C-terminal HA-tag	pCS2+	mRNA *in vitro* transcription (SP6 RNA-Polymerase + NotI restriction)	*this study, see chapter 6.6*
zebrafish H3.1 *zgc:173552*, K4-to-E substitution, C-terminal HA-tag	pCS2+	mRNA *in vitro* transcription (SP6 RNA-Polymerase + NotI restriction)	*this study, see chapter 6.6*
zebrafish eGFP + *nanos1*-3'UTR	pCS2+	mRNA *in vitro* transcription (SP6 RNA-Polymerase + NotI restriction)	Schier Lab, Harvard

zebrafish Kdm5Bb	pSport6.1	subcloning	Schier Lab, Harvard
zebrafish KDM5Bb, C-terminal Flag-Tag	pCS2+	mRNA *in vitro* transcription (SP6 RNA-Polymerase + NotI restriction)	*this study, see chapter 6.6*
mouse KDM5A/Jarid1A	pSp1	cell culture expression	Anastassiadis Lab, Dresden
mouse KDM5B/Jarid1B	pSp1	cell culture expression	Anastassiadis Lab, Dresden
mouse KDM5C/Jarid1C	pSp1	cell culture expression	Anastassiadis Lab, Dresden
mouse KDM5D/Jarid1D	pSp1	cell culture expression	Anastassiadis Lab, Dresden
mouse KDM5A/Jarid1A, C-terminal Flag-Tag	pCS2+	mRNA *in vitro* transcription (SP6 RNA-Polymerase + NotI restriction)	*this study, see chapter 6.6*
mouse KDM5B/Jarid1B, C-terminal Flag-Tag	pCS2+	mRNA *in vitro* transcription (SP6 RNA-Polymerase + KpnI restriction)	*this study, see chapter 6.6*
mouse KDM5C/Jarid1C, C-terminal Flag-Tag	pCS2+	mRNA *in vitro* transcription (SP6 RNA-Polymerase + NotI restriction)	*this study, see chapter 6.6*
mouse KDM5D/Jarid1D, C-terminal Flag-Tag	pCS2+	mRNA *in vitro* transcription (SP6 RNA-Polymerase + NotI restriction)	*this study, see chapter 6.6*

Antibodies

Epitope	Type	Working concentration	Supplier	Order number
Flag-Tag	mouse monoclonal (M2)	WB 1:500	Sigma-Aldrich	F1804
HA-Tag, ChIP-Grade	rabbit polyclonal	WB 1:5000	Abcam	ab9110
Ash2l	rabbit polyclonal	WB 1:100	Abcam	ab192906
Ash2l	rabbit polyclonal	WB 1:500	Abcam	ab84732
Mll1	rabbit affinity-purified	WB 1:500	Bethyl Labs	A300-375A-1
Mll/HRX	rabbit polyclonal	WB 1:500	Active Motif	61295
H2A.z, ChIP-Grade	rabbit polyclonal	WB 1:500, ChIP 10ug	Abcam	ab4174
pan-H3	rabbit polyclonal	WB 1:10000, ChIP 10ug	Abcam	ab1791
H3.1	rabbit polyclonal	WB 1:500	Abcam	ab183627
H3K4me1	rabbit polyclonal	WB 1:500	Abcam	ab8895
H3K4me2	rabbit polyclonal	WB 1:500	Millipore	07-030
H3K4me3	rabbit polyclonal	WB 1:500, ChIP 10ug	Millipore	07-473
H3K9me3	rabbit polyclonal	WB 1:500	Abcam	ab8898
Tubulin	mouse monoclonal	WB 1:10000	Sigma-Aldrich	T6074

HRP-coupled anti-mouse IgG	rabbit affinity-purified	WB 1:20000	Jackson ImmunoResearch	315-035-003
HRP-coupled anti-rabbit IgG	goat affinity-purified	WB 1:20000	Jackson ImmunoResearch	111-035-003
IRDye 800CW anti-rabbit IgG	donkey affinity-purified	WB 1:20000	LI-COR	925-32213
IRDye 680RD anti-mouse IgG	goat affinity-purified	WB 1:20000	LI-COR	925-68070

Primer Sequences

Name	Sequence	Annealing temperature
Jarid1A fw (KDM5A)	CCCCCATCGATGCCA CCATGGCGTCCGTGG GCCCG	67°C
Jarid1A-Flag rev (KDM5A)	GCTCTAGATTACTTG TCATCGTCATCCTTG TAATCTTTTTTATTC TTCTCTTCTTCCGTT ACAAATACAAC	59/81°C
Jarid1B fw (KDM5B)	CCCCGCTCGAGGCCA CCATGGAGCCGGCCA CCACG	64°C
Jarid1B-Flag rev (KDM5B)	GCTCTAGATTACTTG TCATCGTCATCCTTG TAATCCTTTCGGCTT GGTGCGTCCTTCC	64/80°C
Jarid1C fw (KDM5C)	CCCCCATCGATGCCA CCATGGAGCTGGGGT CCGAC	60°C

Jarid1C-Flag rev (KDM5C)	GCTCTAGATTACTTG TCATCGTCATCCTTG TAATCCAATTGTTGC AGAGATGGCTGC	59/81°C
Jarid1D fw (KDM5D)	GAAGGCCTGCCACCA TGAAGCCAGGATCTG ACG	55°C
Jarid1D-Flag rev (KDM5D)	GCGCCTACGTATTAC TTGTCATCGTCATCC TTGTAATCTAATTTT TGCTGATGAAAATAA GATAGGTG	55/78°C
KDM5Bb fw	CCCCCATCGATGCCA CCATGACCCAGCAGG GCCCG	64°C
KDM5Bb-Flag rev	GCTCTAGATTACTTG TCATCGTCATCCTTG TAATCCTTCCGGCGG TCCATGTGAC	62/81°C
H3.1 zgc:173552 fw (gene-specific)	CCATGCAGCGGCGTA TTCAG	56°C
H3.1 zgc:173552 rev (gene-specific)	GCCTTTGGGGTTTGA CGTGT	54°C
H3.1 fw	ATGGCCAGAACCAAG CAGACCGC	67°C
H3.1-HA rev	TTAAGCGTAGTCTGGC ACGTCGTAGGGGTAG CCTGATCCTGATCCA GCGCGCTCTCCGCGG	58/71°C
H3.1 K4-E fw	CTTGGCATGGCCCGT ACTGAGCAGACCGCG CGTAAATC	72°C

H3.1 K4-E rev	GATTTACGCGCGGTC TGCACAGTACGGGCC ATGCCAAG	72°C
H3.1 K4-M fw	CTTGGCATGGCCCGT ACTATGCAGACCGCG CGTAAATC	72°C
H3.1 K4-M rev	GATTTACGCGCGGTC TGCATAGTACGGGCC ATGCCAAG	72°C
ash2l F2	ATCATTTACTCACCC TCCACTTGTC	60°C
ash2l R2	AAAGCAAAATGAAAC TGATGTCCAG	59°C
mll TALEN F2	ACCCACACTGAGCTC CTGATCC	65°C
mll TALEN R2	GGTCCCATAGTAAGA AAAACAGCAC	63°C
mll 5-1 fw	TCAGCATTCTCACCT CTGTG	54°C
mll 5-4 rev	TTGTTGCTGATCTCT CAAAGC	55°C
mll2 TALEN seq-fw	TCCATTGCCTCATTT CAGATG	57°C
mll2 TALEN seq-rev	GTCCCATGCCTACTT AAAAGAAC	57°C
eif4g2a qfw	GAGATGTATGCCACT GATGAT	57°C
eif4g2a qrev	GCGCAGTAACATTCC TTTAG	57°C
ash2l qfw	CACCAGTACCAGCAG AATGAT	57°C
ash2l qrev	CAACCCATATCGCTC ATTGG	57°C

mll1 qfw	CTGTGGGTGTCTATA GATCAG	57°C
mll1 qrev	GAGTGGTTAATGAAA CGGGC	57°C
mll2 qfw	TCTGTACGCTGCTAA AGACC	57°C
mll2 qrev	GAATGATTGACATAA CGAGCAGG	57°C
efna1 qfw	TGTTGGACCTCACGG TTC	57°C
efna1 qrev	AGACCAGTGTCCATA AAGCC	57°C
skib qfw	TACTCCTCACAGATT GAGGATC	57°C
skib qrev	TCTTGATCTCCTTTG GTGGG	57°C
slc30 qfw	GAAGGCTGCCGATAT GTGT	57°C
slc30 qrev	CAATGTTGACCCGAA CAAGG	57°C
akap12b qfw	AGAATCCAGAGGTCC AGACT	57°C
akap12b qrev	TTACAGGCCGACTGC TC	57°C
rpl26 qfw	CTGCGGCAGAAATAC AACG	57°C
rpl26 qrev	CTGGTGATCACAACC TTGC	57°C
rpl26 upstream qfw	GGTTGTAGAGCGAAG ATGC	57°C
rpl26 upstream qrev	GAAGGTAAATCTGTT GCGTAAC	57°C
rpl26 intron qfw	AGTTCTCGAGTCTCTT TTCCG	57°C

rpl26 intron qrev	TGTGTGAGTGTTTTC ATGTCG	57°C

All Primers were obtained from Europrim-Invitrogen via Thermo Scientific in the minimal scale and quality possible for the respective oligo length.

Buffers

Zebrafish medium (for 20 liter)

Chemical	Amount	Comments
Red sea salt Methylene-blue, 10% solution fill up to 20 l with desalted water	5 g 20 ml	test pH for being between 7 and 8

TBS (Tris-buffered saline)

Chemical	Final concentration	Comments
Tris-chloride NaCl	50 mM 150 mM	adjust pH to 7.5 with HCl

TBS-Tween (Tris-buffered saline with 1/Tween20)

Chemical	Final concentration	Comments
Tris-chloride NaCl Tween20	50 mM 150 mM 1% v/v	adjust pH to 7.5 with HCl

Deyolking Buffer

Chemical	Final concentration	Comments
NaCl	55 mM	add 10% Glycerol
KCl	1.8 mM	for deyolking of stages
$NaHCO_3$	1.25 mM	prior to MZT

Deyolk Wash Buffer

Chemical	Final concentration	Comments
NaCl	110 mM	add 10% Glycerol
KCl	3.5 mM	for deyolking of stages
$CaCl_2$	2.7 mM	prior to MZT
Tris-Cl, pH 8.5	10 mM	

SDS Loading Buffer

Chemical	Final concentration	Comments
Tris-Cl	70 mM	
Glycerol	5% v/v	
SDS	0.7 mM	
DTT	1.5 mM	
Bromophenol blue	0.3 nM	

ChIP Cell Lysis Buffer

Chemical	Final concentration	Comments
Tris-Cl	10 mM	pH to 7.5,
NaCl	10 mM	add 1x Protease Inhibitors
NP-40/IGEPAL	0.5% v/v	fresh for each experiment

ChIP Nuclear Lysis Buffer

Chemical	Final concentration	Comments
Tris-Cl	50 mM	pH to 7.5,
EDTA	10 mM	add 1x Protease Inhibitors
SDS	1% v/v	fresh for each experiment

IP Dilution Buffer

Chemical	Final concentration	Comments
Tris-Cl NaCl SDS EDTA	16.7 mM 167 mM 0.01% v/v 1.2 mM	pH to 7.5, add 1x Protease Inhibitors fresh for each experiment

ChIP Washing Buffer

Chemical	Final concentration	Comments
HEPES EDTA DOC NP-40/IGEPAL LiCl	50 mM 1 mM 0.7% v/v 1% v/v 500 mM	pH to 7.6, add 1x Protease Inhibitors fresh for each experiment

ChIP Elution Buffer

Chemical	Final concentration	Comments
$NaHCO_3$ SDS	50 mM 1% v/v	pH must be above 10.5

MNase Digestion Buffer

Chemical	Final concentration	Comments
Tris-Cl $CaCl_2$ Triton X-100	50 mM 1 mM 0.2% v/v	pH to 7.5, add 1x Protease Inhibitors and PMSF fresh for each experiment

MNase Dilution Buffer

Chemical	Final concentration	Comments
Tris-Cl $CaCl_2$ NaCl Glycerol	50 mM 126 mM 10 mM 5% v/v	pH to 8

MNase Stop Buffer

Chemical	Final concentration	Comments
Tris-Cl	10 mM	pH to 7.6
EDTA	20 mM	

Chromatin IP Buffer

Chemical	Final concentration	Comments
Tris-Cl	27.8 mM	pH to 7.5
NaCl	111 mM	
SDS	0.34% v/v	
EDTA	4.1 mM	
Triton X-100	0.8% v/v	

2.2 Methods

2.2.1 Zebrafish husbandry and care

Zebrafish were maintained at 28°C according to standard procedures. For mating, male and female fish were separated overnight. After combining individual fish of opposite sex, embryos were collected after 15 minutes. If necessary for experimental procedures, embryos were dechorionated by brief Pronase treatment (~3 minutes) and subsequent washes with Zebrafish Medium (see Table 2.1). Embryos were left to develop to the desired stage at 28°C in agarose-coated culture dishes. Staging was performed by developmental time and according to morphological criteria based on Kimmel et al. 1995.
To assess the genotype of mutant zebrafish, adult fish were anesthetized by Tricaine treatment (0.0001% in Zebrafish medium) and fin biopsies were taken. Fins were lyzed in 100mM NaOH for 15 minutes at 95°C, and fin lysate was neutralized by addition of 10mM of 1M Tris (pH 7.5). After brief spin-down, 2ul of fin lysate were used for regular PCR using genotype-specific primers. The resulting PCR products were incubated with specific restriction enzymes (see below), and the resulting band pattern was analyzed by DNA gel electrophoresis.

2.2.2 Generation of zebrafish knock-out lines by TALEN mutagenesis

TALEN constructs were generously provided by the Schier lab at Harvard. mRNA synthesized from provided constructs were injected into one-cell stage embryos through the chorion, and embryos raised under standard conditions. To optimize TALEN construct concentration, different amounts of TALEN mRNA were injected, and mutagenesis activity was assessed by T7E1 restriction. In brief, DNA of multiple TALEN-injected embryos was extracted and PCR was performed with primers flanking the targeted mutation site. The obtained PCR product was denatured at 95°C for 10 minutes, and hybridized to wildtype PCR product by slow cool-down to 65°C on a PCR machine. After hybridization, T7E1 enzyme was added. This enzyme would cause DNA restriction if basepair mismatches between hybridized strands occur. The amount of degraded DNA was assessed by Agilent DNA High Sensitivity kit analysis according to manufacturers protocols.
The workflow to generate mutant zebrafish by TALEN expression is depicted in Fig-

Genotyping protocols

Genotype	Primers	PCR program	Enzyme	Expected bands
ash2l TALEN	ash2l F2 + ash2l R2	40 cycles; 95°C 20sec, 60°C 20sec, 72°C 30sec	PstI, 37°C, 2h	wt: 250 + 400bp; mutant: 640bp
mll TALEN	mll TALEN F2 + mll TALEN R2	40 cycles; 95°C 20sec, 65°C 20sec, 72°C 30sec	SacII, 37°C, 2h	wt: 470bp; mutant: 160 + 310bp
mll sa137	mll 5-1 fw + mll 5-4 rev	40 cycles; 95°C 20sec, 54°C 20sec, 72°C 30sec	BsrI, 65°C, 2h	wt: 250bp; mutant: 500bp
mll2 TALEN	mll2 TALEN seq-fw + mll2 TALEN seq-rev	40 cycles; 95°C 20sec, 57°C 20sec, 72°C 30sec	Hin4L, 37°C, 2h	wt: 200 + 300bp; mutant: 500bp

ure 3.1. Optimal TALEN construct concentrations were injected into one-cell stage embryos through the chorion. Subsequently, these fish were raised to adulthood. T7E1 restriction assay (described above) was performed on biopsies of the adult fin, to identify individual fish harboring mutated alleles. These fish were mated with wildtype fish, and the resulting offspring was used for an T7E1 restriction assay, to test for germline transmission of mutated alleles. Validated founder fish were mated with wildtype fish, and the resulting F1 generation was raised to adulthood. F1 fish were individually tested for mutant alleles by sequencing of PCR products spanning the mutation site. F1 fish that harbor validated protein-disruptive mutations were crossed against wildtype to obtain stable lines of a single mutated allele. These fish lines were crossed at least twice against wildtype fish before incrossing them to obtain homozygous mutants.

2.2.3 Generation of plasmids for mRNA production

Plasmids coding for mouse Jarid1/KDM5 constructs were generously provided by the Anastassiadis (Biotec, Dresden) and Neugebauer (MPI-CBG, Dresden) labs. The coding sequence from KDM5 constructs was amplified by PCR with Pfu-Polymerase. Construct-specific primers were designed that further encoded sequences for zebrafish KOZAK sequence and restriction enzyme sites. Furthermore, C-terminal Flag-tag sequence was encoded in reverse primers (Table 2.9). Specific PCR products were identified by DNA gel electrophoresis and purified by Qiagen Gel Extraction kit according to manufacturers protocols. Restriction digest of PCR products was performed for 3 hours at 37°C (KDM5A, KDM5C and zebrafish KDM5Bb: XbaI+ClaI; KDM5B: XbaI+XhoI; KDM5D: StuI+SnaBI) and products were purified by Qiagen PCR purification kit according to manufacturers protocols. The purified products were ligated with empty pCS2+ vector backbone by T4 DNA Ligase overnight at 16°C and cultured in Strata-Clone competent bacterial cells.

Zebrafish sequence for H3.1 was obtained by PCR for the gene locus zgc:173552 from genomic DNA, and integration of the PCR product into pJet vector using the CloneJET PCR Cloning kit according to manufacturers protocols. From this construct, H3.1 was amplified by Pfu-Polymerase PCR with primers further encoding a C-terminal HA tag. The resulting PCR product was used for restriction digest with XhoI and XbaI at 37°C for two hours, and inserted into empty pCS2+ vector-backbone by T4 ligation at 25°C for two hours.
The resulting wildtype histone construct was used for primer-guided mutation. For each substitution, two primers were designed that overlap the site of interest, and single basepair substitutions were used to change amino acid 4 (Lysine, AAG) to either Methionine (ATG) or Glutamate (GAG). With these primers, Pfu-Polymerase PCR was performed to amplify the complete vector. The resulting PCR product was digested with XhoI for two hours at 37°C and religated by T4 DNA ligase at 25°C for two hours to produce a circular vector.

2.2.4 Microinjection

mRNA was synthesized *in vitro* from respective linearized plasmids using the Ambion mMessage mMachine kit with respective RNA polymerase (mostly SP6 polymerase) according to manufacturers protocol. The resulting RNA was purified using Qiagen RNeasy Mini kit according to manufacturers protocols, and RNA quality was assessed by gel electrophoresis.
For injection, dechorionated one-cell stage embryos were lined up in agarose-molds to allow for embryo positioning. If not stated otherwise, 1ng of RNA (or 100ng peptide) were injected via filamented glass capillaries into the animal cap. Microinjections were performed at 20°C. After injections, embryos were staged and raised at standard conditions described in subsection 2.2.1.

2.2.5 Germline transplantation

Germline transplantation was performed according to published protocols (Ciruna et al., 2002). Embryos of the genotype of interest were used to extract germline progenitors (referred to as 'donors') that were transplanted into wildtype embryos with interfered germline development ('hosts').
In detail, an incross of heterozygous fish for the respective genotype was performed. One-cell stage embryos were dechorionated and injected with 50pg RNA coding for GFP with the 3'UTR of *nanos1* mRNA (subsection 2.2.3). This construct restricts RNA localization and thus GFP expression to germline progenitors. This labelling approach allows to follow integration of GFP-positive donor germline into GFP-negative host embryos. In parallel, wildtype embryos were dechorionated at the one-cell stage and injected with 1nl of 0.25mM *dead end* Morpholino. This Morpholino inhibits translation of *dead end* mRNA. Inhibition of *dead end* expression results in mislocalization of germline-specific mRNAs and consequently absence of germline progenitors, resulting in infertility of *dead end*-Morpholino-injected embryos.
Both host and donor embryos were raised until sphere stage, and transplantation was performed between sphere stage and 30% epiboly stage. For each donor, 5 host embryos were lined up in agarose molds with animal cap facing up. Glass capillaries without filament were cut to match cell size of zebrafish embryos. Care was taken to achieve blunt needles. From donor embryos, the cells around the margin towards the yolk were sucked into the needle by careful underpressure applied by 10ml sy-

ringe, avoiding yolk contamination. After cell uptake, each 20-30 cells were carefully transplanted to the middle of the animal cap of 5 host embryos. Immediately after transplantation, the respective donor embryo was transferred into an 8well-stripe for subsequent genotyping. Only host embryos with transplanted cells from homozygous mutant donors were further raised under standard conditions.
Successfully transplanted germline progenitor cells migrate to the correct position for developing germline in the wildtype host, which do not contain germline progenitors themselves due to *dead end* Morpholino interference. At 24-36 hours post fertilization, transplanted embryos were screened for the presence of GFP-positive germ cells at the expected position of the germline. These fish were further raised to adulthood. An outcross against wildtype fish was used to test for fertility and to validate the genotype of the transplanted germline.

***dead end* Morpholino:** 5'-GCTGGGCATCCATGTCTCCGACCAT-3'

2.2.6 Western Blot Assays

To generate protein lysates of zebrafish embryos, the animal cap has to be separated from the yolk cell, otherwise protein gels are overloaded with yolk material (Link et al., 2006). Embryos were dechorionated at one-cell stage (subsection 2.2.1) and raised to respective developmental stages. At each stage, 10 or more embryos were transferred into ice-cold Deyolking Buffer. Embryos were homogenized by pipetting, and 20 seconds agitation at 1000rpm. Cells were spun down by centrifugation at 300g for one minute, washed with Deyolk Washing Buffer, subjected to agitation and spun down as before. The resulting cell pellet was resuspended in SDS-Loading Buffer and inactivated by incubation at 95°C for 10 minutes. Sample material was collected by brief centrifugation, and stored at -80°C until use.
For standard electrophoresis, 4-12% gradient gels and 1xMOPS Running Buffer were used to probe for multiple proteins at the same time. For histone modification blotting, 12% gels and 1xMES Running buffer were used to achieve highest resolution at small molecular weight proteins. A sample equivalent to 10 embryos was loaded per lane onto NuPage Bis-Tris gels. Electrophoresis was performed at 150V for 70 minutes at room temperature. Gels were blotted onto PVDF membrane by semidry transfer with 1X Transfer buffer containing 15% methanol at 150mA for 3 hours. Membranes were dried on Whatman Blotting paper until further use.

Membranes were blocked in TBS-Tween with 5% dry milk powder for one hour at room temperature, washed three times for 5 minutes with TBS-Tween, and treated with the respective primary antibody for one hour on Parafilm sealing foil at room temperature. Membranes were washed again three times for 5 minutes with TBS-Tween, and then treated with respective secondary antibody for one hour at room temperature on Parafilm sealing foil.
For standard protein detection, membranes were washed three times for 5 minutes with TBS-Tween, and subjected to ECL detection reagents for three minutes at room temperature. After removal of excess detection reagents, membranes were sealed into foil. Photosensitive film was exposed to the membranes for different time intervals in the dark, and developed using the Kodak Photofilm Developer X-Omat 2000.
For quantitative protein detection, secondary antibodies coupled to fluorophors were used. After secondary antibody treatment, the membranes were washed three times for 5 minutes with TBS-Tween, and rinsed with desalted water. Membranes were dipped into Methanol to allow for quick drying. Dry membranes were scanned and quantified using the Li-Cor Odyssey Infrared imaging systems with the provided analysis software.
If multiple immunoblotting of one membrane was necessary, membranes were washed three times 5 minutes in TBS and subjected to antibody stripping using ReStore Stripping Buffer for 15 minutes at room temperature. Membrane was rinsed twice with desalted water and washed three times 5 minutes with TBS before treatment with respective primary and secondary antibodies as described above.

2.2.7 RNA extraction and quantification assays

qRT-PCR analyses

Total RNA was extracted from zebrafish embryos by homogenization of 25 embryos per sample in 350ul RLT buffer (from Qiagen RNeasy Mini Kit) complemented with 3.5ul β-mercapto-Ethanol. Subsequent RNA cleanup was performed using the RNeasy Mini Kit protocols according to manufacturers protocols. RNA quality was assessed by gel electrophoresis.
cDNA synthesis was performed on 500ng of RNA by using the iScript cDNA synthesis kit following the maufacturers protocol. The resulting cDNA was 10 times diluted before use in qRT-PCR experiments.
For qRT-PCR quantification of mRNA transcript levels, primers were generally de-

signed to span at least one exon-exon border within the 3'-end of the coding sequence. For qRT-PCR, 5ul of respective cDNA samples were mixed with each 0.625ul 10mM forward and reverse primer, and 6.25ul Roche FastGreen MasterMix. PCR was performed in white 96-well plates on the Roche LightCycler following a three-step amplification protocol with each 20 seconds denaturation (95°C), 57°C annealing and 72°C elongation for 45 cycles. Melting peaks were tested for the presence of a single specific PCR products before subsequent analyses.

Nanostring assays

To allow for a direct quantification of multiple lowly-expressed mRNAs in parallel, a fluorescent probe-based Nanostring assay was applied to zebrafish genome activation (Geiss et al., 2008). Probes were designed to capture mRNA molecules from 12 genes that are stably expressed during zebrafish embryogenesis ('Normalizers'), and 84 genes that are not present prior to genome activation, but actively transcribed after MZT (Aanes et al., 2011; Harvey et al., 2013; Lee et al., 2013b; Pauli et al., 2012). Custom oligo probes matching these genes were obtained from IDT, and were used in a custom RNA Nanostring assay.
Total RNA was extracted as described above (subsection 2.2.7), and quality was assessed by performing Agilent RNA 6000 Nano Kit electrophoresis. RNA samples with RNA Integrity Number above 8 (Agilent BioAnalyzer) were subsequently used for Nanostring analyses. 100 ng of respective total RNA samples were mixed with Probe Sets and hybridized according to manufacturers protocols. The RNA samples were loaded onto Nanostring cartridge via Nanostring consumables by an Nanostring pipetting robot. Analysis was performed by nCounter Digital Analyzer with maximum scanning fidelity. Assay-internal negative controls were used for background substraction, and counts were normalized by Nanostring nCounter software to internal positive controls as well as to zebrafish-specific normalizer genes. The resulting normalized counts were used for subsequent analyses.

2.2.8 Chromatin immunoprecipitation (ChIP)

Preparation of spike-in Lambda Phage DNA

Lambda Phage DNA was used as exogenous DNA added to immunoprecipitated DNA to allow for normalization of overamplification during sequencing (Figure 4.4). In the process of library preparation, DNA samples are subjected to size-selection, which would exclude non-fragmented DNA. To achieve Lambda DNA fragments comparable to fragment sizes obtained by immunoprecipitation protocols, Lambda DNA was subjected to restriction digest.
DNA was extracted from commercially available DNA ladders, where Lambda Phage DNA was fragmented by EcoRI+HindIII double-restriction. DNA purification was performed by Qiagen PCR purification kit according to manufacturers protocols. The purified DNA was subjected to HincII restriction for 3 hours at 37°C and subsequent Tsp45I restriction for 3 hours at 65°C. Reaction was inactivated by incubation for 10 minutes at 95°C and gel electrophoresis was performed. From a 1% agarose gel, a region matching 100-700bp DNA was excised and DNA was extracted by Qiagen Gel Extraction kit. Successful digest was validated by performing DNA quantification in fragment sizes of interest by Agilent DNA High Sensitivity kit analysis. For ChIP-Seq experiments, an amount of fragmented Lambda DNA representing 5% of chromatin sample was added to immunoprecipitated DNA.

Sonication-based ChIP-Seq

To monitor local enrichment of chromatin features, Chromatin Immunoprecipitation (ChIP) was performed. To provide enough material for immunoprecipitation, around 1000 embryos of stages up to oblong, and around 500 embryos for later stages were used. For conventional sonication ChIP, embryos of the respective stage and treatment were fixed in PBS containing 2% formaldehyde for 15 minutes at room temperature under regular agitation. Formaldehyde was quenched by adding 1M Glycine (1/8th of the sample volume) for 5 minutes. Fixed embryos were washed three times with PBS. Afterwards, embryos were homogenized by pipetting in 1ml ice-cold ChIP Cell Lysis Buffer containing Protease Inhibitors. While 15 minute incubation on ice, homogenization was repeated regularly by pipetting through a 200ul tip. Cells were collected by centrifugation at 1300g for 5 minutes at 4°C, and nuclear lysis was performed by resuspension in 650ul ChIP Nuclear Lysis Buffer containing Protease Inhibitors. After

incubation on ice for 10 minutes, lysis was stopped by addition of 1350ul of IP Dilution Buffer containing Protease Inhibitor. The sample was transferred into a 15ml tube and subjected to probe-based sonication for a total of 2:45 minutes at 30% amplitude with one-second pulses of sonication. After sonication, 160ul of 10% Triton X-100 was added to 2ml of chromatin, and the cell debris was precipitated by high-speed centrifugation of 13000g for 10 minutes at 4°C. 50ul of the resulting chromatin sample were used for Input DNA extraction. The chromatin sample was incubated overnight with respective antibody-coated magnetic beads. The target-bound beads were washed six times with ChIP Washing Buffer, and once with 1xTBS, while beads were collected via a magnetic stand. Afterwards, chromatin was eluted from beads by two subsequent incubation steps with each 150ul of Elution Buffer for 15 minutes at 25°C. Beads beads were collected via high-speed centrifugation at 13000g for 5 minutes. Input samples were filled up to 300ul with Elution Buffer, and 20ul of 5M NaCl was added to each input and eluate sample. Samples were reverse-crosslinked overnight at 65°C, and DNA was precipitated at -80°C for 2 hours by addition of 800ul 100% Ethanol. After centrifugation at 13000g for 10 minutes at 4°C, the resulting DNA pellet was resuspended in 100ul desalted water, and purified by using the Qiagen MinElute PCR purification kit according to manufacturers protocol.
DNA was eluted in 12ul TE buffer, from which 1ul was used to quantify DNA content using the Qubit DNA High Sensitivity kit, and 1ul was used to control for ChIP efficiency by standard qPCR of validated ChIP target regions.
For Next generation sequencing, the calculated amount of fragmented Lambda Phage DNA (see subsection 2.2.8) was added to the samples, and the whole 10ul were used for Library preparation and Sequencing according to manufacturers protocols at the Deep Sequencing Facility of the BioTec, Dresden. Sequencing was performed to a depth of 50 Mio reads of 75bp length, which is twice the amount of data that is recommended for locally enriched histone modifications (Jung et al., 2014).

MNase-ChIP-Seq

For nucleosome-resolution enrichment analysis of chromatin features, chromatin was digested by Micrococcal Nuclease (MNase) prior to immunoprecipitation. While the general procedures were similar to sonication-based ChIP experiments (subsection 2.2.8), the nuclear lysate preparation and MNase treatment differ. Embryos were fixed in PBS containing 1% formaldehyde for 5 minutes under constant agitation. Formaldehyde

was quenched by adding 1M Glycine (1/8th of the sample volume) for 5 minutes. Fixed embryos were washed three times with PBS. Afterwards, embryos were homogenized by pipetting in 1ml ice-cold ChIP Cell Lysis buffer containing Protease Inhibitors, and incubated for 5 minutes on ice with regular homogenization. Cells were collected by centrifugation at 1300g for 5 minutes at 4°C, washed once with ice-cold PBS and collected by centrifugation at 1300g for 5 minutes at 4°C. The resulting cell pellet was resuspended in MNase Digestion Buffer in a volume to match one embryo per microliter. Chromatin samples were divided into 200ul aliquots, and heated to 25°C for two minutes. 0.08U of MNase were added to each aliquot of chromatin, and digestion was performed for 8 minutes for high and oblong samples, 10 minutes for sphere samples, and 15 minutes for dome samples. MNase activity was stopped by adding 100ul of MNase Stop Buffer to each aliquot, and pooling aliquots into one tube. The sample was dialyzed against 250ml Chromatin IP Buffer for 4 hours at 4°C using Slide-A-Lyzer dialysis chambers. After dialysis, samples were spun at 13000g for 10 minutes at 4°C to separate cell debris, and supernatant was used for chromatin immunoprecipitation as described before (subsection 2.2.8).

2.3 Bioinformatics Analyses

2.3.1 Quality control, alignment and peak calling

Raw Sequencing reads were obtained from the Deep Sequencing Facility at the Biotec Dresden. Reads were demultiplexed and converted to FastQ format by Offline Linux Basecalling (OLB, version 1.9) and CASAVA (version 1.7) software packages provided by Illumina Inc. Sequencing quality was assessed for each sequencing experiment by using FastQC (version 0.11.1). Reads were mapped to the zebrafish genome Zv9/danrer7 (Ensembl release v79, March 2015) using bowtie (version 2.2) with local alignment (option –local) to allow for soft trimming of reads. Reads that were annotated to more than one position in the zebrafish genome ('multimappers') were removed by only retaining reads with a reported unique mapping. For visualization, a coverage track from the alignment was built using bedtools (version 2.25) and UCSC wigtobigwig (version 4). Lambda Phage library size was used to scale samples of different treatment to allow for direct comparison.
MACS2 (version 2.1) was used to identify enriched genomic regions from the generated bam files for each sample separately. Sequencing libraries for input chromatin samples were merged for all stages, to be used as consistent control background sample for peak calling. The following options were applied:
-g 1.59e9 (genome size); –bdg (report score values); -q 0.01 (q-value cutoff); –nomodel (no fragment size calculation); –extsize 160 (experimental fragment size).

2.3.2 Lambda normalization

Lambda DNA was added to immunoprecipitated materials to allow for correction of overamplification during sequencing. Each sample was mapped to Lambda Phage DNA separately, using bowtie2 as described above. The number of uniquely mapped reads was reported per sample ('lambda library size'). The lambda library size for two samples of the same stage were used to calculate the overrepresentation factor by dividing the two lambda library sizes. This factor was used to scale graphical representations (bigwig files), and was used during calculation of H3K4me3 enrichment in the following analyses.

2.3.3 Differential ChIP enrichment analysis

To identify H3K4me3-positive regions, peak clusters identified by MACS2 were further merged and filtered. In experiments comparing wildtype and mll-/- fish, peak clusters of all stages were merged by bedtools (version 2.25), to obtain maximal-inclusive peak regions. Bedtools and ChipPeakAnno (version 3.7.3) were used to further annotate peak clusters to the closest gene. The annotated peak clusters were filtered to retain the peaks within +/-5kb distance to annotated transcription start sites.
For experiments comparing H3HA and K4M overexpression, peak regions of all stages of H3HA injection were merged by bedtools (version2.25), and only regions retained that are supported by a peak in at least 3 of 5 samples (custom R script). These regions were annotated by bedtools and ChipPeakAnno, and further filtered to be within +/-5kb distance to annotated transcription start sites.
Reads falling within the obtained merged peak regions were counted for each stage. To analyze differential binding, a fold change calculation was applied based on Pataskar et al. 2016. To each count, a constant value of 8 was added, to avoid overrepresentation of low-enrichment regions. Counts were normalized by lambda library size for each sample, and fold change was calculated by division of normalized read counts for samples of different treatment.
To allow for filtering of regions with low enrichment, FPKM values for each peak cluster and sample were calculated by division of read counts with the respective library size and peak cluster width. Additionally, a p-value to the fold-change was calculated for each peak cluster based on Bayesian statistics published in Shao et al. 2012.

2.3.4 Data integration

Enrichment for H3K4me3 and H3K36me3 for dome stage, as well as promoter CpG density were obtained from Zhang et al. 2014. Promoter GC content was calculated using EMBOSS geecee for regions +/-2.5kb of annotated transcription start sites. For H2A.z enrichment, data obtained by a sonication-ChIP during the course of this study was used. Enrichment was calculated for regions +/-2.5kb of annotated TSS based on Zhang et al. 2014. Datasets were integrated to the results obtained in this study by matching of gene loci via ENSEMBL gene-IDs.

2.3.5 Gene classification

Genes were classified according to their transcription levels based on Pauli et al. 2012. Grouping was performed in the following order:

Not transcribed:	FPKM value for all stages <1
Zygotic late activated:	FPKM value <1 until dome stage earliest, and FPKM>2 for all subsequent stages
Zygotic early activated:	FPKM value <1 until 1k stage, FPKM>4 for any stage until dome, and FPKM>2 for all subsequent stages
Stable transcript levels:	FPKM>2 for all stages, and transcript level covariant of variation of <30% over all stages
Maternal late degraded:	FPKM>2 for all stages until dome, and decreasing FPKM for later stages
Maternal early degraded:	FPKM>2 for all stages until oblong, and decreasing FPKM for later stages
Maternal late activated:	FPKM>2 for all stages until dome, and at least 2fold increase at any stage later
Maternal early activated:	FPKM>2 for all stages until oblong, and at least 2fold increase at any stage later
Random expression:	Genes that do not match any of the criteria above. Either genes with strong transcript level fluctuations, or weak expression at random stages.

3 Results I: H3K4me3 interference by Histone methyltransferase mutation

To investigate the role of histone H3 lysine 4 trimethylation (H3K4me3) in embryonic transcription activation, I aimed to interfere with H3K4me3 establishment during early zebrafish development. My first approach to achieve H3K4me3 reduction was to genetically interfere with expression of proteins involved in H3K4me3 establishment. I generated mutations in genes of the histone methyltransferases (HMT) *mll1* and *mll2*, as as well as the essential HMT complex component *ash2l* by TALEN mutagenesis (section 3.1). I analyzed the generated mutants for resulting phenotypes (section 3.1) and possible effects on total H3K4me3 levels in zebrafish embryos (section 3.2). I further investigated whether *mll1* mutation results in locus-specific changes of H3K4me3 levels in zebrafish embryos (section 3.3). Finally I analyzed whether changes in H3K4me3 promoter levels resulted in a transcriptional change in *mll1* mutant embryos (section 3.4).

3.1 Generation and phenotypic description of histone methyltransferase mutants

3.1.1 HMT TALEN mutagenesis workflow

H3K4me3 is absent from early embryos, and is only established during MZT (Lindeman et al., 2011; Vastenhouw et al., 2010; Zhang et al., 2014). To allow for rapid H3K4me3 establishment at zygotic genome activation, it is likely that histone methyltransferases (HMTs) and their complex partners are maternally provided. Due to a genome duplication in fish (Postlethwait et al., 2000), orthologs of Mll3 and Mll4 as well as Set1b are found in two copies in the zebrafish genome. All components of H3K4-specific HMT complexes have been shown to be ubiquitously expressed throughout early embryonic development (Pauli et al., 2012; Sun et al., 2008). Among these, the Set1-orthologs *setd1ba, setd1bb* and the Trr-ortholog *mll3b* show weak transcription levels, while *setd1a* and *mll4b* show highest expression levels (Table 3.1).

The establishment of H3K4me3 coincides with transcription activation, and many pro-

Table 3.1: Expression of COMPASS components in zebrafish embryogenesis Expression levels were derived from Pauli et al. 2012. Displayed are RPKM values for transcripts of each gene over early stages until gastrulation (dome stage).

Gene	2-4cell stage	256cell stage	1K stage	oblong	dome
mll1	13.47	8.95	16.30	9.74	4.31
mll2	32.13	26.09	33.28	28.00	29.28
mll3a	7.03	9.76	16.73	11.16	7.47
mll3b	0.84	2.96	2.96	2.27	1.53
mll4a	115.07	53.17	56.57	36.48	13.06
mll4b	7.70	7.94	9.27	7.65	4.98
setd1a	82.07	55.88	75.25	53.06	35.87
setd1ba	3.68	1.97	3.99	2.53	2.70
setd1bb	0.00	0.01	0.01	0.01	0.01
wdr82	86.23	91.38	62.01	82.97	55.74
ash2l	253.46	95.52	72.22	69.71	32.02
wdr5	77. 46	138.95	108.76	141.89	100.40
dpy-30	377.01	168.99	84.32	168.87	108.40
rbbp5	184.55	77.95	57.24	50.48	15.36
menin1	6.46	2.64	2.44	2.43	0.76
paxip1	134.00	109.75	89.52	93.62	34.95
PA1	132.88	30.66	19.75	17.58	12.33

moter regions are reported to gain H3K4me independent of transcription (Vastenhouw et al., 2010). Therefore I hypothesized that H3K4-specific HMTs that are able to bind chromatin independent of transcription are necessary for bulk H3K4me3 establishment during early zebrafish embryogenesis. While Set1 proteins are thought to be the main H3K4-specific methyltransferase (Hallson et al., 2012), they have been shown to be recruited to chromatin via active RNA-Polymerase II (Krogan et al., 2003; Ng et al., 2003). Mll3 and Mll4 have been reported to be restricted to H3K4 mono-methylation *in vivo*. Therefore I decided to target Mll1 and Mll2. Additionally, I chose to target the essential COMPASS component Ash2l, which is shared by all H3K4-specific HMT complexes (Cao et al., 2010).

To generate HMT and HMT complex mutants, I targeted the first coding exon of respective genes for mutagenesis by transcription activator-like effector nuclease (TALEN) constructs (obtained from the Schier lab, Harvard). TALEN constructs were injected into one-cell stage embryos, and the resulting fish were screened for mutated alleles by

T7E1 restriction (see subsection 2.2.2 and workflow in Figure 3.1). To validate germline

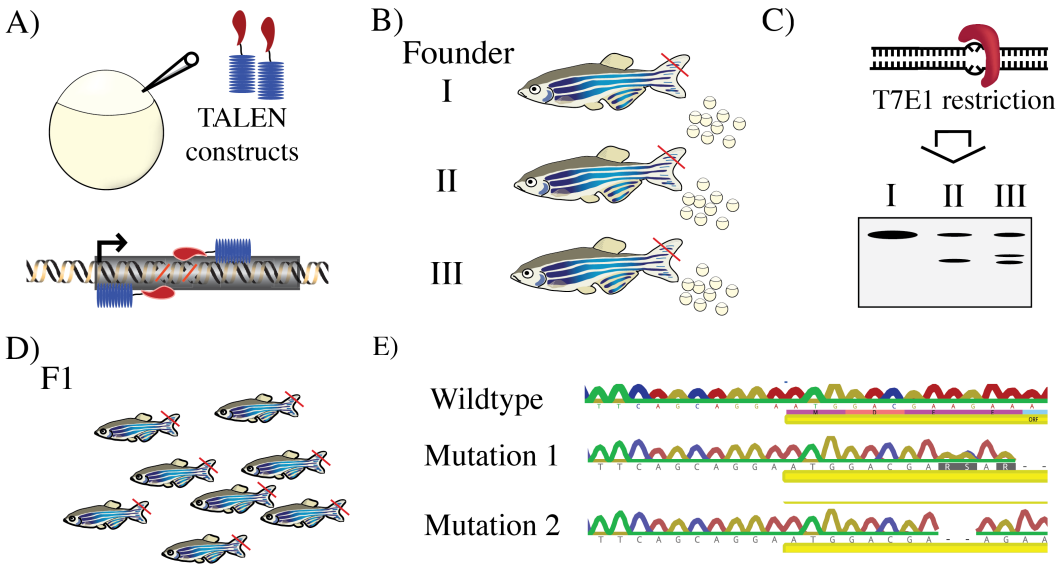


Figure 3.1: Description of TALEN mutagenesis A) TALEN constructs were designed to bind within the first coding exon of targeted genes. TALEN constructs were injected into one-cell stage zebrafish embryos. Upon expression, they induce DNA doublestrand breaks that are repaired by non-homologous end joining. B) TALEN-injected founder fish (F0) were raised to adulthood, and fin biopsies as well as embryo clutches taken for C) T7E1 mutation screen. PCR products spanning the putative mutation site were hybridized to wildtype PCR products, and subjected to T7E1 restriction. This enzyme only digests unpaired DNA, resulting in restricted DNA if a mutation was induced. D) F1 of mutant Founder fish was raised to adulthood, and fin biopsies were taken. E) Fin biopsies were screened for mutations that destroy the coding potential of the targeted genes by Sanger sequencing.

transmission, identified fish fish with mutant alleles were mated with wildtype fish, and the resulting offspring was screened for mutations by T7E1 restriction. Fish that carried mutations and that passed these alleles on to their offspring were kept as founders (F0). TALEN mutagenesis results in DNA double-strand breaks that are repaired by imprecise non-homologous end joining. Therefore, each cell within the developing embryo could harbor mutated alleles, which are different to other cells within the embryo. Consequently, multiple mutations can be introduced within a single organism, and founder fish could be mosaic for their mutated alleles. Indeed, I observed up to three different mutations in offspring from individual founder fish. To obtain a fish line with a

single mutation, the F1 generation of founder fish was raised to adulthood, and screened for frameshift-mutations by sequencing of PCR products that span the targeted genomic locus. For each identified mutation, at least two outcrosses against wildtype were performed before analysis, to reduce the chance of retaining off-target mutations that could possibly contribute to phenotypic changes. The identified mutations for genes of interest are described in the following paragraphs.

3.1.2 *Ash2l* TALEN mutation does not result in a larval or adult phenotype

Using TALEN mutagenesis, I was able to establish two separate lines with mutations in the *ash2l* gene (see Figure 3.2 A):
(1) an 11bp-deletion that deletes the Start codon of *ash2l* coding sequence (referred to as ATG deletion), and
(2) an 8bp deletion within the first exon, that results in a frameshift and a premature stop of the coding sequence within the first exon.
The start codon within the first exon of *ash2l* is essential to define the full-length protein-coding sequence of the mRNA. I therefore focused on the mutation that deletes the start codon for further analyses. Heterozygous *ash2l* mutant fish did not display any phenotypic differences. To determine whether homozygous *ash2l* mutants manifest a phenotype, I performed an incross of heterozygous fish. Until 5 days post fertilization (dpf), no phenotypic difference became apparent when comparing homozygous mutant embryos with their heterozygous siblings and wildtype fish. During further development, homozygous mutant offspring did not show phenotypic changes compared to wildtype fish. The adult fish were fertile and produced viable offspring.
Because zebrafish embryos rely on maternally provided mRNA and protein for their early development, the presence of maternally-provided RNAs and proteins could compensate for zygotic mutations during early embryonic development (Perrimon et al., 1984). To study early effects of mutations, it is therefore essential to generate embryos devoid of wildtype maternal contribution - maternal-zygotic mutants. This can be achieved by an incross of homozygous-mutant parents. *Ash2l* mRNA is maternally provided to zebrafish embryos (Table 3.1; Sun et al. 2008). However, maternal-zygotic *ash2l* mutants did not display an early embryonic phenotype, and did not show increased mortality until fertility age when compared to wildtype fish (see Figure 3.2 B). Thus, deletion of the start codon of zebrafish *ash2l* does not appear to affect zebrafish

development.

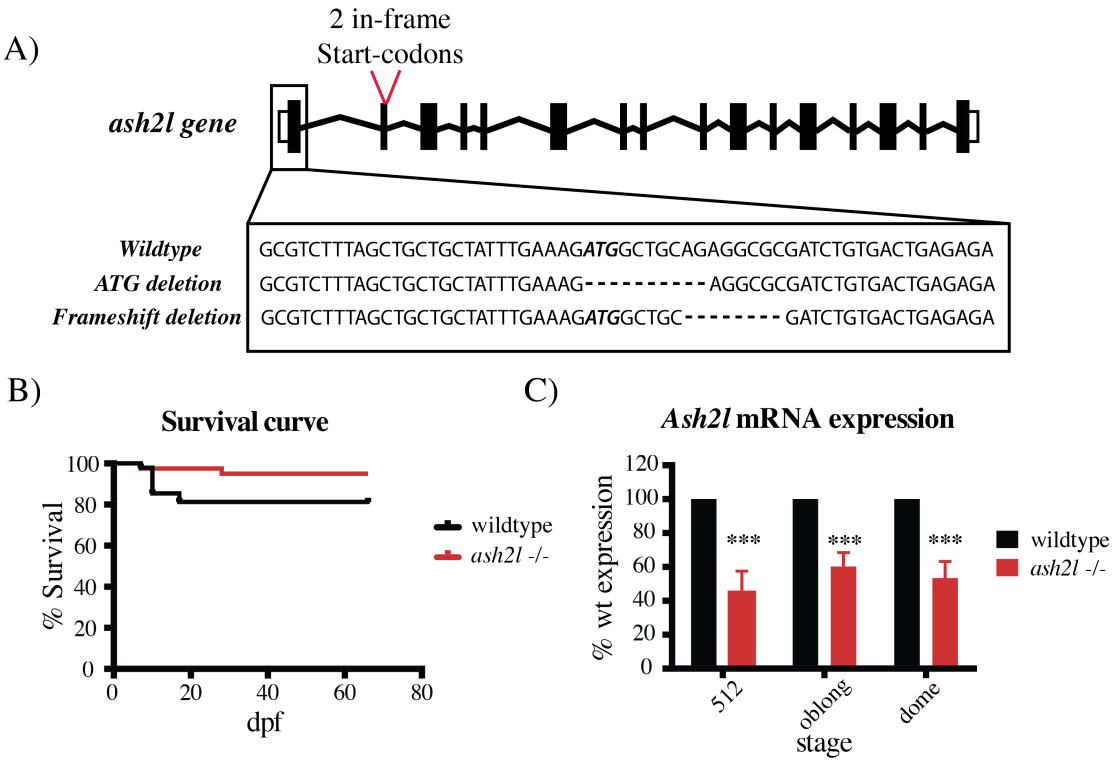


Figure 3.2: *Ash2l* TALEN mutations A) *Ash2l* gene structure, with mutations within the first exon generated by TALEN mutagenesis. Possible rescueing in-frame Start codons are indicated. B) Survival curve of wildtype versus *ash2l* maternal-zygotic mutants. n = 40. C) Relative expression of *ash2l* mRNA in *ash2l* mutants compared to wildtype embryos at stages around genome activation. n = 4, error bars = s.d., *** p<0.001

Ash2l is an essential component of all H3K4-specific HMT complexes (Roguev et al., 2001; Southall et al., 2009), and mutation of *ash2l* in mouse leads to embryonic lethality (Stoller et al., 2010). Therefore it is surprising that the here described mutation of *ash2l* does not have an influence on zebrafish development. One reason could be that the deletion of the start codon does not result in a complete loss of functional Ash2l protein in zebrafish embryos.

RNAs that lose their coding potential can be degraded by nonsense-mediated decay (Chang et al., 2007). To test whether the introduced mutation results in a loss of *ash2l* mRNA, I isolated RNA from maternal-zygotic *ash2l* mutant embryos at stages before (512 cell stage) and after genome activation (high and dome stage). I performed RT-

qPCR on the resulting cDNA with primers specific for *ash2l*. In mutants, I observed a reduction of *ash2l* mRNA to approximately 50% of wildtype RNA levels (see Figure 3.2 C). Thus, even though the deletion of the start codon was validated by sequencing of cDNA, *ash2l* mRNA is still produced from the mutated allele. It cannot be excluded that the remaining RNA retains coding potential. Indeed, *ash2l* has two in-frame start codons within the second exon, which encode a truncated Ash2l protein with all known functional domains. It is important to note that the second mutant line described above results in a truncated open reading frame that ends before these inframe start codons. Thus, this mutation will probably not lead to a different outcome than the described ATG deletion. While no increased larval mortality was observed for these frameshift mutants as well, further phenotypic tests remain to be performed.
Since mRNA levels are still present in *ash2l* mutants and could encode for a truncated protein, it would have been important to test for Ash2l protein levels in *ash2l* mutants. I tested two antibodies designed against mouse Ash2l C-terminal domains which has high homology to zebrafish Ash2l. However, neither of these antibodies could detect zebrafish Ash2l protein. Therefore it cannot be excluded that a truncated Ash2l protein from remaining mRNA functionally compensates the introduced start codon deletion within the *ash2l* gene. Since *ash2l* mutation is thought to be detrimental for embryonic development (LaJeunesse and Shearn, 1995; Stoller et al., 2010), it is likely that the here described mutations did not result in a complete loss of functional Ash2l protein. Since the *ash2l* gene is rich in in-frame ATG througout its gene body, a whole gene deletion approach would be a better strategy to achieve *ash2l* mutants. Recent advances in targeted deletion by homology-guided CRISPR-Cas9 in zebrafish might allow for this gene deletion approach (Hwang et al., 2013).

3.1.3 *Mll2* mutation results in increased larval mortality, while adult fish are healthy and fertile

By introduction of TALEN constructs targeting the first exon of *mll2*, I was able to generate three different gene-disruptive mutations. These mutations were deletions of different lengths, that all resulted in an early premature stop codon within the first exon of the coding sequence (Figure 3.3 A).
I focused further analysis on the 5bp frameshift mutation. To analyze potential phenotypes induced by mutating *mll2*, I performed an incross of heterozygous F2 fish. No phenotypic differences became apparent when comparing homozygous mutant embryos

with heterozygous siblings and wildtype embryos. However, in adult fish, homozygous mutants were underrepresented, even though there was no apparent adult phenotype. The fish were fertile and produced viable offspring.

To address whether *mll2* is important for embryonic development, I next analyzed

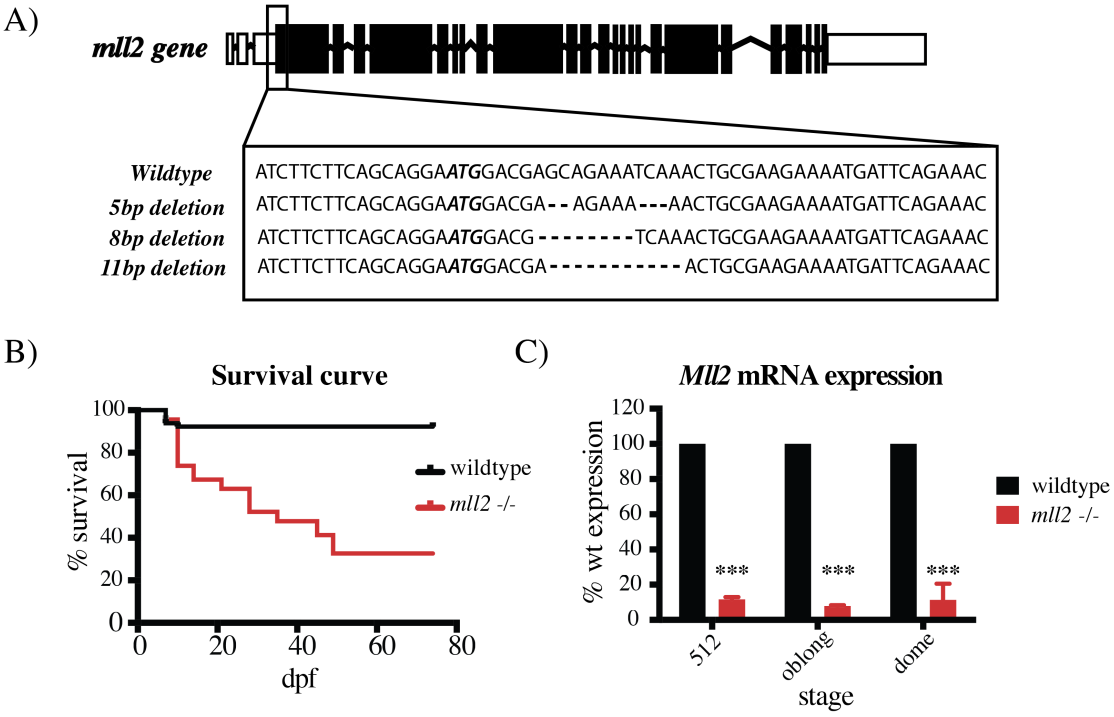


Figure 3.3: *Mll2* TALEN mutations A) *Mll2* gene structure, with mutations within the first coding exon generated by TALEN mutagenesis. B) Survival curve of wildtype versus *mll2* maternal-zygotic mutants. n = 65 C) Relative expression of *mll2* mRNA in *mll2* mutants compared to wildtype embryos at stages around genome activation. n = 4, error bars = s.d., *** $p<0.001$

maternal-zygotic *mll2* mutants. Mutant embryos did not show an early developmental phenotype nor increased mortality until 5 days post fertilization when compared to wildtype embryos. However, during further larval development, homozygous mutants showed an increased mortality rate. Only 30% of mutant fish survived until 3 months post fertilization (Figure 3.3 C). The surviving adult fish did not show significant morphological differences compared to wildtype fish, and were able to produce offspring. Thus, while *mll2* mutation apparently does not lead to defects during embryonic development, it seems to make larval devlopment less robust, leading to increased mortality in the absence of a morphological phenotype.

To test whether the introduced mutation has an effect on *mll2* mRNA levels, I performed

RT-qPCR experiments as described above (subsection 3.1.2). RNA levels of *mll2* were reduced to 10% of wildtype levels in *mll2* mutants, suggesting that mRNA is cleared due to nonsense-mediated decay (Figure 3.3 C). This suggests that similar to *mll2* RNA levels, Mll2 protein is depleted in *mll2* mutant embryos as well. Unfortunately, I was unable to validate the absence of Mll2 protein due to the lack of commercial antibodies that recognize zebrafish Mll2 protein.
In conclusion, the frameshift mutation within the first exon of *mll2* results in an almost complete loss of *mll2* mRNA in maternal-zygotic mutant embryos. Maternal-zygotic *mll2* mutants do not show an early embryonic phenotype, however increased mortality is observed during larval stages. Taken together, and in contrary to mouse knock-out studies (Glaser et al., 2009, 2006), *mll2* mutation does not lead to embryonic lethality and infertility. Instead, *mll2* mutant embryos develop normally, but increased mortality during larval development is observed, while escapers do not manifest an apparent adult phenotype and are able to produce fertile offspring.

3.1.4 *Mll1* mutation results in increased larval mortality and a severe adult phenotype

Using TALEN mutagenesis, I generated two different frameshift mutations for *mll1* within the first exon. These mutations lead to a premature stop codon within 60 base-pairs of the coding sequence (Figure 3.4 A). In addition, I obtained an *mll1* mutant (sa137) generated by the Wellcome Trust Sanger Institute. This mutation was generated by treatment of zebrafish embryos with *N*-ethyl-*N*-nitrosourea (ENU), followed by a sequencing-based screening for knock-out alleles of annotated protein-coding genes (Kettleborough et al., 2013). The obtained point mutation in *mll1* results in a premature stop (TGG to TGA) of the protein coding sequence in exon 10. Therefore, the resulting mRNA encodes a truncated Mll1 protein that contains DNA binding domains, but neither protein interaction domains that are necessary for complex formation, nor the enzymatic active C-terminal SET-domain (mutant sa137, Figure 3.4 A).
For further analysis, I focused on analyzing the Sanger mutation. To analyze a possible phenotype resulting from *mll1* mutation, I performed an incross of heterozygous mutants. Homozygous mutant fish were severely underrepresented in several incrosses of heterozygous mutants. The escapers displayed a severe phenotype, which manifests in an overall smaller size, small waist circumference and an atypical yellow color (Figure 3.5). While heterozygous mutant fish were viable and fertile, homozygous

mutants died before reaching fertility age. This suggests that *mll1* is essential for zebrafish larval development.

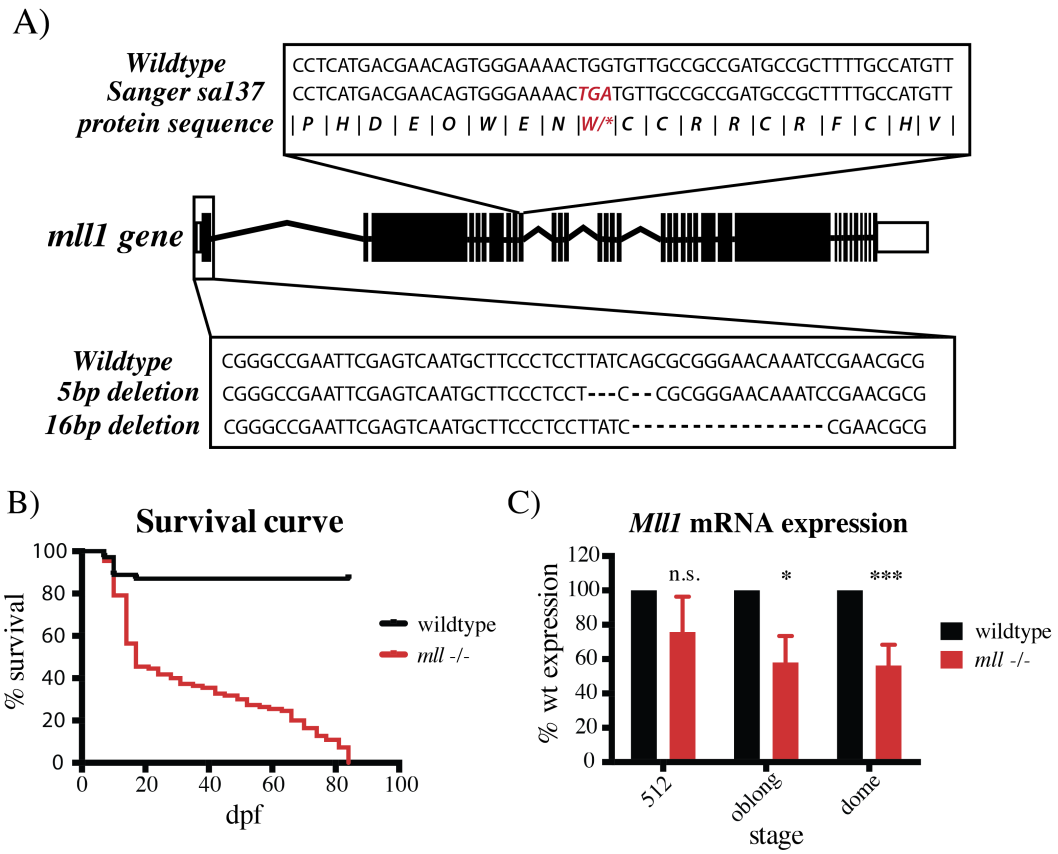


Figure 3.4: *Mll1* mutations A) *Mll* gene structure, with sequences generated by TALEN mutagenesis within the first coding exon (below gene), as well as Sanger point mutation within exon 10, leading to a premature stop codon (above gene). B) Survival curve of wildtype versus *mll1* maternal-zygotic mutant embryos. n = 110 C) Relative expression of *mll1* mRNA in *mll1* mutants compared to wildtype embryos at stages around genome activation. n = 4, error bars = s.d., * $p<0.05$, *** $p<0.001$

To analyze whether *mll1* is also involved in early embryonic development, I next set out to generate maternal-zygotic mutants. Since raising homozygous mutant fertile fish was not possible due to the severe adult phenotype, I performed germline transplantation (described in (Ciruna et al., 2002)). This technique allows for the introduction of homozygous-mutant germline progenitors into wildtype embryos where germline progenitor development was ablated (see subsection 2.2.5). In spite of germline transplantation carrying the *mll1* mutation, these fish developed without phenotypic abnormalities,

and reached fertility age. To test whether germline replacement was successful, these fish were incrossed and the resulting offspring screened for giving rise to homozygous mutant embryos. Using this technique, I was able to generate both germline-transplanted fertile males and females (16 and 3, respectively).
To validate that germline-transplanted females do not contribute wildtype mRNA coding for *mll1* to their offspring, I tested whether RNA in maternal-zygotic embryos contains the mutation. I extracted mRNA from maternal-zygotic mutants, and performed a PCR for *mll1* on the resulting cDNA. The obtained PCR products from stages before and after genome activation were sequenced, and confirmed that no wildtype RNA was present in these embryos.
Next I tested whether *mll1* mutation leads to clearance of the respective mRNA by nonsense-mediated decay. I performed RT-qPCR experiments as described above (subsection 3.1.2). While *mll1* RNA levels were not significantly changed before genome activation (512 cell stage), RNA was reduced to approximately 50% of wildtype levels in *mll1* mutants at stages post genome activation (Figure 3.4 C). This suggests that the mRNA might be partly subjected to nonsense-mediated decay during later stages of development.
The analyzed Sanger mutation results in a premature stop codon, that allows for the expression of a truncated N-terminal protein. However, while this protein still contains domains involved in DNA association, it lacks essential domains to form an enzymatic functional complex, and lacks the enzymatic active Set1-domain. In order to validate the absence of Mll1 protein with H3K4 methylation activity, I obtained two different antibodies that were raised against C-terminal domains of mouse Mll1 protein with high homology to zebrafish Mll1. However, both antibodies failed to detect a specific band at the expected height in zebrafish embryos. Therefore no validation of the absence of the protein could be performed. Thus it is possible that two truncated versions of Mll1 are made, one each for the N-terminus and the C-terminus. It cannot be excluded that functional complexes are assembled with truncated Mll1 proteins. Further efforts should focus on validating the absence of the Mll1 C-terminal protein in zebrafish mutant embryos, either by raising a zebrafish-specific antibody, or by targeted Mass-Spectrometry.

Next I set out to determine whether maternal-zygotic mutants show an early embryonic phenotype. I focused on analyzing morphological differences in embryonic development until bud closure. I was unable to observe any differences in embryonic

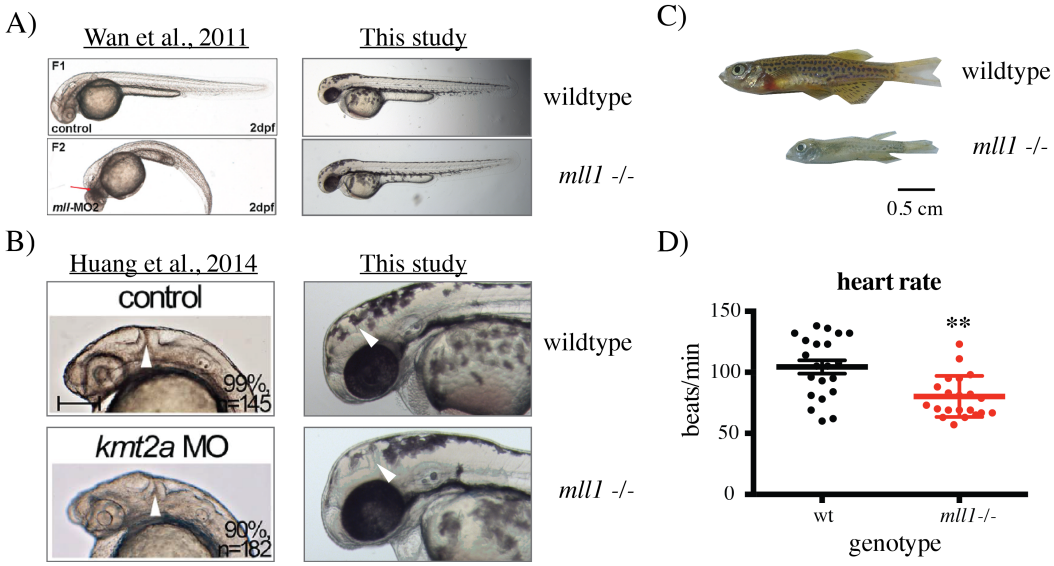


Figure 3.5: *Mll1* mutation results in a severe adult phenotype A) Zebrafish embryo body shape at 48hpf. Left figures from Wan et al. 2011, wildtype (top) and *mll1* Morpholino-treated (bottom); right figures from this study, comparing wildtype (top) and *mll1* maternal-zygotic mutants (bottom). n = 60. B) Brain development in 48 hpf embryos. Left figures from Huang et al. 2014, wildtype (top) and *mll1* Morpholino-treated (bottom); right figures from this study, comparing wildtype (top) and *mll1* maternal-zygotic mutants (bottom). White arrowheads indicate midbrain-hindbrain boundaries. n = 14. C) Adult zebrafish at 4 months age. Top, wildtype male; bottom, *mll1* maternal-zygotic mutant. D) Heart rate in anesthetized fish at 48 hpf. n = 12, error bars = s.d., p = 0.007

morphology over the first 12 hours of development (data not shown). Additionally, I assessed larval morphology at 24 hpf and 48 hpf. However, maternal-zygotic mutants were indistinguishable from wildtype mutants in general morphology, but also in embryonic brain structures (see Figure 3.5). *Mll1* mutants developped without any obvious morphological phenotype until 5 dpf. These results are contradictory to previous knockdown experiments using synthetic oligonucleotides (Morpholino). Interference with *mll1* expression by two different Morpholinos resulted in severe defects during early embryonic development. At 24 hpf and 48 hpf, zebrafish injected with *mll1* Moropholinos displayed an abnormal body shape with a kinked tail. Furthermore, phenotypic differences in brain folding (Huang et al., 2014) as well as a significant reduction in embryonic hematopoiesis were observed (Wan et al., 2011). However, recently it was reported that embryonic phenotypes described by Morpholino injections are seldom recapitulated in maternal-zygotic *mll1* mutants, suggesting that the previously described

phenotypes of *mll1* knock-down might have been the result from off-target effects (Kok et al., 2014).
Even though no apparent phenotype could be observed in early embryonic development, severe premature death was observed in maternal-zygotic mutants from 2 weeks of development onwards, with no mutant fish developing to fertility age (Figure 3.4).
Interestingly, the onset of increased mortality coincides with the time when zebrafish larvae become dependent on blood circulation to provide the body with oxygen. To understand whether defects in blood circulation could be the cause for increased larval mortality, I analyzed the heart rate of zebrafish embryos at 28 hpf, when blood circulation is fully established. I mounted embryos for imaging in Zebrafish medium containing Tricaine, a drug that blocks signal transduction in motor neurons and thus prevents larval movement. Interestingly, the heart rate of *mll1* mutants was reduced by 35% compared to wildtype fish of the same age (Figure 3.5 C). In rare occassions (1 out of 20 observed fish), the embryonic heartbeat became asynchronous, with sometimes no heartbeat for up to one minute before resuming to slower, synchronous heartbeat again. These observations indicate that maternal-zygotic *mll1* mutants might have a weakened heart that could result in heart failure and ultimately death during later larval development.
Together, these results show that *mll1* is dispensable for early embryonic development. However, *mll1* mutants die during larval development from 5 days post fertilization, and do not reach fertility age. While the molecular mechanism remains to be understood, initial observations suggest that *mll1* mutants display heart failures from 1 day post fertilization onwards.

3.2 HMT mutations do not affect global H3K4me3 levels in early zebrafish embryos

H3K4me3 is absent from early zebrafish embryos, and is established during genome activation (Lindeman et al., 2011; Vastenhouw et al., 2010; Zhang et al., 2014). To robustly establish H3K4me3, zebrafish embryos thus rely on the presence of HMT complex proteins. Therefore I hypothesized that *ash2l* knock-out results in strong reduction of H3K4me3 levels, as was shown in mouse embryonic stem cells (Wan et al., 2013). Furthermore, since H3K4me3 is also found at genes that are not transcriptionally active (Vastenhouw et al., 2010), I hypothesized that Mll1 and Mll2 are involved in establishment of H3K4me3 during early embryonic development.

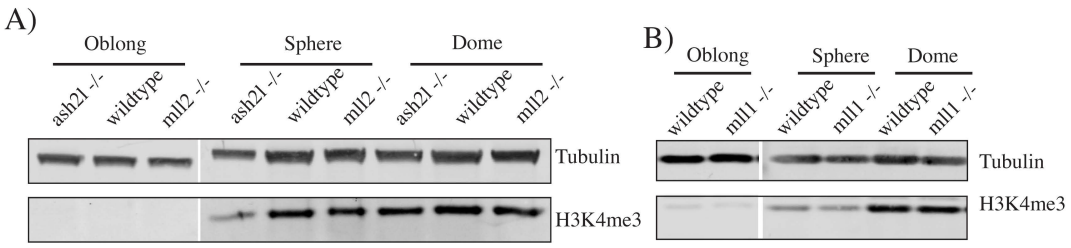


Figure 3.6: HMT mutation does not result in reduction of global H3K4me3 levels A) A representative Western Blot of oblong, sphere and dome stages comparing wildtype H3K4me3 levels to levels in *ash2l* and *mll2* mutant embryos. Tubulin serves as loading control. *ash2l* sphere stage shows reduced protein content by Tubulin blotting. B) As in A) comparing *mll1* maternal-zygotic mutants to wildtype embryos.

To screen for global H3K4me3 reduction in HMT mutants, I performed Western Blot analyses at different stages after genome activation. I chose to test H3K4me3 at oblong stage, the earliest stage after genome activation when H3K4me3 can be detected, and the following two developmental stages, sphere and dome. H3K4me3 levels in maternal-zygotic mutants at oblong stage are barely detectable, and do not allow for a conclusion about H3K4me3 reduction. At sphere and dome stages, H3K4me3 levels did not show significant differences for any of the mutations when compared to H3K4me3 levels in wildtype embryos (representative Western Blot see Figure 3.6).
The unchanged H3K4me3 levels in *ash2l* mutants could be explained by a possible truncated protein produced in the here used TALEN mutants (subsection 3.1.2). However, the reduction of RNA levels and the abrogated coding potential of *mll* and *mll2* mutant mRNA suggest against a functional rescue from remaining RNA in these mutants. One possible explanation is that other H3K4-specific HMT are functionally redundant to the mutated genes. Mll1 and Mll2 share the same complex components, and thus could compensate for their mutual loss (Dou et al., 2006; Miller et al., 2001; Yokoyama et al., 2004). Indeed it has been shown that loss of Mll1 in mouse cell culture experiments results in changes of less than 5% of all genes (Wang et al., 2009; Yu et al., 1998). Furthermore, a recent study found that loss of Mll1 is compensated by the presence of other HMTs, and thus is dispensible for H3K4me3 establishment in mouse embryonic stem cells (Denissov et al., 2014). However, the same study has shown that Mll2 has a unique role in establishing H3K4me3 at bivalent promoters. Thus, even though Mll2 and Mll1 seem not to be involved in global H3K4me3 establishment in zebrafish, they might be uniquely responsible for local H3K4me3 establishment at a subset of genes.

To identify local changes of H3K4me3 against global unchanged H3K4me3 levels, Western Blotting is not sensitive enough. Thus, the question of whether loss of HMT results in local changes of H3K4me3 levels in a distinct set of genes has to be analyzed by other experimental approaches.

3.3 *Mll1* mutation results in local H3K4me3 reduction of a small subset of genes

To identify specific promoters whose H3K4me3 levels may be affected by *mll1* mutation, I performed Chromatin immunoprecipitation followed by Deep Sequencing (ChIP-Seq). In ChIP-Seq experiments, proteins were covalently crosslinked to DNA. The crosslinked material was subjected for Micrococcal nuclease (MNase) digestion, which restricts accessible DNA and results in chromatin fragmentation to nucleosomal size. Next, immunoporecipitation with antibodies specific for H3K4me3 allowed for the enrichment of nucleosomal regions marked with H3K4me3. The crosslinked DNA was purified and subjected to Next-generation sequencing, which allowed for the identification of genomic regions enriched for H3K4me3. The resulting DNA was sequenced to a depth of 50 million reads of 75 bp length, which is twice the recommended sequencing depth for this histone modification (Jung et al., 2014). These experiments were performed in parallel for wildtype and maternal-zygotic *mll1* mutant embryos for two developmental stages:
(1) Oblong stage, when less than 10% of zebrafish genes are actively transcribed, but H3K4me3 is already detectable, and
(2) Dome stage, when H3K4me3 is present on more than 60% of all genes (Zhang et al., 2014).
Because active transcription can result in local H3K4me3 establishment (see subsection 1.3.1, Miller et al. 2001; Ng et al. 2003), it is important to analyze oblong stage. At this stage, most genes are not transcribed yet, thus transcription would contribute to H3K4me3 establishment at a small set of genes. Analyzing dome stage allows for the analysis of the dynamics of H3K4me3 establishment during embryonic development. Furthermore, the two different stages serve as biological replicates to identify regions specifically affected by *mll1* mutation. To account for the quantitative differences in cell number - and thus target DNA for immunoprecipitation - between these stages, I performed ChIP on 1200 embryos at oblong stage, and 400 embryos at dome stage.
H3K4me3 has been shown to be enriched at promoter regions, but depleted from most

enhancers, where H3K4me1 is found (Barski et al., 2007; Heintzman et al., 2009). To analyze whether the H3K4me3 is enriched at promoters in zebrafish development, I plotted the density of obtained sequencing reads over annotated gene bodies. As expected, a strong enrichment of H3K4me3 at transcription start sites (TSS) was observed, both for oblong and dome stage (Figure 3.7 A and B). H3K4me3 enrichment was generally higher at dome stage. Next I only considered the region of +/-3kb around TSSs. Interestingly, read enrichment was slightly higher in *mll1* mutants compared to wildtype samples (Figure 3.7 C and D). It can be excluded that this is a bias in general higher read number in *mll1* mutant samples, as read density was normalized to uniquely mappable reads. Furthermore, in both stages H3K4me3 levels are higher in *mll1* mutants, suggesting that it is a reproducible observation. However the difference in H3K4me3 enrichment is rather small. The density plot does not allow to speculate whether this increase is due to more H3K4me3 at all loci, or whether a gain in otherwise H3K4me3-negative TSSs accounts for the observed difference.

I further tested for H3K4me3 enrichment at enhancer elements. I extracted zebrafish-specific enhancer regions from a recent study that used chromatin modifications for *de novo* annotation of putative enhancers (Bogdanovic et al., 2012). However,while a slight enrichment of H3K4me3 could be found, these levels are marginal when compared to average enrichment at transcription start sites (Figure 3.7 E and F). Thus I conclude that H3K4me3 is preferentially enriched at promoter elements, as described before (Barski et al., 2007; Zhang et al., 2014), and *mll1* mutation leads to slightly higher H3K4me3 levels at promoter elements.

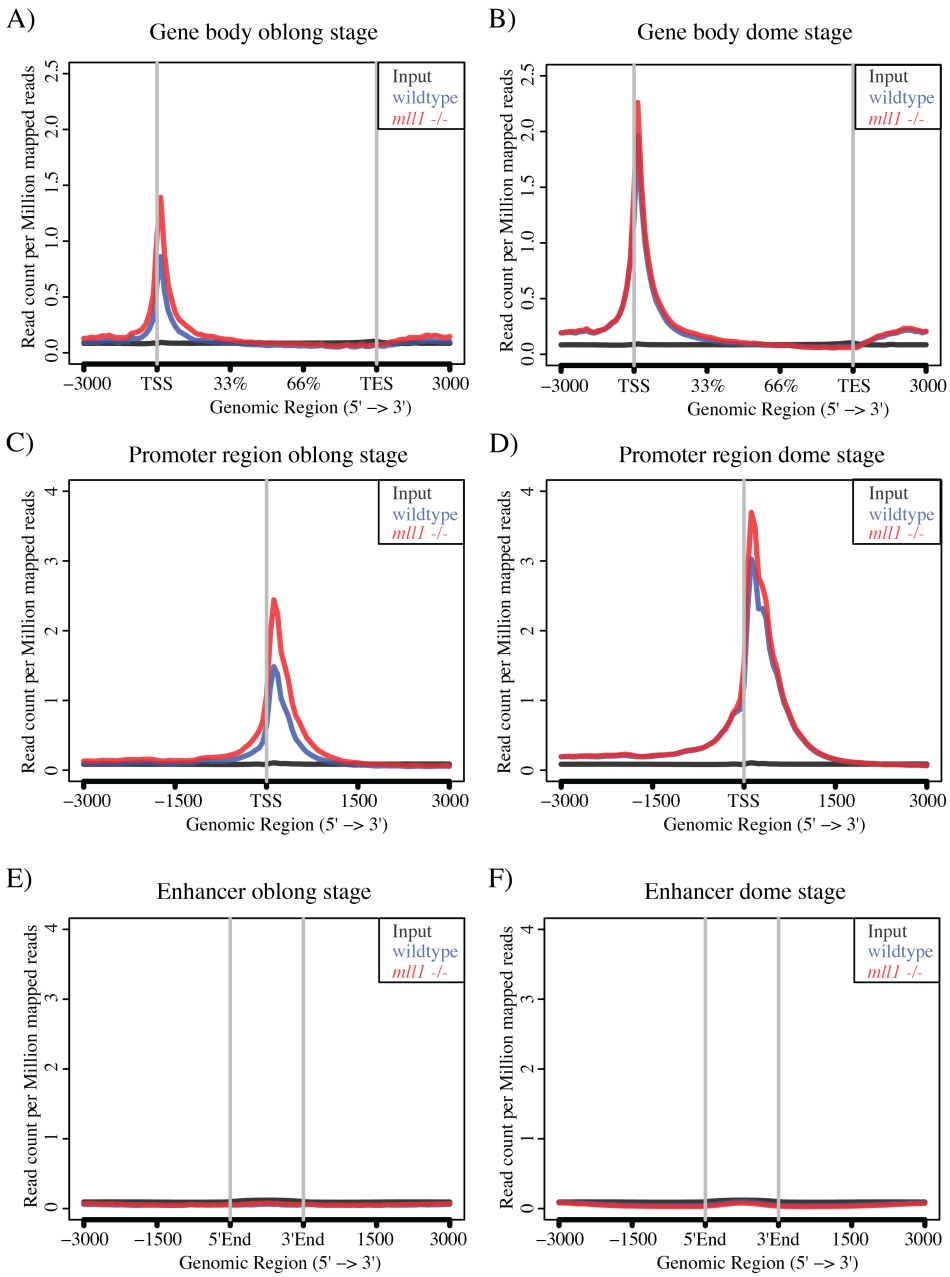


Figure 3.7: H3K4me3 is enriched at promoters in wildtype and *mll1* mutant embryos Metagene profiles displaying enrichment of ChIP-Seq reads for input chromatin (black), wildtype (blue) and *mll1* mutant (red) H3K4me3 ChIP-Seq samples over different genomic locations. A,B) H3K4me3 enrichment over gene bodies of annotated zebrafish genes at oblong and dome stage, respectively. C,D) H3K4me3 enrichment at TSS +/- 3kb at oblong and dome stage, respectively. E,F) H3K4me3 enrichment over enhancer regions (Bogdanovic et al., 2012) at oblong and dome stage, respectively. Plot scaled to TSS enrichment.

Identification of H3K4me3-reduced loci in mll1 mutant fish

Next I analyzed whether H3K4me3 levels at specific promoter regions are affected by *mll1* mutation. To determine quantitative differences in H3K4me3 enrichment, I tested previously established bioinformatics tools (reviewed in Steinhauser et al. 2016). However, none of these tools were applicable to the here obtained data. First, most of the tools are designed for transcription factor ChIP, thus they rely on locally more defined binding profiles than achieved histone modification ChIP data. Second, many tools rely on the presence of replicates to estimate biological variation between similar data. While I included two different stages in my experimental design, their general H3K4me3 enrichment varies dramatically, since H3Kme3 becomes increasingly established. Thus, samples of different stages cannot be treated as replicates, correct estimation of biological variation cannot be performed.
Instead of using published tools, I implemented a custom fold change analysis, that uses predefined regions of interest to identify differentially bound regions in each stage individually.
I excluded reads mapped to multiple positions from further analyses, as these cannot be confidently included into locus-specific quantification of H3K4me3 enrichment. The remaining reads were used to identify enriched regions for H3K4me3 (called 'peaks' from now on) by using the peak calling tool MACS2. Next I merged overlapping peaks from all obtained samples, and filtered for regions in +/-5kb distance to annotated TSS. I then counted sequencing reads falling into identified H3K4me3-positive regions, and normalized these counts by the sample-specific read library size (the number of all reads mapping to the zebrafish genome). The fold change ratio was calculated between wild-type and maternal-zygotic *mll1* mutant samples from the same stage at each genomic region. In total, 20,625 H3K4me3-positive regions were compared between wildtype and *mll1* mutants.

At oblong stage, 10% of those regions showed increased H3K4me3 levels, while 1.8% were decreased in H3K4me3 levels (Figure 3.8 A). At dome stage, only 0.98% showed increased H3K4me3 levels, while only 0.54% have decreased H3K4me3 levels (Figure 3.8 B). Since I hypothesized that *mll1* mutation leads to reproducible changes at specific loci, I consider regions as true target, if one region is significantly reduced (or increased) in H3K4me3 levels at both developmental stages. However, only 8 out of 20265 analyzed promoter regions show reduced H3K4me3 levels at both stages (Figure 3.8 C). In contrast, and opposite to my expectations, 40 of the changed re-

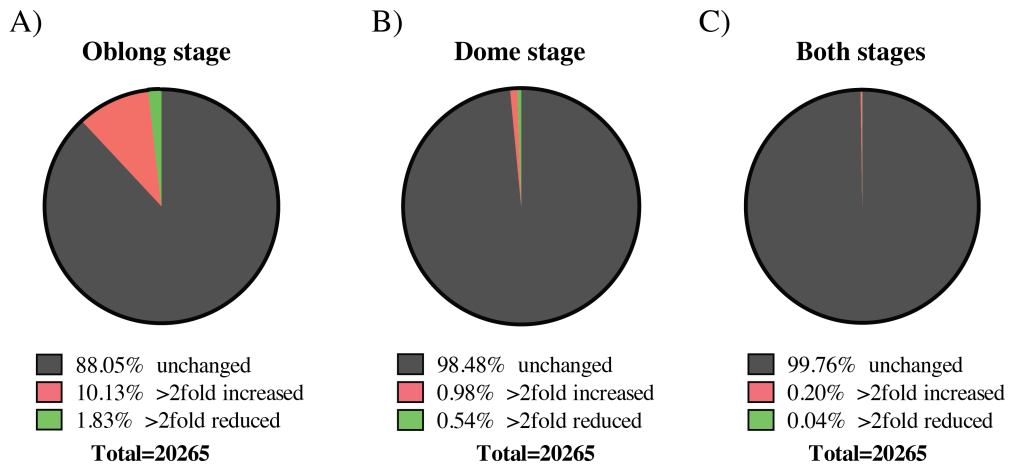


Figure 3.8: Fraction of promoters with changes in H3K4me3 levels Depicted are the fractions of H3K4me3-enriched regions that are unchanged (grey), more than 2fold increased (red) or more than 2fold decreased (green) in *mll1* mutants when compared to wildtype samples. H3K4me3-positive regions within 5kb of annotated TSS were considered. A) Oblong stage, B) Dome stage and C) Fractions of regions that show the same trend in both stages.

gions in both stages show higher H3K4me3 peaks at promoter regions in *mll1* mutants. These results indicate, that H3K4me3 levels are rather increased than decreased in *mll1* mutants, however most of H3K4me3 regions do not change significantly in maternal-zygotic *mll1* mutant embryos.

In conclusion, ChIP-Seq analysis revealed that only a small subset of zebrafish promoter regions displayed a change in H3K4me3 levels (12% at oblong and 1.5% at dome). Interestingly, H3K4me3 levels seem to be rather increased than decreased, both genome-wide (Figure 3.7) as well as on specific promoters (Figure 3.8). It is possible that the loss of Mll1 allows for more efficient recruitment of other H3K4-specific HMTs to these targeted loci. Alternatively it is possible, that shared HMT complex proteins are free to be recruited to other H3K4-specific HMT complexes, and thus stimulate their H3K4me3 activity. It is further possible that functions of Mll1 are redundant to other HMTs (as described in section 3.2).

While the original aim - to significantly reduce H3K4me3 levels at many promoters - was not achieved, the presence of a subset of promoters significantly changed in H3K4me3 levels allowed for further investigation of the role of H3K4me3 in their transcription activation.

3.4 Early embryonic transcription is not altered in *mll1* maternal-zygotic mutants

Mll1 mutation did not lead to a reduction of H3K4me3 levels at many genes in early embryos. However, I was able to identify changes in H3K4me3 levels at individual promoters. To understand whether these changes in H3K4me3 levels result in changes in transcriptional activity, I analyzed the onset of transcription of chosen target genes.

H3K4me3-affected genes do not alter their transcription levels

Of eight genes that had significantly reduced H3K4me3 levels in *mll1* mutant embryos, only *efna1b* is actively transcribed during early embryogenesis. Four of the other genes were not transcribed until 24 hpf, while three others were maternally provided and show decreasing transcript levels in early zebrafish embryos. To increase the number of genes that could be studied, I included two genes that showed significantly increased H3K4me3 levels at promoters in maternal-zygotic *mll1* mutants. These should show the opposite trend in transcriptional output compared to *efna1b*, thus might show higher transcription levels in *mll1* mutants. I extracted RNA at 6 stages post genome activation, when the respective genes become transcribed. I performed RT-qPCR on the resulting cDNA with primers targeting the last coding exons of the chosen target genes. Expression levels were normalized to *eif4g2a*, a gene that shows stable expression throughout zebrafish embryonic development. When comparing the relative expression levels for each of the genes, no significant difference in mRNA abundance could be seen for any of the three genes at all stages (Figure 3.9). Thus, changes in H3K4me3 levels seem not to influence transcription activation of respective genes in *mll1* mutants.

Genes activated during maternal-zygotic transition are not affected by mll1 mutation

Mll1 mutation resulted in changes of H3K4me3 levels in a small subset of genes (Figure 3.3). These changse in H3K4me3 levels did not result in transcriptional changes of the respective genes (section 3.4). However, I observed larval mortality and a severe adult phenoptype in escapers (subsection 3.1.4). These results suggest that the larval phenotype of *mll1* mutants might not be caused by the loss of H3K4me3 establishment,

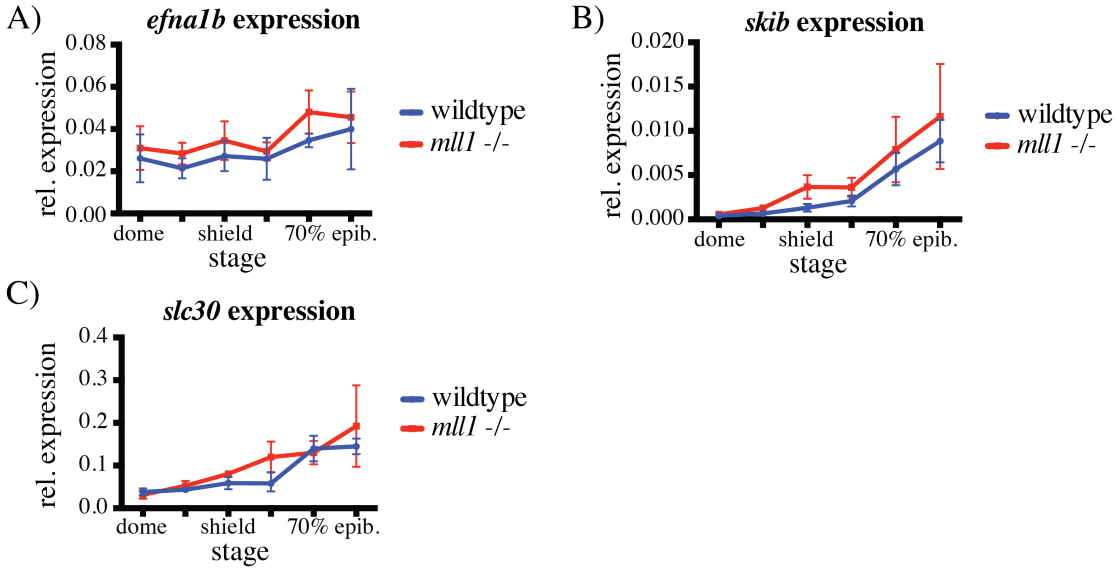


Figure 3.9: Change of H3K4me3 levels in *mll1* mutants does not affect transcription Relative expression levels of efna1b (A)(reduced H3K4me3 in *mll1* mutants), skib (B) and slc30a1a (C) (both increased H3K4me3 in *mll1* mutants) to normalizer gene eif4g2a from dome to 70% epiboly. n = 3, error bars = s.d.

but by other molecular mechanisms. Several studies in mouse have shown that deletion of the enzymatic functional Set1-domain of Mll1 does not recapitulate phenotypes of full *mll1* knock-out (Jude et al., 2007; Mishra et al., 2014; Terranova et al., 2006).

To address whether *mll1* mutation has an influence on embryonic transcription activation, I applied an RNA-quantification assay from NanoString Technologies. This technology allows for direct quantification of multiple specific RNA molecules in parallel, while not being disturbed by the abundance of other RNAs that are not of interest for the respective assay. It is based on immobilization of target RNAs by synthetic oligonucleotide probes, and subsequent fluorescent labelling of individual RNAs by specific fluorescent probes (Geiss et al., 2008).

To quantify transcriptional changes during zebrafish development, I focused on genes that are not maternally provided, but are activated early during zebrafish development (termed 'zygotic genes'). This allows for the detection of subtle changes in transcription activation that is solely influenced by processes in the developing embryo. Together with Mate Palfy from our lab, I chose 84 zygotic genes by integrating transcription data from four different studies (Aanes et al., 2011; Harvey et al., 2013; Lee et al., 2013b; Pauli et al., 2012). To control for differences in sample RNA quantity, we further included 12 'normalizer genes' that show stable transcription throughout early

embryonic development until gastrulation stages. We included genes with different stable transcript levels to account for possible biases in transcription analysis based on RNA abundance.

To test whether *mll1* mutation has an influence on early transcription activation, I per-

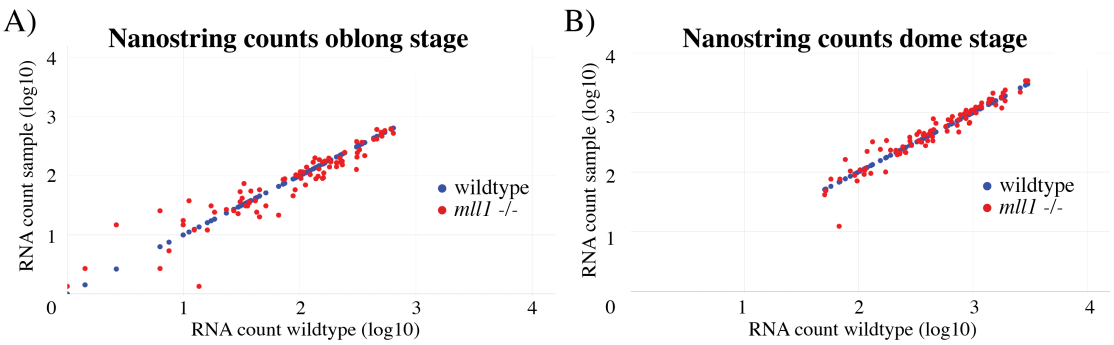


Figure 3.10: Transcription of early activated genes is not changed upon *mll1* mutation Depicted are scatterplots comparing mRNA counts of 84 zygotic genes in wildtype (x-axis) and *mll1* mutant (y-axis) samples. (A) Oblong and (B) dome stage. RNA quantification was performed by custom NanoString assays.

formed the described NanoString analyses comparing RNA samples from wildtype and *mll1* mutant embryos on six subsequent stages at zygotic genome activation (512cell to dome stage). Correct staging between mutants and wildtype was controlled by cell number counting at each stage by whole-embryo fluorescent microscopy.

RNA counts obtained by NanoString analysis were normalized over internal standards and the normalizer genes. Counts for wildtype and *mll1* mutants were compared at all stages. As expected, RNA counts gradually increased from 512cell stage to dome stage for all zygotic genes, confirming that the chosen target genes are indeed transcriptionally activated during early embryonic development.

To identify transcriptional differences, I plotted normalized count values for wildtype samples against *mll1* mutant samples in a scatterplot with logarithmic scale. Examples for oblong and dome stage are shown in Figure 3.10, which are the stages that were chosen for H3K4me3 ChIP-Seq experiments. No significant difference could be observed for any of the analyzed stages. While at individual stages, a few genes seem to be affected, these changes were not reproducible over other developmental stages, suggesting that *mll1* mutation has no influence on zygotic genome activation. Further detailed, genome-wide transcription analysis are necessary to identify, whether specific

target genes are affected by *mll1* mutation.

3.5 Conclusion

Concluding this chapter, I established maternal-zygotic mutant zebrafish for *ash2l, mll2* and *mll1*. *Ash2l* mutation did not result in phenotypic differences compared to wildtype fish, possibly due to the expression of an N-terminal truncated, yet functional protein from the remaining mRNA. Both *mll1* and *mll2* mutation resulted in increased larval mortality, while early embryonic development occured without phenotypic dfferences, suggesting for their importance in zebrafish larval development. However, none of the mutations led to global reduction of H3K4me3 levels. Thus, the here generated mutants do not allow to assess the influence of H3K4me3 on transcription activation during early embryonic development. I therefore decided to reduce H3K4me3 levels by introduction of inhibitors of H3K4me3 establishment in early zebrafish embryos.

4 Results II: H3K4me3 interference by introduction of HMT inhibitors

Genetic ablation of the histone methyl transferases *mll1*, *mll2* and the complex component *ash2l* did not result in global changes of H3K4me3 levels, and in *mll1* mutants only few loci with changed H3K4me3 levels could be observed.
Because these genetic interference approaches did not allow me to analyze the role of H3K4me3 in zebrafish embryonic genome activation, I set out to identify a method that inhibits genome-wide establishment of H3K4me3 in zebrafish embryos.
I established a Western Blot assay to identify global changes in H3K4me3 levels (section 4.1) and used it to test the efficiency of different H3K4me3 interference approaches (section 4.2, section 4.3, section 4.4). Introduction of H3K4-methylation-defective K4M histones was identified as a method to drastically reduce global H3K4me3 levels (subsection 4.4.2). I analyzed the resulting H3K4me3 reduction genome-wide by comparative ChIP-Seq experiments (section 4.5). Finally I analyzed whether local changes in H3K4me3 promoter levels have an effect on transcription activation during early zebrafish development (section 4.6).

4.1 Establishing a Western Blot assay to monitor H3K4me3 reduction

In order to achieve global H3K4me3 reduction, I applied different strategies. These follow a similar scheme: (I) Interfering constructs were injected into dechorionated one-cell stage embryos, (II) controlling for equal developmental progression, and (III) assessing the effect of interfering constructs at stages shortly after genome activation (Figure 4.1). To test for the effectiveness of H3K4me3 interference, I established a Western Blot assay. H3K4me3 is established during genome activation, and levels rapidly increase during embryonic development. As shown in Figure 4.1, H3K4me3 levels significantly change even within 15 minutes of developmental progression. Therefore, differences in H3K4me3 levels between samples might erroneously be due to misstaging rather than due to experimental treatment. Thus, precise staging is very important to identify successful H3K4me3 reduction. However, small staging differences between

treatment cannot be excluded. Thus, I developed a strategy to stage-match samples of different treatments during analysis. I probed for the histone modification H3K9me3 in parallel to H3K4me3. H3K9me3 is established during zebrafish development with similar dynamics as H3K4me3, but should not be affected by interference with H3K4-specific methylation establishment (Figure 4.1). To probe for H3K4me3 reduction, 20 embryos of each sample were deyolked and subjected to Western Blotting on two separate denaturing protein gels at selected developmental stages. The resulting blots were probed for H3K4me3 and H3K9me3, respectively. H3K9me3 levels were used to identify samples of the same developmental stage, before H3K4me3 levels were compared between different treatments. H3K4me3 interference approaches were considered successful if a reduction of >50% in H3K4me3 levels was achieved.

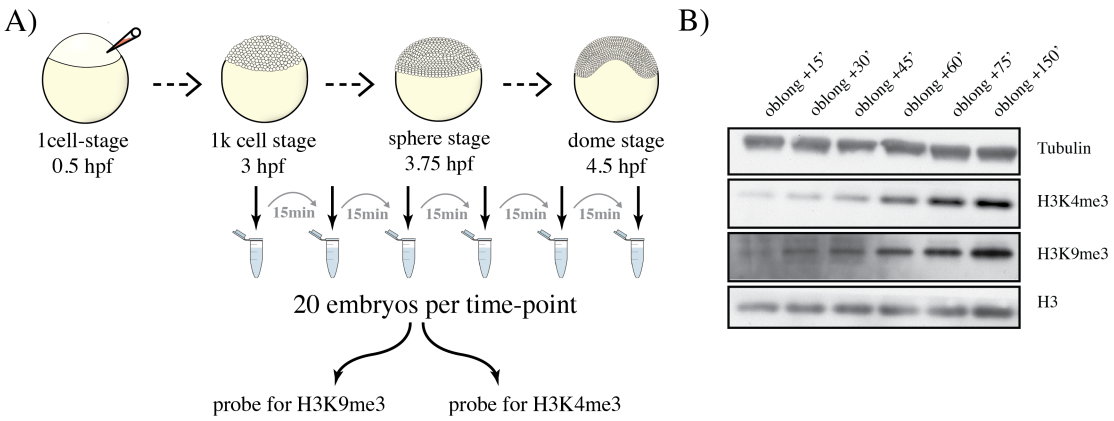


Figure 4.1: Western Blot assay to monitor H3K4me3 reduction A) Embryos were injected at the one-cell stage with respective constructs, and left to develop until genome activation (1000cell stage). A sample of 20 embryos per stage and treatment was taken every 15 minutes until dome stage. Samples were split in half and used for Western blotting on two separate denaturing gels to probe for H3K9me3 and H3K4me3 in parallel. B) Western Blot results for H3K4me3 and H3K9me3 over developmental time. A sample of 10 wildtype embryos was used per lane. Tubulin and H3 serve as controls for equal loading, H3K4me3 and H3K9me3 were probed on separate denaturing gels.

4.2 Overexpression of H3K4-specific histone demethylases does not result in global H3K4me3 reduction

In contrast to other histone modifications, H3K4me3 is biochemically robust and removal relies on active processes (Byvoet et al., 1972; Duerre and Lee, 1974). Recently, specific histone lysine demethylases (KDM) were found to remove H3K4me3 from its genomic targets, thus counteracting H3K4me3 establishment (Seward et al., 2007; Shi et al., 2004). In vertebrates, four different isoforms of the H3K4-specific KDM5 family are known (Iwase et al., 2007). To remove newly established H3K4me3 in zebrafish embryos, I obtained constructs encoding each of the four KDM5 homologs found in mouse (KDM5 A through D), which were found to reduce global H3K4me3 levels in mouse stem cell cultures (Bieberstein et al., Neumann et al., *unpublished*). To monitor for successful expression, KDM5 constructs were subcloned and fused with an N-terminal Flag-tag.

To analyze whether overexpression of KDM5 constructs resulted in H3K4me3 reduction, I injected *in vitro*-transcribed RNA from respective KDM5 constructs into dechorionated one-cell stage embryos. Samples of uninjected embryos were taken as wildtype control. To control for effects resulting from RNA injection alone, I used RNA coding for GFP with nanos-3'UTR, which should not be involved in H3K4me3 establishment. For each KDM5 construct, 3 biological replicates of the described Western blot assay were performed over multiple stages. However, when comparing H3K4me3 levels between uninjected, mock RNA-injected and KDM5-injected embryos, neither of the four mouse KDM5 constructs alone resulted in a significant reduction of H3K4me3 levels during early embryogenesis (Figure 4.2 A and B).
It is possible that the four KDM5 isoforms have different DNA recruitment mechanisms. Therefore overexpression of one KDM5 isoform could result in local demethylation at a subset of genes. To test whether a combination of KDM5 isoforms results in global H3K4me3 reduction, I overexpressed a mix of all for mouse KDM5 isoforms. However, injecting a KDM5 construct mix did not result in significant reduction of H3K4me3 levels either (Figure 4.2 C).
Because KDM5 recruitment could be regulated by species-specific mechanisms, it is possible that mouse KDM5 proteins were unable to demethylate H3K4me3-marked regions in zebrafish embryos. I therefore I obtained a construct coding for zebrafish

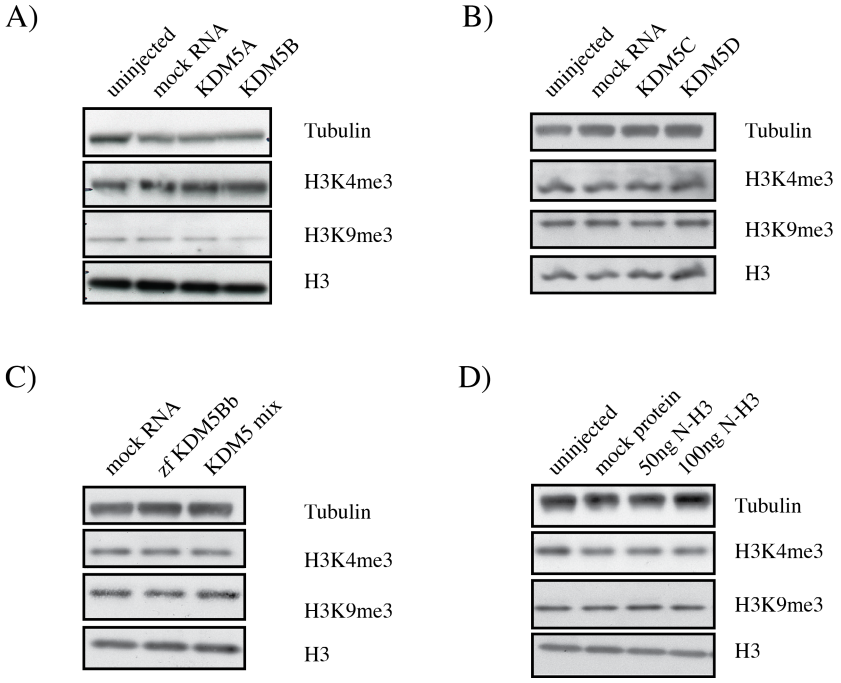


Figure 4.2: Global levels of H3K4me3 are not affected by KDM5 overexpression or injection of N-terminal histone tails Representative results from Western Blot assays for H3K4me3 reduction. Each lane represents a sample of 10 embryos. H3K9me3 serves as staging control, Tubulin and H3 serve as loading control between samples. A-C) Comparison of uninjected and mock RNA (GFP with nanos-3'UTR) injection samples with overexpression KDM5A and KDM5B (A), KDM5C and KDM5D (B) or overexpression of zebrafish KDM5Bb and a mix mouse KDM5 construct overexpression samples (C). For each construct, 1 ng of RNA was injected. D) Representative Western Blot comparing uninjected samples to samples injected with mock protein (BSA, 100 ng) and injection of 50 ng and 100 ng of inhibitory N-terminal H3 peptide.

KDM5Bb from the Schier lab at Harvard. I subcloned the coding sequence of KDM5Bb and fused an N-terminal Flag-tag to the coding sequence. I used the resulting *in vitro*-transcribed RNA to perform the Western Blot reduction assay as described above. However, overexpression of zebrafish KDM5Bb did not result in global H3K4me3 reduction when compared to uninjected and mock RNA injected embryos.

From these results I conclude that overexpression of KDM5 constructs is not a valuable strategy to achieve global H3K4me3 reduction in early embryos. There may be several reasons for this. First, although overexpression of these constructs could be confirmed by the presence of Flag-tagged proteins of the expected height, their levels might be too low to achieve efficient reduction of H3K4me3 levels. Additionally, the presence

of the protein monitored by Western Blotting does not allow for the conclusion that this protein is fully functional in the embryo. It cannot be excluded that the short time from RNA injection to collection of samples for monitoring of H3K4me3 interference (approximately 4 hours) is too little to allow for complete protein folding of these large proteins (>150kDa). Finally, while it has been shown that KDM5 proteins have chromatin binding domains (Liu and Secombe, 2015), it is possible that they need further complex partners to achieve efficient H3K4me3 demethylation in zebrafish embryos.

4.3 Global reduction of H3K4me3 could not be achieved by small-molecule inhibition of HMT activity

H3K4me3 establishment relies on the recognition of histone H3 Lysine 4 by specific histone methyl transferase complexes, and the transfer of a methyl-group from the co-substrate S-Adenosyl-methionine (SAM) to the lysine side chain. HMT specificity towards H3K4 is achieved by binding to the protein N-terminus of H3, as well as to amino acids in close vicinity to lysine 4.

To achieve reduction of H3K4me3 levels, I attempted to inhibit H3K4-specific HMTs by competing with histone H3 substrate binding. To that end I obtained a peptide that is equal to the 8 N-terminal amino acids of H3 from GenScript. I injected 50 and 100 ng N-terminal peptide into 1-cell-stage embryos. These amounts represent 57- and 115-fold excess of inhibitory peptide compared to H3 protein levels in the zebrafish embryo, respectively (100 ng peptide are $6.46 * 10^{13}$ molecules, while zebrafish embryos contain $5.58 * 10^{11}$ H3 proteins; Joseph et al., *in preparation*). Peptide-injected embryos were subjected to the Western Blot assay described above. As a control for possible erroneous effects of protein injection, 100 ng of BSA was injected. Protein injection itself seems to have an influence on H3K4me3 establishment, as peptide-injected as well as BSA-injected embryos show lower levels of H3K4me3 (one representative blot shown in Figure 4.2 D). However, when comparing peptide-injected embryos to BSA-injected embroys, no significant change in H3K4me3 levels could be observed over three biological replicates. Higher amounts of protein could not be injected without causing developmental defects both in peptide-injected as well as BSA-injected embryos. Thus, competitive inhibition of H3K4me3 establishment by peptides was not successfull to achieve H3K4me3 reduction.

Another strategy to inhibit enzymatic activity is to introduce cofactor analogues. However, SAM is a widely used co-metabolite that is not only used in H3K4-methylation,

but also in other protein methylation processes, as well as for DNA methylation and multiple metabolic pathways (Etchegaray and Mostoslavsky, 2016; Kalhan and Marczewski, 2012; Moussaieff et al., 2015). Thus, interference with SAM is not a promising strategy to specifically interfere with H3K4me3 establishment. Recently, two H3K4-specific inhibitors were published to reduce H3K4me3 levels *in vitro* (Cao et al., 2013; Grebien et al., 2015). Both drugs target the interaction interface of the essential cofactor Wdr5 with Mll1 HMT complexes. However, I have shown that interference with *mll1* does not result in global H3K4me3 level changes (section 3.3), making these drugs unlikely candidates for global H3K4me3 reduction. Additionally, recent reports have shown that inhibition of Mll1 complexes by these drugs resulted in a decrease of H3K4me1 - but not H3K4me3 - levels in mouse epiblast stem cells (Zhang et al., 2016). Thus, no H3K4me3-specific inhibitor could be found to reduce H3K4me3 levels in zebrafish embryos.

4.4 Overexpression of K4-specific methylation-defective H3 results in global H3K4me3 reduction

Inhibition of global H3K4me3 establishment could not be achieved by small-molecule interference with H3K4-specific HMTs. However, inhibition of H3K4me3 establishment could also be achieved by inhibition of H3K4me3 establishment within its genomic recruitment sites. Introducing full length histone proteins that cannot be methylated at amino acid 4 could be a valuable strategy to prevent H3K4me3 establishment. To this end, I introduced histone proteins with amino acid substitutions at lysine 4. I tested two constructs with different amino acid substitution for their effect on global H3K4me3 establishment.

4.4.1 Overexpression of H3K4-to-E constructs does not affect global H3K4me3 establishment

I chose to substitute H3 lysine 4 by glutamate (K4E). Glutamate has a similar size but an opposing charge compared to lysine. K4E has been predicted to prevent the HMT enzymatic pocket from binding to these histone proteins by the negative charge. Replacement of H3 proteins with K4E constructs was used previously to reduce H3K4me3 specifically on the histone variant H3.3 in Xenopus embryos (Ng and Gurdon, 2008). To test whether H3 K4-to-E substitution results in a reduction in global H3K4me3

levels, I overexpressed K4E constructs C-termianlly tagged with HA-tag in zebrafish embryos. To that end I injected *in vitro*-transcribed RNA for full length H3 with K4-to-E substitution into one-cell-stage embryos. To control for effects resulting from RNA injection alone, I used RNA coding for GFP with nanos-3'UTR, which should not be involved in H3K4me3 establishment. Histone proteins have been described to affect embryonic genome activation and development (Amodeo et al., 2015, Joseph et al., in preparation). To control for developmental effects caused by histone overexpression alone, I further injected embryos with RNA coding wildtype H3 with C-terminal HA tag (H3HA). For each experiment, embryos were injected with either mock RNA, H3HA RNA or K4E RNA at one-cell stage and H3K4me3 levels were monitored by Western Blotting as described above (section 4.1). When comparing stage-matched H3HA-injected and K4E-injected samples for their H3K4me3 levels, no significant difference in H3K4me3 levels were observed in three biological replicates of multiple stages (one representative example shown in Figure 4.3 A). These results showed that K4E overexpression did not result in global H3K4me3 interference in zebrafish embryos. A possible reason is that the achieved protein levels from overexpression are not enough to outcompete the maternal load from wildtype histone proteins. Despite the injection of high amounts of RNA, only 20% of additional H3 protein was added to already present histone pools.

4.4.2 H3K4-to-M constructs act as dominant-negative substrate for H3K4me3 establishment

Recently, spontaneous mutations resulting in a lysine-to-methionine (K-to-M) substitution within histone H3 were identified in multiple cancer samples (Fang et al., 2016a). Interestingly, already if this mutation occurs in only one out of multiple gene copies, methylation levels of the respective lysine residue were significantly reduced. These observations were found to target H3K27 (Chan et al., 2013; Lewis et al., 2013) as well as H3K36 (Shah et al., 2014) methylation marks. These findings suggest that K-to-M substitution within histones act as dominant-negative inhibitor of histone methylation processes, and might be applicable to other histone lysine residues as well. To test whether K-to-M substitution results in global reduction of H3K4me3 levels in zebrafish, I designed constructs coding for Histone H3 with K4 to M substitution (referred to as K4M). I introduced a C-terminal HA-tag to allow for detection of protein overexpression.

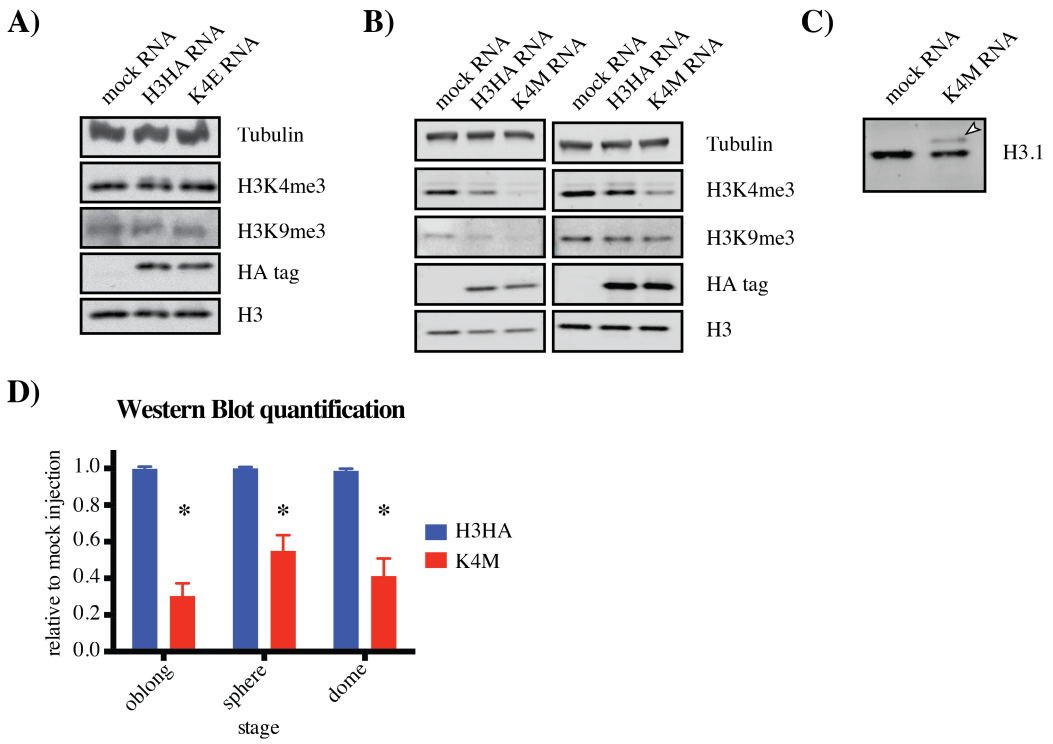


Figure 4.3: H3K4M, but not H3K4E introduction leads to global H3K4me3 reduction A,B) Representative Western Blots comparing samples of mock RNA injection (GFP with nanos-3'UTR) with injection of H3HA and H3K4E (A) or H3HA and H3K4M constructs (B). Each lane represents a sample of 10 embryos. HA-tag controls for equal expression of H3 constructs, Tubulin and H3 serve as loading control, H3K9me3 serves as developmental control. C) Western Blot of mock-injected and H3K4M-injected samples at sphere stage probed for H3.1 protein content with an antibody that recognizes C-terminally tagged histone H3. Arrowhead indicates band resulting from introduced C-terminal HA-tagged histone H3. D) Quantification of H3K4me3 level reduction upon K4M (red) overexpression when compared to H3HA overexpression (blue) at stages oblong, sphere and dome. n = 4, error bars = s.d., * $p<0.05$

Histone H3 proteins and mRNA are maternally provided, and histone genes belong to the earliest activated genes during early zebrafish development (Heyn et al., 2014; Pauli et al., 2012). To analyze how the protein levels of introduced K4M constructs compare to wildtype H3 in early embryos, I injected *in vitro*-transcribed mRNA into one-cell stage embryos, and I tested histone H3 expression at sphere stage. To detect both wildtype H3 and K4M proteins, I used an antibody that recognizes the globular domain of H3. The C-terminal HA-tag is shifting the molecular weight of introduced K4M constructs by 1kDa, which allows for differential detection of wildtype and K4M

proteins with one antibody (Figure 4.3 C). Quantitative comparison of the two distinct bands by fluorescent Western Blotting revealed that K4M proteins comprise around 20% of wildtype H3 levels.
Next I set out to test whether K4M overexpression results in H3K4me3 reduction. I injected wildtype embryos with RNA encoding K4M constructs in parallel with RNA coding for wildtype histone H3 tagged with HA (H3HA). Similar expression levels were monitored by HA-tag expression (Figure 4.3 B, HA band intensity). Again, H3K9me3 levels were used to stage-match samples of different treatment (section 4.1). While only 20% of extra histone protein could be achieved by injection H3 RNA injection, a drastic effect on H3K4me3 levels could be observed upon K4M overexpression when compared to H3HA injection (Figure 4.3 B). Four independent experiments consistently resulted in more than 50% reduced H3K4me3 levels over multiple stages (Figure 4.3 D). These results suggest that K4-to-M substitution indeed acts as a dominant negative inhibitor of global H3K4me3 establishment in zebrafish embryos. These results are in line with recent findings where H3K27me3 and H3K9me3 levels can be reduced in Drosophila imaginal wing discs by overexpression of respective lysine-to-methionine substitutions (Herz et al., 2014). H3K4-to-M substitution is regularly used in protein chrystallography to mimic dimethylated lysine residues (Fang et al., 2013). Therefore it is hypothesized that the presence of methionine in the context of full-length histones is a dominant substrate for the enzymatic pocket of histone methyltransferases (Lewis et al., 2013). In conclusion, overexpression of K4M constructs in zebrafish embryos results in a drastic reduction of H3K4me3 levels, possibly by representing a dominant-negative substrate for H3K4-specific HMTs.

4.5 H3K4me3 levels at gene promoters are reduced upon introduction of methylation-defective Histone H3

Overexpression of K4M histone constructs resulted in a significant reduction of global H3K4me3 levels when analyzed by Western Blotting. Western Blot analysis do not allow for an interpretation of the local distribution of H3K4me3 reduction. The drastic reduction in total H3K4me3 levels could be either the result of a complete loss of H3K4me3 levels on a significant fraction of genes, or a significant reduction on all gene promoters. These options are not mutually exclusive. To understand how the achieved reduction of global H3K4me3 levels is reflected in local changes in H3K4me3 levels, I performed Chromatin Immunoprecipitation experiments followed by Deep

Sequencing (ChIP-Seq). ChIP-Seq experiments were performed at four subsequent stages after genome activation (high, oblong, sphere dome) in parallel for H3HA and K4M overexpression, to control for effects that could be caused by overexpression of histone H3 proteins alone.

Introduction of Lambda phage-DNA allows for correction of overamplification

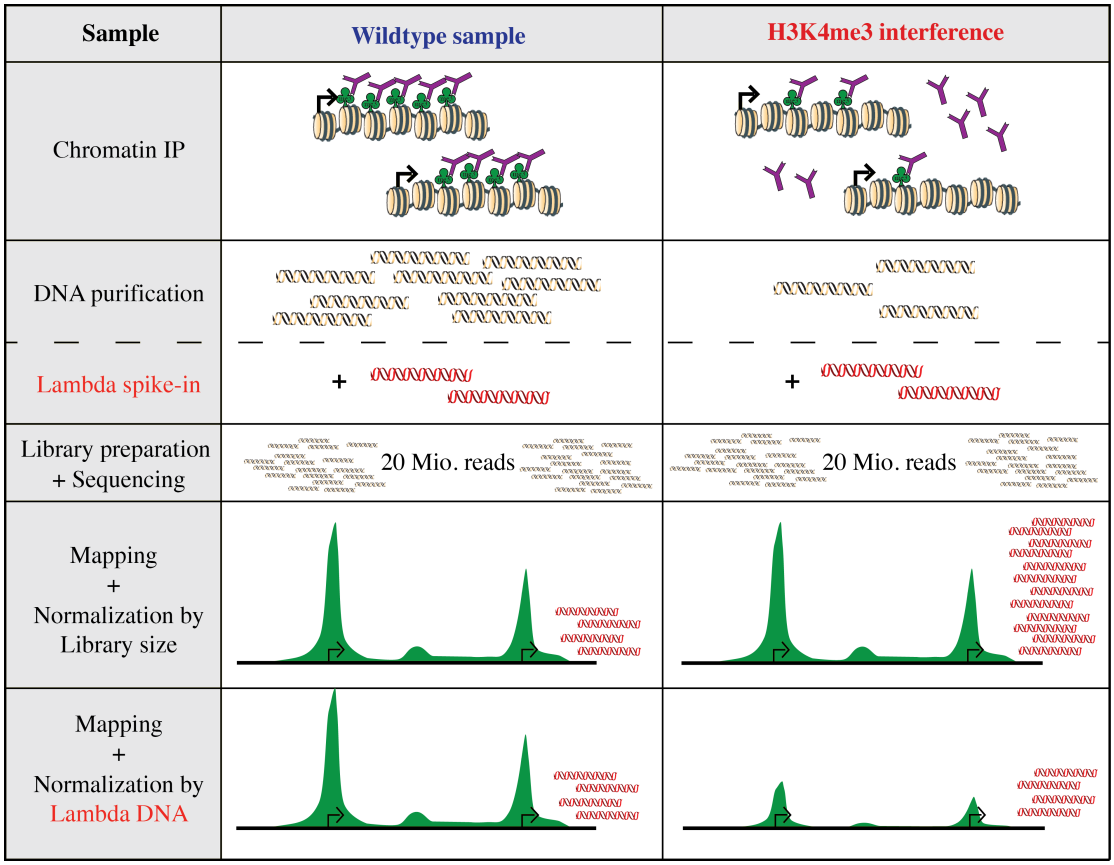


Figure 4.4: Lambda DNA spike-in allows for correction of overamplification during Sequencing If interference with a histone modification (here H3K4me3) results in a global reduction, Chromatin Immunoprecipitation results in less purified DNA. During Library Preparation and Next Generation Sequencing, sample DNA is amplified to reach a predefined amount of Sequencing reads. Samples with initially less DNA is overamplified to reach the same information. Thus, normalization of sample data by library size does not represent changes in initial DNA quantity. Introduction of a given amount of spike-in DNA (here Lambda DNA) allows for monitoring overamplification during sequencing, and is a valid tool to allow normalization for DNA quantity differences in initial samples.

Multiple recent studies found that global reduction of histone modification is not reflected in standard ChIP-Seq approaches (Bonhoure et al., 2014; Chen et al., 2016; Orlando et al., 2014). This is explained by increased amplification of material in samples with lower DNA quantity during the process of library preparation and sequencing. This overamplification is not reflected in the sample library size (the amount of reads that can be mapped to genomic positions) after sequencing, which is usually used to normalize quantitative differences in subsequent data analyses.

Most studies address sample normalization by post-procession of the acquired data by bioinformatics analysis. These strategies rely on normalization over specific genomic locations that are known to be not affected by the experimental treatment. Therefore, this post-processing does not allow for analysis of experiments where global changes in histone modifications are expected, and no prior knowledge about not-targeted regions exist. Therefore, spike-in strategies to correct for overamplification of immunoprecipitated material are essential for a successful quantitative comparison between different samples.

To allow for quantitative comparison of my ChIP experiments, I added a given amount of fragmented Lambda phage DNA to each sample (10% of reference sample DNA). After sequencing, the amount of reads mapped to Lambda phage were used to normalize quantitative differences between samples of the same stage (subsection 2.3.2). If Lambda DNA spike-in allows for normalization of overamplification during library preparation and sequencing, the ratio of Lambda Phage reads between two samples should reflect the quantitative difference of DNA prior to sequencing better than the actual library size. Indeed, my data showed that the ratio of Lambda Phage-specific reads after sequencing directly reflects the quantitative differences in input DNA, especially in early stages (Table 4.1, compare Lambda ratio to DNA ratio at individual stages). Therefore I conclude that the amount of Lambda Phage-specific reads is a good reference to normalize for quantitative differences of immunoprecipitated DNA that are the result of global reduction of targeted histone modifications.

Table 4.1: Lambda Phage library size reflects differences in sample DNA content The amount of DNA in each sample (in ng) compared to total mapped reads (Total library), Lambda library size, and their relative ratios to respective H3HA samples. Lambda library factor reflects best the observed differences in DNA content.

sample	DNA content	Total library	Lambda library	DNA factor	Library factor	Lambda factor
H3HA high	12.01	6,165,525	937,701	1.000	1.000	1.000
K4M high	1.09	6,410,505	6,809,612	0.091	1.040	0.138
H3HA oblong	11.69	3,927,642	206,535	1.000	1.000	1.000
K4M oblong	0.70	5,050,954	11,949,412	0.060	1.286	0.078
H3HA oblong2	2.80	13,954,143	7,276,341	1.000	1.000	1.000
K4M oblong2	n.d.	6,910,026	31,579,042	n.d.	0.495	0.230
H3HA sphere	4.01	11,770,538	4,331,846	1.000	1.000	1.000
K4M sphere	3.80	12,482,032	4,663,928	0.951	1.060	0.929
H3HA dome	16.70	12,193,878	221,680	1.000	1.000	1.000
K4M dome	13.80	9,257,326	261,696	0.834	0.759	0.847

K4M overexpression results in a strong reduction of promoter-associated H3K4me3 peaks

I have shown before that H3K4me3 levels are enriched at gene promoters during early embryonic development (Figure 3.7). Western blot experiments have shown that K4M overexpression results in drastic reduction of global H3K4me3 levels (Figure 4.3). To analyze whether this reduction is also reflected in H3K4me3 levels at gene promoters, I analyzed the obtained ChIP-Seq data for its enrichment at promoter regions. I plotted the density of reads within +/-3kb of transcription start sites (TSS) annotated in the zebrafish genome assembly. Read density was normalized to Lambda phage library, to allow for cross-sample and cross-stage comparison. Enrichment of H3K4me3 around TSS increased dramatically over developmental stages (Table 4.1, note the different y-axis scale for dome stage). While almost no enrichment at TSSs is found at high stage, H3K4me3 are high at dome stage when averaged over all genes. This is in line with the increase of global H3K4me3 levels over developmental stages observed by Western Blot analyses (Figure 4.1). When comparing samples from H3HA-injected embryos (blue line) with samples from K4M overexpression (red line), it becomes apparent that

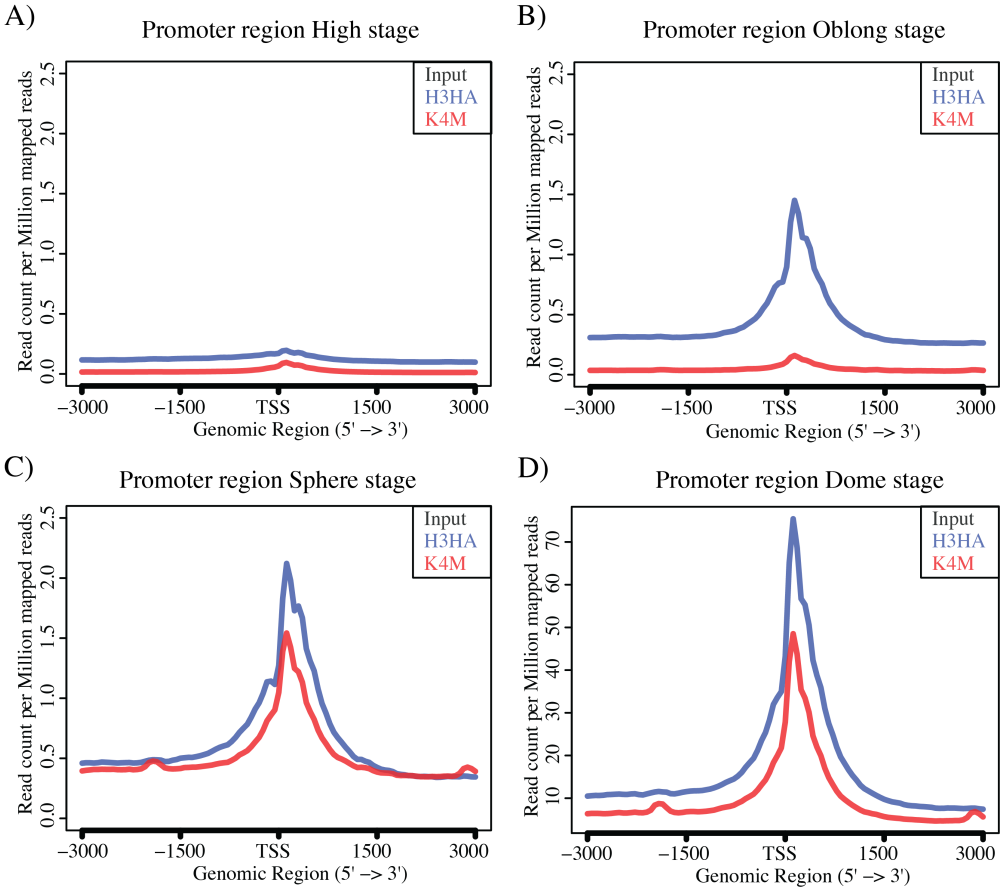


Figure 4.5: H3K4me3 is reduced at promoter regions upon K4M introduction Metagene profiles displaying enrichment of H3K4me3 ChIP-Seq reads over promoter regions. Depicted are input chromatin enrichment (black), and H3K4me3 enrichment upon H3HA (blue) or K4M introduction (red) at high (A), oblong (B), sphere (C) and dome (D) stage. Enrichment plots were scaled to Lambda DNA library size.

H3K4me3 enrichment is significantly reduced at all stages upon K4M overexpression (Figure 4.5). Especially at early stages shortly after genome activation, H3K4me3 levels are dramatically reduced at TSS. While the difference in H3K4me3 levels is not as strong for later stages, H3K4me3 stages are significantly reduced in all stages. It is possible that the initial loss of H3K4me3 is compensated by H3K4me3 establishment coupled to active transcription at a subset of genes.

To understand whether the reduction in H3K4me3 enrichment at promoters is also reflected in H3K4me3 reduction at individual promoters, I analyzed Lambda Phage-normalized H3K4me3 enrichment tracks of H3HA-overexpressed and K4M-overexpressed samples. By manual inspection I observed that the majority of promoter regions showed

a significant H3K4me3 reduction. If a locus was reduced in H3K4me3 levels at one stage, this locus showed H3K4me3 reduction at other stages, too (one representative example shown in Figure 4.6 A). Interestingly, not all loci show H3K4me3 reduction. A small subset of genomic regions was unaffected by K4M overexpression (Figure 4.6 B) over different stages, suggesting that these regions are not susceptible to K4M introduction.

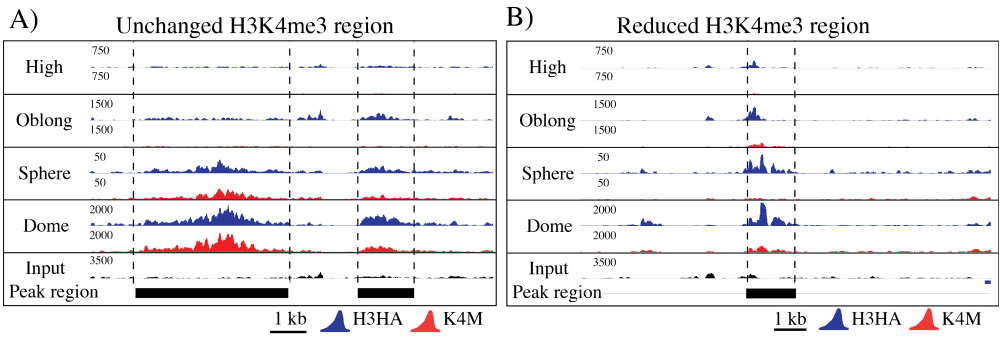


Figure 4.6: K4M overexpression reduces H3K4me3 levels reproducibly at TSS A) Representative genomic region showing a region (in dashed lines) that is signficantly reduced upon K4M overexpression (red) in all stages. B) Representative genomic region showing a region that is not significantly changed between H3HA (blue) and K4M overexpression (red). Dashed lines indicate analyzed regions.

To quantify the differences of H3K4me3 levels for individual regions over the whole genome, and to allow for the identification for loci with strongest H3K4me3 reduction, I developed a bioinformatics pipeline. This work was performed in collaboration with Holger Brandl from MPI-CBG. The following steps were applied to ChIP-Seq data from all stages and both treatments (H3HA and K4M overexpression).

1. I filtered sequencing reads based on standard quality criteria. Furthermore, I rejected reads that map to more than one position in the genome (genome version danRer7/Zv9 from April 2010 was used).
2. Using the remaining reads, I performed peak calling by MACS2.
3. To reduce false-positive calls by MACS2, I reasoned that true H3K4me3-enriched regions should be present at the same genomic location in more than one sample. Since no biological replicates for single stages were performed, I reasoned that true peak regions should be stable over developmental stages. Since K4M results

in strong reduction of H3K4me3 levels, I only considered samples from H3HA injection, and rejected peak regions that are present in less than 3 of 5 ChIP-Seq datasets.

4. The retained peak regions from all samples were merged to obtain a set of regions that reflects a common peak region over all developmental stages.
5. Additionally, I only retained peak regions which were within +/-5kb distance to annotated transcription start sites, to allow for inclusion of promoter-associated H3K4me3 peak in the vicinity of annotated transcription start sites.
6. Next, for each treatment and developmental stage, the number of reads that cover each of the identified peak regions was assessed.
7. The resulting counts were normalized by Lambda library size, and fold change was calculated comparing K4M overexpression samples versus H3HA overexpression samples at each stage (Pataskar et al., 2016).

In my experimental set-up, different developmental stages reflect biological replicates. By manual inspection I observed that local H3K4me3 level changes induced by K4M overexpression are reproducible over developmental stages. To assess whether this observation is true when considering all promoter regions, I correlated local H3K4me3 reduction over successive stages (high versus oblong, and sphere versus dome). Scatterplot analysis of the fold change (K4M versus H3HA overexpression) are shown in Figure 4.7 A and B.

A clear correlation between fold changes can be observed, suggesting that local H3K4me3 reduction is reproducible over developmental stages. Taken together, these results suggest that K4M overexpression results in local H3K4me3 reduction that is reproducible over developmental stages and biological replicates.
Next I wanted to understand, how many regions are reduced upon H3K4me3 levels. I plotted the number of regions that show more than 2fold and 4fold reduced (or increased) levels of H3K4me3 enrichment. At none of the analysed stages, regions that are more than 2fold increased in H3K4me3 levels in K4M-overexpressed samples were identified. In contrast, upon K4M overexpression around 90% of genomic regions show more than 2fold reduced H3K4me3 levels at each individual stage. At high and oblong stages, the majority of regions is even more than 4fold decreased (58 and 64%, respectively). I further assessed which regions show a reproducible decrease in

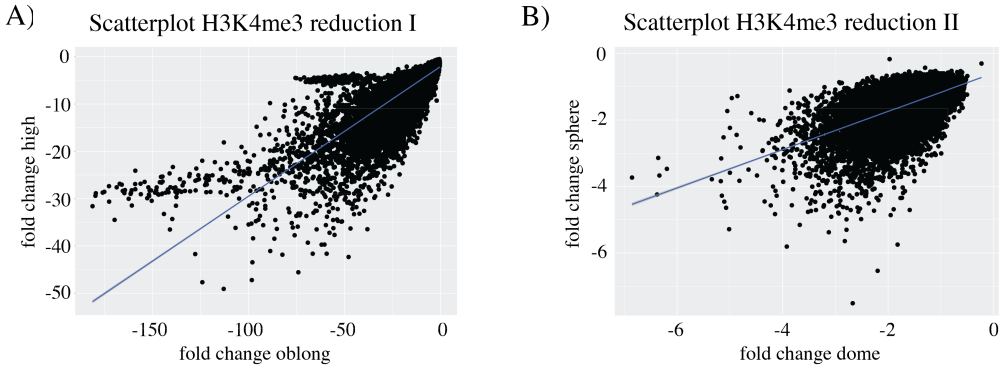


Figure 4.7: K4M overexpression reduces H3K4me3 levels reproducibly at TSS Scatterplots of H3K4me3 fold changes observed at high (y-axis) and oblong (x-axis) (C), as well as sphere (y-axis) and dome (x-axis) stage (D). Blue line represents linear fit.

H3K4me3 levels in all the analyzed stages. Of almost 33,000 analyzed region, 1,264 regions (3.84%) show more than 4fold reduced H3K4me3 levels at each analyzed stages, and more than 75% of these regions are at least 2fold reduced. This again reflects that K4M robustly and reproducibly reduces H3K4me3 levels in early zebrafish development.

Taken together, these results identify overexpression of histone H3 constructs with lysine-4-to-methionine substitution as a dominant negative histone methyl transferase inhibitor in zebrafish development. Overexpression of these histone constructs resulted in a drastic and reproducible reduction of H3K4me3 levels at individual promoters throughout the zebrafish genome. Thus, K4M overexpression is a valuable tool to study the effects of H3K4me3 reduction on transcription initiation in zebrafish development.

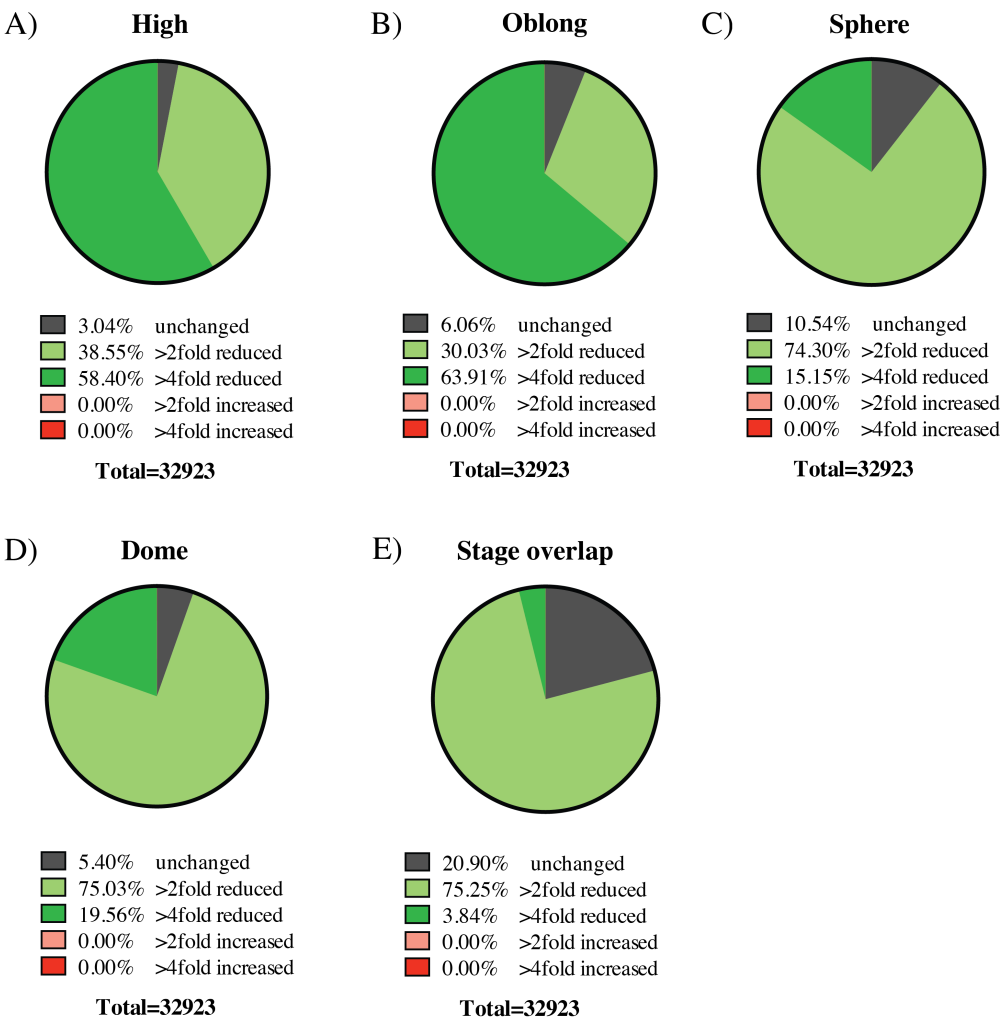


Figure 4.8: The majority of promoters shows reduced H3K4me3 levels Depicted are the fractions of genes with indicated fold change difference of H3K4me3 levels at promoter regions. Percentage of promoters that are significantly changed are shown at high (A), oblong (B), sphere (C) and dome stage (D). Peak regions that show the same trend in all stages are shown in (E).

4.6 Early transcription activation is not altered upon K4M overexpression

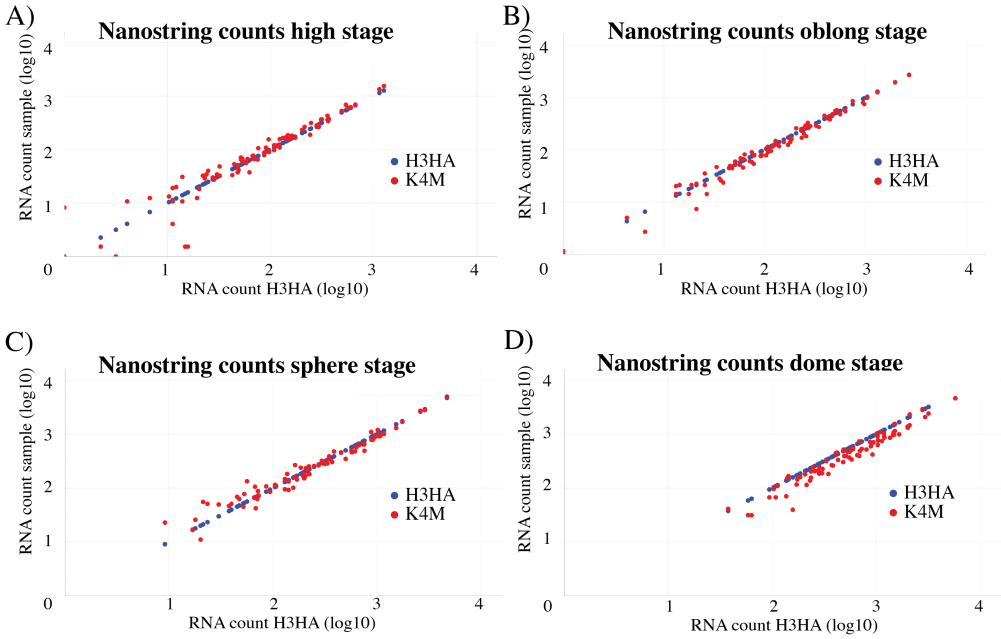


Figure 4.9: K4M overexpression does not result in transcriptional change of zygotic genes Depicted are Scatterplots correlating transcript counts of 84 zygotic activated genes in H3HA overexpression (x-axis, blue) versus K4M overexpression (y-axis, red). Counts were measured by Nanostring assays at high (A), oblong (B), sphere (C) and dome (D) stage.

K4M overexpression resulted in at least two-fold reduced H3K4me3 levels on more than 90% of gene promoters in stages shortly after genome activation (high and oblong, see Figure 4.8). To investigate whether H3K4me3 reduction affects transcription already at the first stages after genome activation, I analyzed the onset of zygotic transcription using the previously described NanoString assay (section 3.4 and subsection 2.2.7). The here applied NanoString assay probes for 12 genes with stable expression levels during early developmental stages ('normalizer genes') and 84 genes that are not maternally provided, but activated early during zebrafish development ('zygotic genes'). To compare expression levels of early activated genes, I extracted RNA from H3HA-injected and K4M injected embryos at six consecutive stages during genome activation (512cell stage to dome stage). RNA levels were quantified by NanoString assay, and normalized RNA counts for each individual gene were compared between H3HA-

injected and K4M-injected samples. No significant difference of transcript levels could be observed at early stages (high, oblong and sphere). At dome stage, slightly reduced transcript levels were observed for the majority of genes (Figure 4.9 D). While the reduction of individual transcript levels was low (less than 20% for each individual gene), the majority of genes showed the same trend (see individual time courses of genes in chapter 6.6).
It is possible that reduction of H3K4me3 levels is causative for the observed subtle changes in RNA transcript levels. To understand whether H3K4me3 reduction directly correlates with changes in transcript levels, I plotted the transcriptional change observed by NanoString assays against the fold change of promoter-associated H3K4me3 levels at oblong (early after genome activation) and dome stage (when strongest transcriptional changes are observed). If H3K4me3 level changes are causative for transcriptional changes, strong reduction of H3K4me3 levels should result in strong reduction of RNA transcript levels, while genes that show no H3K4me3 reduction should not show transcription level changes. However, scatterplot analyses that correlate H3K4me3 level changes against transcriptional change did not reveal a significant correlation of transcriptional changes to changes in H3K4me3 (Figure 4.10). Instead, strong reduction in transcription levels was observed for both, genes that show increased as well as decreased transcription levels. Furthermore, genes with increased and decreased transcription levels were found for almost similar H3K4me3 level changes both at oblong and dome stage. Thus, no correlation between H3K4me3 levels and transcriptional activity could be observed, suggesting that H3K4me3 reduction does not cause the observed changes in transcript levels.

Early zygotic genes need to be robustly activated to allow for further developmental progression (Kane et al., 1996; Newport and Kirschner, 1982). Therefore it is possible that early activated genes have robust activation mechanisms that overcome interference with chromatin modifications. Thus, early transcription activation might not rely on H3K4me3 to be established. However, H3K4me3 has been suggested to poise developmental regulators for transcriptional activation, as genes important for embryonic development are enriched in H3K4me3 levels prior to their activation (Bernstein et al., 2006; Vastenhouw et al., 2010). Many of these genes are only activated at later stages, when the embryonic genome might be less permissive for transcription activation (Bogdanovic et al., 2011; Schneider et al., 2011; Wen et al., 2009). I therefore reasoned that perhaps H3K4me3 levels might have an effect on transcription activation

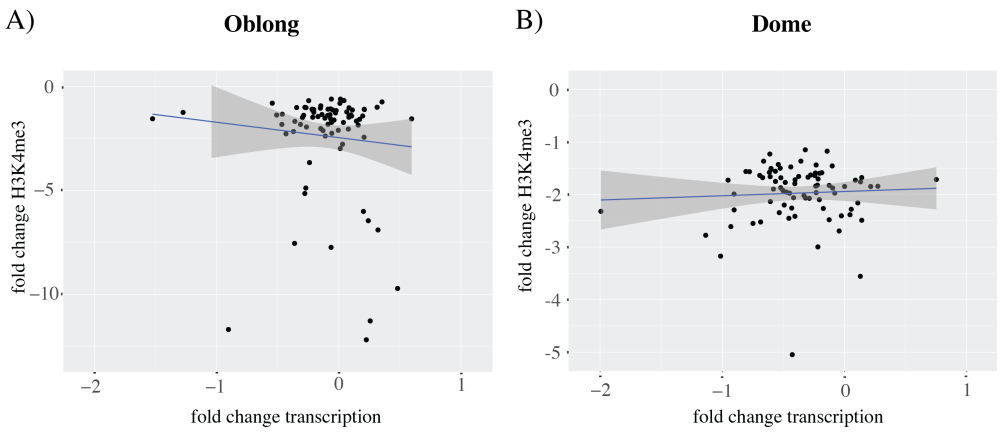


Figure 4.10: Transcriptional change does not correlate with change in promoter H3K4me3 levels Depicted are Scatterplots of transcriptional change (x-axis) measured by Nanostring assay versus change in H3K4me3 levels at promoter regions of respective genes (y-axis) measured by ChIP-Seq analysis for oblong (A) and dome stage (B). Blue line represents linear regression, grey shaded area represents 95% confidence interval to linear regression.

of genes that are activated later during early gastrulation stages.

To analyze the effect of H3K4me3 reduction on later activated genes, I first chose genes that showed strongest reduction (more than 2fold reduction at all stages, and more than 4fold reduction at dome stage). From these, I chose genes that are transcriptionally activated at different stages: I chose 17 genes that show first increase in transcript levels at dome stage or later. If H3K4me3 affects genes that are later activated, these should show significant changes in transcription activation during H3K4me3 reduction. Furthermore I chose 14 genes that are activated early after genome activation, until oblong stage, and thus should not be affected by H3K4me3 reduction as observed by NanoString assays.

To test for changes in transcription activation of these genes, I performed RT-qPCR experiments. For each gene, expression levels were monitored for one stage before, and at least three consecutive stages after their respective transcription activation. Experiments were performed for at least three biological replicates, monitoring transcription levels in H3HA-injected and K4M-injected embryos. H3K4me3 reduction at respective gene promoters, and the transcription profile upon K4M overexpression for one representative gene of each class are shown in Figure 4.11.

When monitoring transcription levels of genes that are activated early during development, I was unable to identfy transcriptional changes. This is consistent with results

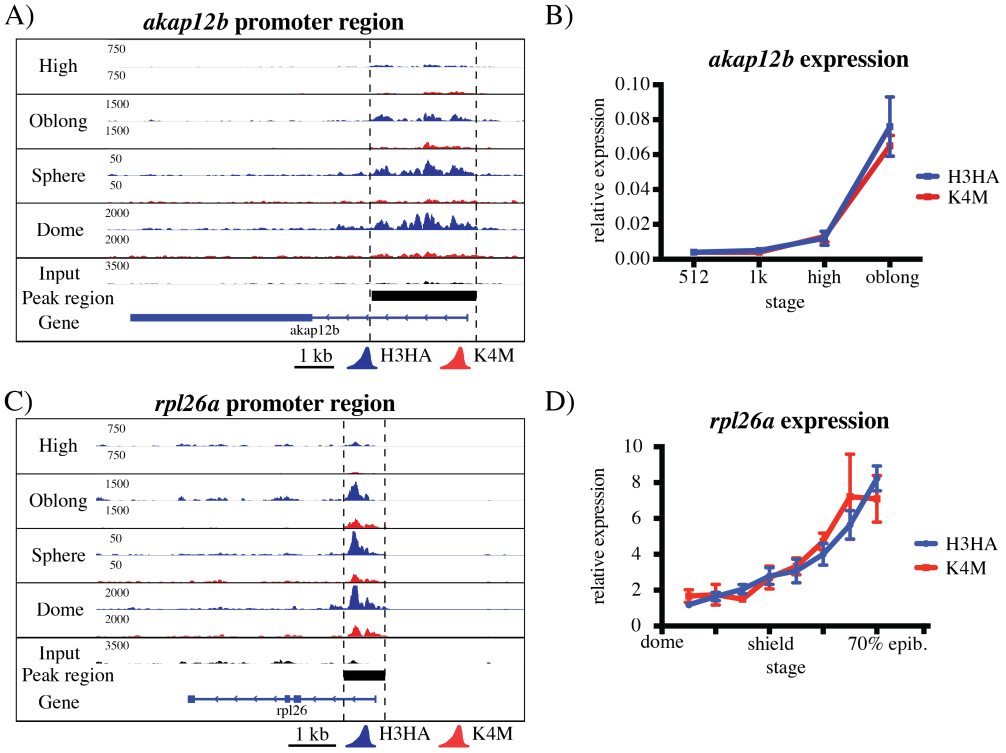


Figure 4.11: Local H3K4me3 reduction does not result in transcriptional change of target genes A) H3K4me3 levels of H3HA (blue) and K4M (red) samples from high to dome at *akap12b* promoter region (dashed line). B) Relative expression of *akap12b* to normalizer gene (eif4g2a) at stages 512 to oblong stage measured by RT-qPCR. H3HA samples are depicted in blue, K4M samples in red. n =4, error bars = s.d. C) H3K4me3 levels of H3HA (blue) and K4M (red) samples from high to dome at *rpl26* promoter region (dashed line). D) RT-qPCR experiments showing the relative expression to normalizer gene for *rpl26a* at stages dome to 70% epiboly. H3HA overexpression samples are depicted in blue, K4M overexpression samples in red. n = 3, error bars = s.d.

obtained by NanoString experiments. Next I assessed transcriptional changes from the 17 genes that are activated later than dome stage. Again, no significant changes in transcription levels could be observed. This suggests that like early activated genes, genes that are activated at dome stage and later do not rely on H3K4me3 establishment for their activation.

K4M overexpression resulted in a strong reduction, but not a complete loss of H3K4me3 levels at promoters, as analyzed by ChIP-Seq experiments. Therefore it cannot be excluded that the remaining levels of H3K4me3 at promoters are sufficient to allow for transcription activation in zebrafish embryos. Additionally it is possible that remaining

levels of H3K4me2 (which was not monitored during this study) might compensate for the lack of H3K4me3.
The experimental procedures did not allow for parallel monitoring of local H3K4me3 reduction by ChIP experiments and transcription analyses by RT-qPCR or NanoString. Therefore it cannot be excluded that in samples taken for transcription analyses, the respective promoters were not reduced for H3K4me3. However, since reduction of H3K4me3 at specific promoter elements was reproducible over developmental stages (Figure 4.6), and Western Blots performed in parallel to RNA extraction confirmed global reduction it is unlikely that variation in H3K4me3 reduction is causal for the absence of a transcriptional effect.

4.7 Conclusion

In conclusion, I could show that overexpression of histone H3 constructs with K4-to-M substitution (K4M) resulted in a drastic reduction of global H3K4me3 levels, possibly by representing a dominant-negative inhibitor for H3K4me3 establishment. K4M overexpression resulted in strong, reproducible H3K4me3 reduction at the majority of gene promoters during multiple stages of embryonic development. However, transcription analysis both on early activated genes as well as genes that are first activated during early gastrulation (dome stage) were unable to identify changes in transcription activation upon reduction of H3K4me3. Thus, H3K4me3 reduction seems not to affect transcription activation during early zebrafish embryogenesis.

5 Results III: Promoters rich in CpG and H2A.z gain H3K4me3 early

H3K4me3 has been shown to be established at more than 60% of gene promoters during zebrafish early embryonic development (Zhang et al., 2014). It has been proposed that H3K4me3 enrichment poises genes for future transcription activation (Bernstein et al., 2006; Lindeman et al., 2011; Vastenhouw et al., 2010). Yet, I was unable to identifiy significant transcriptional changes upon H3K4me3 reduction during early zebrafish development. This raises the question, why H3K4me3 is established at the majority of promoters during early developmental stages, if it is not involved genome activation. To better understand, which genes gain H3K4me3 during zebrafish embryogenesis, I grouped genes by their embryonic transcriptional profile and analyzed which group is favored for H3K4me3 establishment (section 5.1). Next, to address whether specific features of chromatin might be involved in H3K4me3 establishment during embryogenesis, I further analyzed H3K4me3 regions for specific features in the underlying DNA sequence (section 5.2), and a possible correlation with histone variant H2A.z (section 5.3). Together, these analysis have the potential to shed light on the underlying cues that lead to H3K4me3 establishment during early embryonic development.

5.1 H3K4me3 levels increase over developmental time at all gene classes

H3K4me3 is known to be enriched at promoters that are actively transcribed. However, recent studies proposed that H3K4me3 enrichment precedes transcription activation in embryonic development, and poises genes for future transcription activation (Bernstein et al., 2006; Lindeman et al., 2011; Vastenhouw et al., 2010).
During analysis of H3K4me3 reduction, I showed that H3K4me3 establishment increases over developmental time (Table 4.5). However, it is not clear whether these increasing levels are caused by Set1-mediated H3K4me3 establishment coupled to active transcription, or whether H3K4me3 becomes continuously established independent of transcription. The here obtained time series of H3K4me3 enrichment over developmental stages shortly after genome activation allows to address, which mechanism of

H3K4me3 establishment dominates in early embryonic development.
If H3K4me3 is established independent of active transcription, I expect that actively transcribed genes are not enriched for H3K4me3 compared to not transcribed genes. On the contrary, if transcription-coupled H3K4me3 establishment predominates in zebrafish embryos, genes with increasing transcript levels should be preferentially marked by H3K4me3, and transcriptionally silent genes should be depleted from H3K4me3. To understand whether H3K4me3 is enriched at transcriptionally activate genes, I classified genes based on their embryonic transcription profiles. Classification of genes into different embryonic transcription clusters was performed before, however either only a small subset of all genes were used for classification (Aanes et al., 2011), or classification was solely based on zygotic transcription activation (Harvey et al., 2013). To allow for the analysis of all known zebrafish genes, I made use of a recently published zebrafish embryonic transcriptome (Pauli et al., 2012) that monitors transcription levels up to 5 days post fertilization. I classified the genes into the following groups (for classification criteria see subsection 2.3.5):

1. Genes that are maternally provided
2. Genes that are not maternally provided, but will be transcribed during embryonic development until day 5
3. Genes neither provided maternally nor transcribed by the embryo until day 5
4. Genes that cannot be classified to either of these groups; this group is considered as a random choice of genes.

To assess H3K4me3 enrichment, I used the obtained ChIP-Seq data for H3HA-injected embryos at all analyzed stages (high to dome). At each stage, H3K4me3-enriched regions (peak) were identified by using MACS2. H3K4me3 density plots have shown that H3K4me3 enrichment is found within +/-2.5kb of TSS (Figure 4.5). To examine whether an individual gene could be considered positive for H3K4me3 enrichment, I tested whether a peak could be found within close proximity to its TSS (+/-2.5 kb) for each of the four analyzed stages. I then analyzed, whether a specific class of genes is enriched for H3K4me3-positive promoters.
When analyzing H3K4me3 enrichment, I observed that over developmental time more gene promoters become H3K4me3-positive (Figure 5.1 A). While at high stage only around 5,000 promoters (13%) are H3K4me3-positive, this number gradually increases to more than 20,000 promoters (61%) at dome stage. Interestingly, at each stage, a proportion of all classes is marked by H3K4me3 (Figure 5.1 A). This suggests that

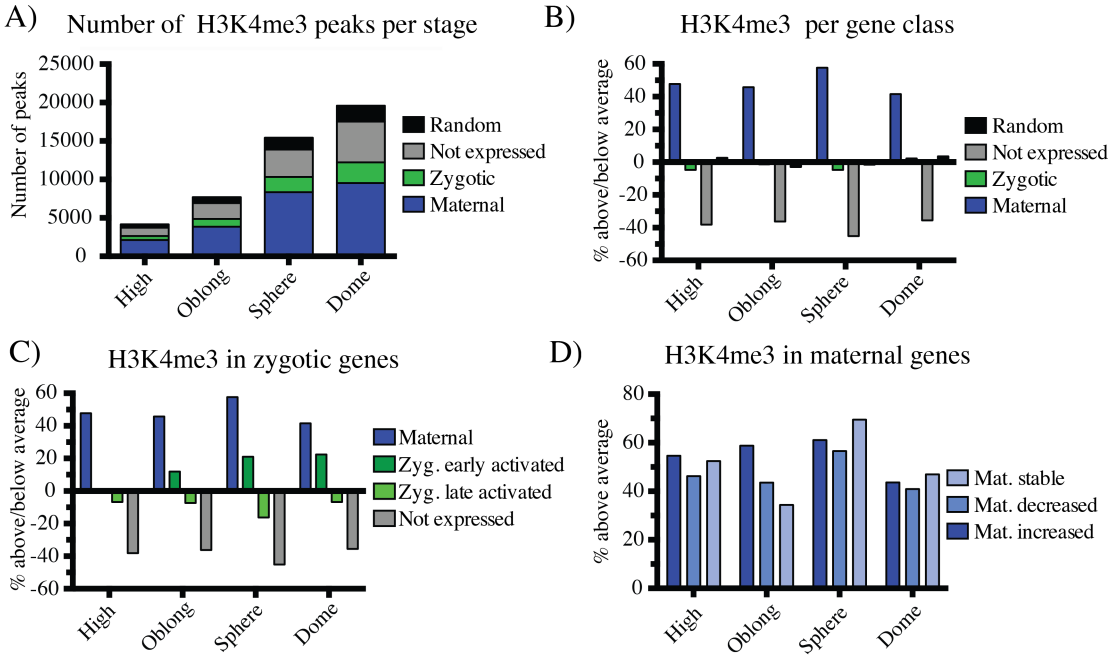


Figure 5.1: Maternal genes are overrepresented in H3K4me3 enrichment A) Number of MACS2-identified H3K4me3 peaks at stages high to dome. The fraction of peaks at gene promoters of maternally provided (blue), zygotic expressed (green), not expressed (grey) and random fluctuating genes (black) are indicated. B) Percentage of genes covered with peaks was calculated for each group, and compared to average percentage at the specific stage. Depicted is the percentage of over- or underrepresentation to average of H3K4me3-peak coverage per gene class. Gene classes are colored as in (A). C) Similar to B), but zygotic genes were split in early activated genes (detectable transcription levels prior to dome, dark green) and late activated genes (light green). D) Percentage of overrepresentation of H3K4me3-peak coverage at maternally-provided genes which have stable (light blue), decreasing (blue) or increasing (dark blue) transcription levels.

H3K4me3 is continuously established over developmental time, even at genes that are not transcribed. Since not-expressed genes become H3K4me3-positive as well, H3K4me3 cannot merely be consequence of active transcription in the zebrafish embryo. These results corroborate with findings of previous studies (Lindeman et al., 2011; Vastenhouw et al., 2010).

Next I addressed whether a specific class of genes is overrepresented in H3K4me3-positive promoters. Therefore I tested whether the percentage of genes positive for H3K4me3 in each class is different to to total percentage per stage. I plotted percentage of over- or underrepresentation compared to average percentage for each gene class on 4 consecutive stages (Figure 5.1 B). This analysis shows that maternally provided

genes are overrepresented in H3K4me3-positive promoters, while not-expressed genes are underrepresented at each analyzed stage. The random outgroup as well as genes zygotic activated genes are neither over- nor underrepresented for H3K4me3-positive promoters. If H3K4me3 is thought to poise genes for activation, zygotic genes should be preferentially marked by H3K4me3. This could not be observed in these analysis, suggesting that transcriptional poising is not the dominating reason for H3K4me3 establishment.
Not-expressed genes are underrepresented for H3K4me3-positive promoters at each stage. Therefore, while not being the only cause for H3K4me3 establishment, active transcription could contribute to H3K4me3 establishment. To address this hypothesis, I subdivided the class of zygotic activated genes. Because these genes have no transcription levels prior to genome activation, their timing of first transcription activation can be assessed based on transcript levels. I subdivided purely zygotic genes in:

- genes that become transcriptionally activate until dome stage and
- genes that have no transcripts until dome, but become transcriptionally activate during embryonic development until 5dpf.

Both subcategories of zygotically activated genes gained in number of H3K4me3 positive genes. However, their enrichment for H3K4me3-positive genes is different (Figure 5.1 C). If other features than transcriptional activity would allow for enrichment of H3K4me3 in specific classes, no change should be observed over developmental stages. While genes activated late after genome activation are not enriched for H3K4me3 at any stage, early activated genes change their likelihood of being marked with H3K4me3 over developmental time. While first their likelihood of being marked with H3K4me3 is not different from average, at dome stage they are overrepresented in H3K4me3 (Figure 5.1). These genes were chosen to be activated until dome, thus more and more genes of this class become transcriptionally active over developmental time. These results suggest that active transcription of these genes leads to their higher chance of gaining H3K4me3 over developmental time.

Interestingly, the class of genes that is strongest overrepresented in H3K4me3-positive genes at all stages are maternally provided genes. This overrepresentation could be explained by a higher fraction of actively transcribed genes among the class of maternal genes. To determine whether transcription activity contributes to the overrepresentation

of maternal genes, I further subclassified maternally provided genes into the following classes:

- genes that have stable transcription levels over time,
- genes that show increasing RNA levels after genome activation, and
- genes that show decreasing RNA levels.

It is likely that genes that show decreasing RNA levels are less transcriptional activity than genes with increasing RNA levels. Thus, this classification allows me to decipher whether active transcription contributes to the likelihood of gaining H3K4me3 at maternal genes.

Interestingly, when analyzing the percentage of H3K4me3-positive promoters in these subclasses, all maternally provided genes show a higher probability of H3K4me3 appearance at their promoters than expected from average percentage (Figure 5.1 C), irrespective of their transcriptional activity. No difference between the subgroups could be observed at any stage. This suggests that the higher fraction of H3K4me3-positive promoters in maternally provided genes is not due to zygotic transcription, but is caused by other mechanisms.

Taken together, these findings suggest that there are two modes of H3K4me3 establishment during early embryonic development:

1. H3K4me3 is established with active transcription. Not-transcribed genes are underrepresented in H3K4me3-positive promoters compared to all other classes. Furthermore, while zygotic genes that become activated in late developmental stages are underrepresented for H3K4me3, zygotic genes genes activated during early stages switch from being not enriched to being overrepresented for H3K4me3-positive promoters. This suggests, that active transcription guides H3K4me3 establishment to promoters that otherwise would not gain H3K4me3. This is in line with a Set1-mediated recruitment of H3K4me3 establishment (Krogan et al., 2003; Ng et al., 2003).

2. H3K4me3 is established independent of transcription during embryonic development. This becomes evident from H3K4me3 being present at genes that are not transcribed in zebrafish embryos. At dome stage, 5,300 of those genes (40% of all not-transcribed genes) are H3K4me3-positive. Furthermore, zygotic activated genes are marked by H3K4me3 prior to their activation. This is in line with previous findings that identified H3K4me3 at transcriptionally inactive loci in

zebrafish development (Lindeman et al., 2011; Vastenhouw et al., 2010; Zhang et al., 2014).

Because I could show that reduction of H3K4me3 levels does not change transcriptional output of affected genes during early embryonic development, it is possible that H3K4me3 establishment is a mere consequence of chromatin accessibility during early embryogenesis (Bogdanovic et al., 2011; Schneider et al., 2011; Wen et al., 2009). However, if H3K4me3 is not solely a cause of active transcription, nor needed for robust transcription activation, it is unclear why H3K4me3 is established to promoters during early embryonic development.

5.2 H3K4me3 is gained at CpG-rich elements

To understand, what makes promoters gain H3K4me3 independent of transcription, I analysed the promoter sequence of genes that are neither provided maternally, nor expressed during embryonic development until day 5. According to embryonic transcriptome data (Pauli et al., 2012), around 13,000 genes have no detectable transcript levels during embryonic development. Of these, 5,300 have an H3K4me3 peak in close vicinity to their transcription start sites at dome stage. I analyzed the GC content as well as the CpG density of the DNA sequence of TSS +/-2.5kb according to Takai and Jones 2002. I plotted the distribution of promoters over the range of GC-content and CpG observed-to-expected ratio (termed 'CpG density'). When comparing H3K4me3-positive to H3K4me3-negative non-transcribed genes, I found that both groups of genes show the same average GC-content (Figure 5.2 A), suggesting that a bias towards specific nucleotide content is not predictive for H3K4me3 establishment.
CpG density shows two population of promoters: genes with low CpG content (below genome-wide average CpG density of 0.7) and genes with high CpG content (above genome-wide average). Interestingly, the majority of H3K4me3-positive promoters have a high CpG-content, while H3K4me3-negative are more likely to have low CpG-content (Figure 5.2 B).
To determine whether this preference for CpG content - but not for GC content - is not only restricted to non-transcribed genes but also reflected in H3K4me3 enrichment genome-wide, I analyzed H3K4me3 enrichment by metagene plots. I divided all genes in low-GC content and high GC-content (below and above average promoter GC content of 0.35, respectively), or low and high CpG density (below or above average

CpG density of 0.7, respectively). I plotted the average density of reads identified in H3K4me3 ChIP-Seq experiments at dome stage around TSS for each of these classes separately. Low GC-content promoters showed lower levels of H3K4me3 when compared to high GC-content promoters (Figure 5.2 C). Because I have shown before that GC-content does not distinguish between the number of H3K4me3-positive and -negative promoters, it is possible that H3K4me3 levels at individual promoters with high GC-content are higher than on low GC-content promoters.

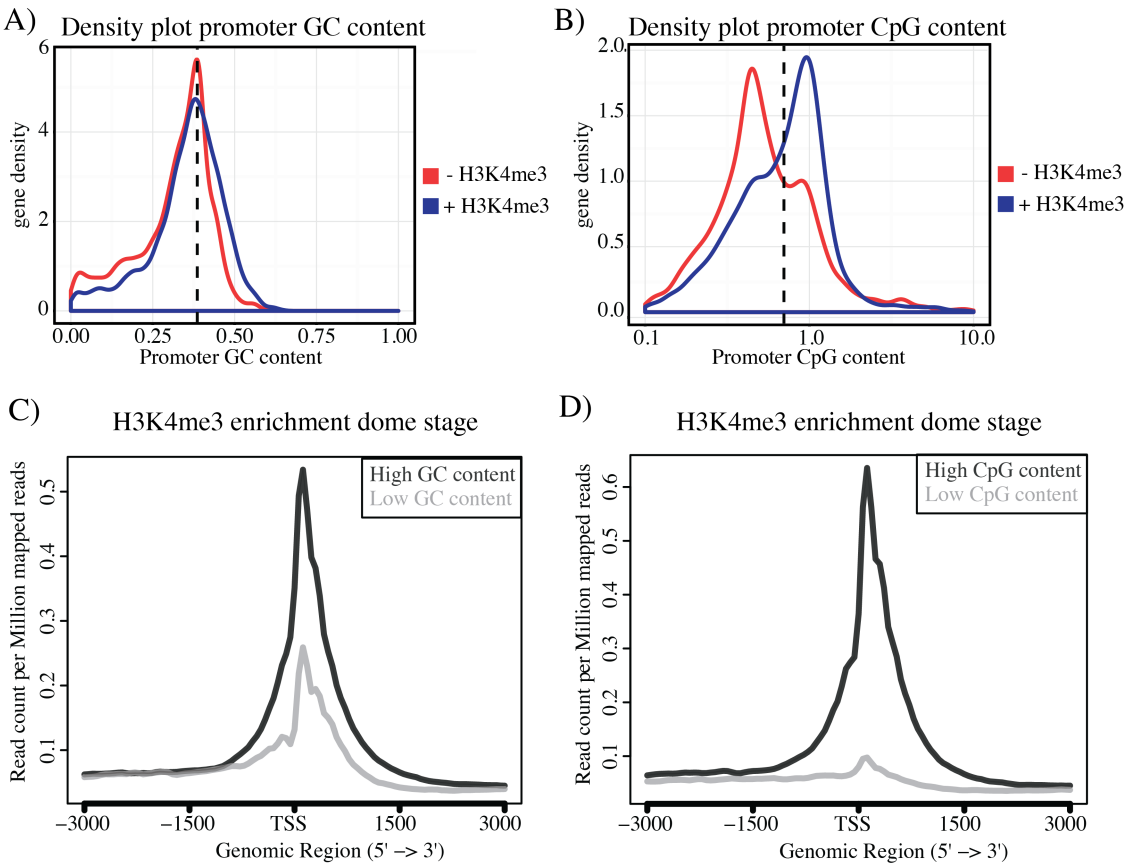


Figure 5.2: H3K4me3 is enriched at CpG-rich promoters A,B) Smoothened histograms representing gene density of non-transcribed genes either negative (red) or positive (blue) for H3K4me3 enrichment at dome stage. Depicted is gene density according to GC content (A) or CpG content (B) at the respective promoter. Dashed line indicates average GC- or CpG content. C,D) Metagene profile over all annotated promoters with either high (black) or low (grey) GC content (C) or CpG content (D) at dome stage.

When comparing average H3K4me3 density on promoters with high CpG density to low CpG density, a striking difference could be observed. At low CpG density promoters, H3K4me3 enrichment is almost undetectable compared to genes with high CpG density promoters (Figure 5.2 D). Together with the observation that not-transcribed genes that are positive for H3K4me3 have in general higher CpG density, this suggests that CpG content might be involved in transcription-independent H3K4me3 establishment.

5.3 H2A.z marks overlaps with H3K4me3 at promoters of non-transcribed genes

H3K4me3 establishment could also be stimulated by other features of promoter chromatin. One candidate that could allow for H3K4me3 establishment is the histone variant H2A.z. This histone variant replaces the canonical H2A at specific genomic features by active nucleosome exchange (Kobor et al., 2004; Mizuguchi et al., 2004). H2A.z-containing nucleosomes were found to be less stable compared to canonical nucleosomes, leading to higher nucleosome turnover and increased DNA accessibility at promoter regions (Jin and Felsenfeld, 2007; McKittrick et al., 2004). To be able to compare H2A.z enrichment to H3K4me3, I performed a ChIP-Seq experiment for H2A.z at dome stage, when H3K4me3 is robustly established. Manual inspection of the genome-wide distribution of H2A.z revealed that H2A.z is enriched at TSS and overlaps with H3K4me3 enrichment at dome stage (Figure 5.3 A). To address whether H2A.z is generally enriched at TSS, I plotted the read density of ChIP-Seq samples over annotated TSS. Indeed, H2A.z is highly enriched within +/-1.5kb of TSS.
I have observed that H2A.z and H3K4me3 are both enriched at TSS of genes during embryonic development, and manual inspection showed that H2A.z and H3K4me3 enrichment overlap at individual loci. Thus, the presence of H2A.z might be predictive for H3K4me3 enrichment. To address this hypothesis, I quantified the level of H2A.z at individual promoters based on an analysis published in Zhang et al. 2014. In brief, all reads that fall within a region of +/-2.5kb around individual TSS were counted, and normalized by library size. These values were used separate genes in a groups of high and low H2A.z levels (above or below 2fold of average H2A.z enrichment). Next I plotted the H3K4me3 read density over TSS for both groups separately. Strikingly, H3K4me3 is significantly more enriched at promoters that have high H2A.z levels. Thus, high levels of H2A.z overlap with, and might influence, H3K4me3 establishment.

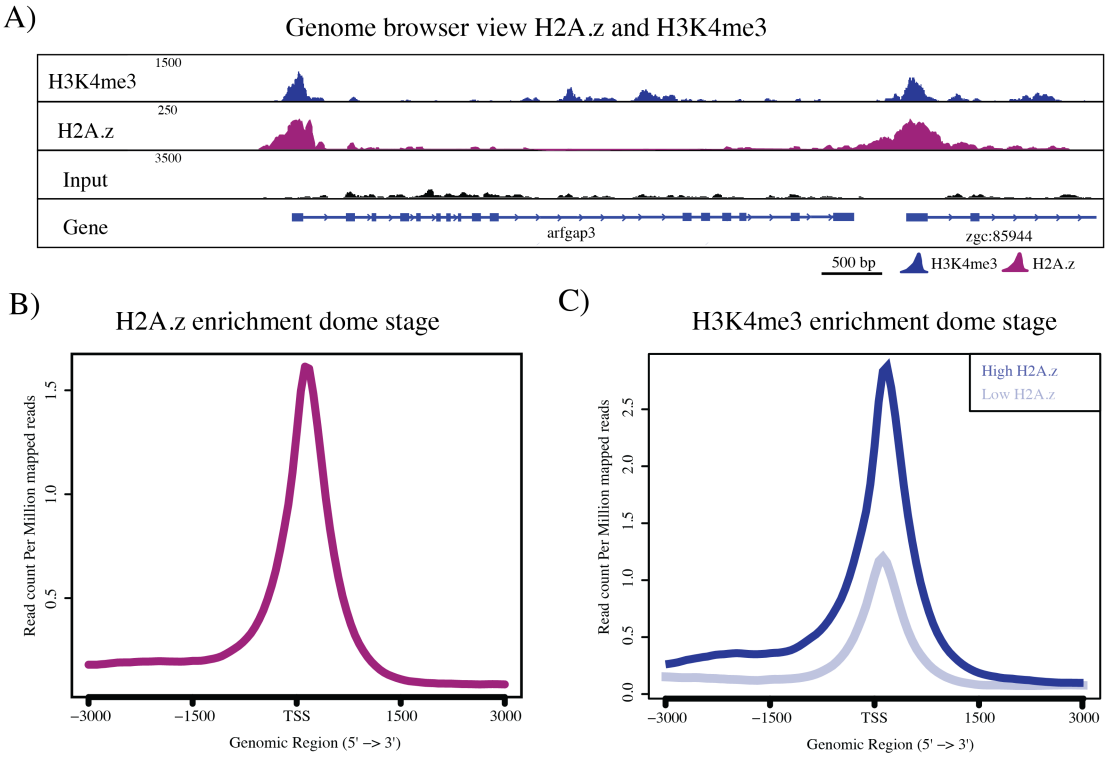


Figure 5.3: H2A.z is enriched at promoter regions and overlaps with H3K4me3 enrichment A) Representative genomic region showing H3K4me3 enrichment (blue) and H2A.z enrichment (violet) overlap at gene promoters. Input chromatin depicted in black. B) Metagene profile over annotated TSS plotting the read density of H2A.z ChIP-Seq reads at dome stage. C) Metagene profile for H3K4me3 enrichment over promoters that are either high (blue) or low (light blue) in H2A.z enrichment.

5.4 High CpG density and H2A.z enrichment are predictive for H3K4me3 establishment

Both high CpG content and high H2A.z levels overlapped with stronger H3K4me3 enrichment. If CpG content and H2A.z enrichment is causal for H3K4me3 establishment, I expect that genes marked by H3K4me3 without transcriptional activity are either rich in CpG content, or enriched for H2A.z, or both.

To test this hypothesis, I grouped genes according to their transcriptional activity. Classification based on RNA-Seq data does not allow to confidently state whether maternally provided genes are actively transcribed, as transcription might contribute to stable RNA levels. To allow for classification of maternally provided genes as well, I used

the H3K36me3 enrichment as a marker for transcriptional activity. Previous studies have shown that H3K36me3 enrichment within the gene body correlates with active transcription RNA-Polymerase II, and thus is an experimental marker for ongoing transcription (Zhang et al., 2014). I used H3K36me3 data obtained for zebrafish dome stage embryos to identify genes without active transcription according to Zhang et al. 2014. Of these genes, I analyzed H3K4me3-positive and -negative loci for their CpG content and H2A.z content. In this analysis I considered promoter regions to have high H2A.z levels, if they have 2fold higher than average H2A.z enrichment, and high CpG content if CpG content was higher than average (0.7 observed-to-expected ratio) in their promoter region.

Most of the 9,147 H3K4me3-positive yet transcriptionally inactive genes are both high

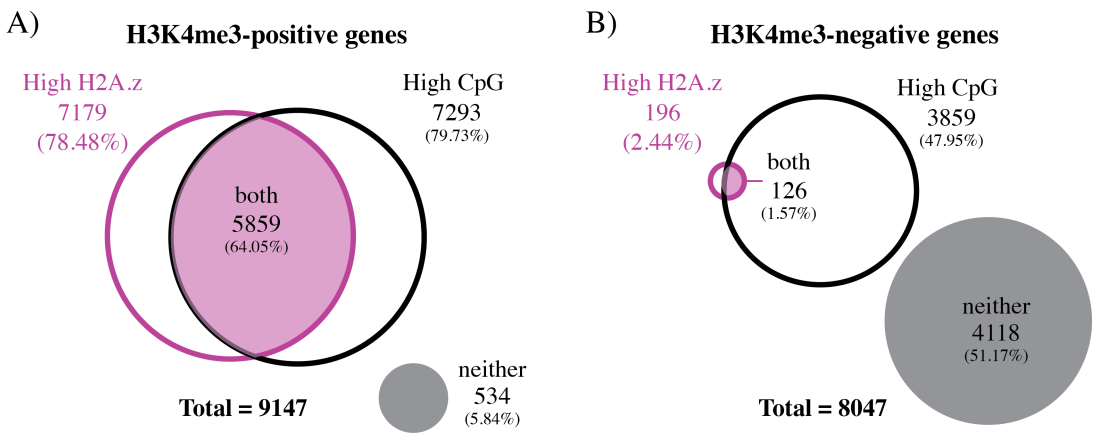


Figure 5.4: H2A.z and CpG density predict H3K4me3 establishment. A) Venn diagram displaying the fraction of genes that are transcriptionally inactive and enriched for H3K4me3 at dome stage based on previously published data (Zhang et al., 2014). B) Venn diagram displaying the fraction of genes that are transcriptionally inactive and are H3K4me3-negative at dome stage based on previously published data (Zhang et al., 2014). CpG content and H2A.z enrichment were considered high, if the values were above the average or higher than 2fold average of all promoter values, respectively.

in H2A.z enrichment and CpG content (64%). 78% of H3K4me3-positive promoters of non-transcribed genes have CpG-levels above average, and 74% show high H2A.z levels. Less than 6% of H3K4me3-positive, non-transcribed genes are neither marked by H2A.z nor contain high CpG content in their promoters. On the contrary, in the analyzed 8047 non-transcribed H3K4me3-negative genes, this fraction is larger than 50% (Figure 5.4). While around 48% of those genes have promoters with high CpG content, only 2.5% are marked with H2A.z, and only 1.5% (126 genes) are marked

with both. Together, these results suggest that H2A.z and CpG content cooperate to allow for H3K4me3 establishment independent of transcription.

5.5 Maternally provided genes are enriched for H2A.z and CpG content

I have shown before that H3K4me3-positive promoters are overrepresented in maternally provided genes, irrespective of their transcription status (Figure 5.1). I could further show that H3K4me3-positive promoters of non-transcribed genes have high CpG content and high H2A.z enrichment (Figure 5.4). If CpG content and H2A.z enrichment are predictive of H3K4me3 establishment, these features should be overrepresented in maternally provided genes as well.

To test this hypothesis, I compared maternally provided genes against all genes that

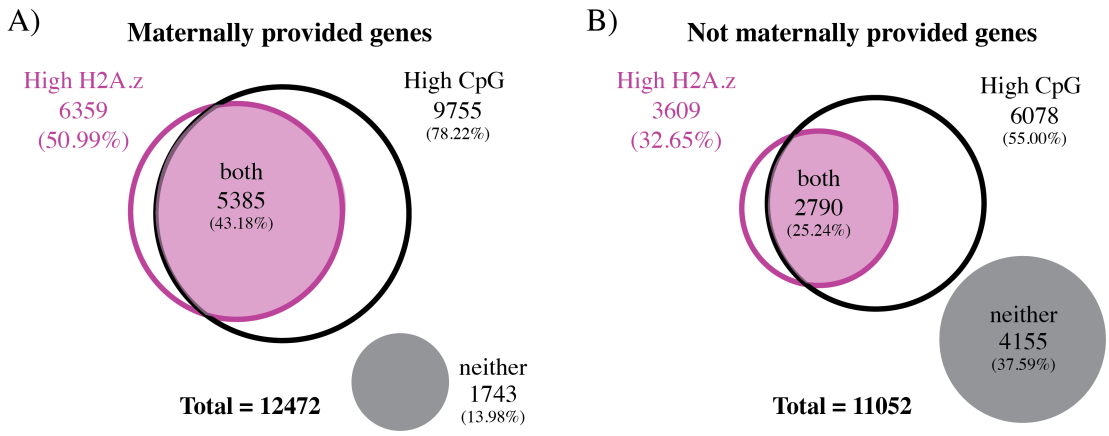


Figure 5.5: H2A.z and CpG density are overrepresented in maternal genes Venn diagram displaying the fraction of genes that are high in CpG content or H2A.z enrichment of maternally provided (A) and not maternally provided genes (B) based on previously published transcription data (Pauli et al., 2012). CpG content and H2A.z enrichment were considered high, if the values were above the average or higher than 2fold average of all promoter values, respectively.

are not maternally provided for CpG content and H2A.z enrichment as described above (section 5.3). I did not consider transcription activity, as this seems not to influence the overrepresentation of H3K4me3 promoters in maternal genes. 78% of maternally provided genes have a high CpG content, and 51% are enriched for H2A.z, with an overlap of 43%. While this is less than observed for H3K4me3-positive yet untranscribed genes (Figure 5.4), this could be explained by the lower fraction of H3K4me3-

positive promoters in maternally provided genes (84% to 100%, respectively). 14% of maternally provided genes have neither of the two chromatin states. When analyzing not maternally provided genes, both the number of promoters with high CpG content (55%) and high H2A.z enrichment (33%) are lower, and only 25% of these promoters have both features. 37% of not maternally provided genes have neither of the features. While the difference between these fractions is not drastic, it very well reflects the difference in H3K4me3 enrichment found at dome. 86% of maternal genes were found to be H3K4me3-positive, while 14% are not marked, and only 48% of not maternally provided genes are positive for H3K4me3 (Figure 5.1).
While these results are based on correlation only, they suggest that high density of CpG and high levels of H2A.z are chromatin features that might explain the increased amount of H3K4me3-positive promoters in maternally provided genes.

5.6 Conclusion

In conclusion, these results show that the number of genes marked with H3K4me3 increases over developmental time. Increased H3K4me3 establishment is found on both, actively transcribed as well as non-transcribed genes. Genes that become activated during embryonic development are not overrepresented for H3K4me3-positive promoters, but become increasingly marked by this modification when they become transcriptionally activate, suggesting that transcription contributes to embryonic H3K4me3 establishment. Interestingly, maternally provided genes are overrepresented for H3K4me3-positive promoters independent of their transcription status, suggesting for transcription-independent H3K4me3 establishment at these genes specifically.
By analyzing chromatin features that correlate with H3K4me3, I could further show that genes marked by H3K4me3 independent of transcription are enriched for CpG density (but not GC-content), and H2A.z enrichment. Both H2A.z enrichment and high CpG content are overrepresented in maternally provided genes as well, suggesting that these features together result in a higher fraction of H3K4me3-positive genes in this subgroup of genes. These results indicate that H2A.z enrichment and high CpG content represent features of embryonic chromatin that might be involved in H3K4me3 establishment independent of transcription.

6 Discussion

Early zebrafish development relies on robust and timely activation of specific genes (Kane et al., 1996; Newport and Kirschner, 1982). While progress has been made to identify key regulators of genome activation (Chan et al., 2009; Lee et al., 2013b; Leichsenring et al., 2013; Liang et al., 2008), it remains elusive which general mechanisms further contribute to activation of transcription during embryonic development. It has been proposed that chromatin plays a crucial role in preparing the embryonic transcriptome for robust activation (Kane and Kimmel, 1993; Prioleau et al., 1994). In this context, the histone modification H3K4me3 might be a crucial player for the onset of transcription, as it is thought to prepare key developmental genes for robust activation during embryogenesis (Bernstein et al., 2006; Lindeman et al., 2011; Vastenhouw et al., 2010).

In this study, I investigated whether H3K4me3 establishment is involved in zebrafish zygotic genome activation. Mutation of the histone methyltransferases Mll1 and Mll2 did not result in loss of H3K4me3 during early embryogenesis, possibly due to compensation of their loss by other H3K4-specific Histone methyltransferases (section 6.1). Instead I was able to achieve robust reduction of H3K4me3 levels in early zebrafish embryos by introduction of H3K4-to-M constructs (K4M). K4M seems to act as dominant-negative substrate for H3K4me3 establishment. I analyzed changes in H3K4me3 levels genome-wide by quantitative ChIP-Seq experiments, which showed that the majority of gene promoters had reduced H3K4me3 levels upon introduction of K4M (section 6.2). Reduction of promoter-associated H3K4me3, however, did not result in significant changes of transcription activation. Neither early zygotic activated genes, nor genes that become transcriptionally active during early gastrulation stages were affected by H3K4me3 reduction (section 6.2). If not involved in transcription activation, it is puzzling why H3K4me3 is established at more than 60% of all promoters at early developmental stages. Future studies have to address whether H3K4me3 establishment is a mere consequence of a permissive chromatin state during early developmental stages (section 6.3), and whether it might be important for later developmental stages, where it might be important for transcription regulation during lineage commitment in

a generally repressive chromatin environment (section 6.4). Finally I analyzed, which genomic features allow for H3K4me3 establishment independent of transcription. I found that H3K4me3 is preferentially established at promoters with high CpG content and strong H2A.z enrichment, suggesting that these features might favor H3K4me3 establishment during early developmental stages (section 6.5). Together, these findings suggest that H3K4me3 is established at CpG-rich promoters during early embryonic stages, but does not influence early transcription activation in zebrafish embryos.

6.1 Neither Mll1 nor Mll2 are the main histone methyltransferase for H3K4me3 establishment in early zebrafish development

During my thesis I aimed to interfere with H3K4me3 establishment by genetic ablation of H3K4-specific histone methyltransferases. I generated zebrafish mutants for the histone methyltransferases *mll1*, *mll2* and the HMT complex protein *ash2l*. However, none of the generated maternal-zygotic mutants showed global reduction of H3K4me3 levels, even though increased mortality during larval development was observed in *mll1* and *mll2* mutants (section 3.2).

Mll1 and Mll2 share components for their enzymatic active protein complex *in vivo* (Dou et al., 2006; Miller et al., 2001; Yokoyama et al., 2004). Additionally, other H3K4-specific HMTs exist that are able to establish H3K4me3 (section 1.3, Shilatifard 2012). Therefore it is possible that H3K4me3 establishment by Mll1 and Mll2 can be compensated for by other HMTs. Indeed it was shown that interference with Mll1 resulted in little changes of H3K4me3 levels at promoters (Guenther et al., 2005; Terranova et al., 2006; Wang et al., 2009, 2012). Furthermore it was shown that loss of Mll1 can be compensated for by the presence of other HMTs in mouse embryonic stem cells (Denissov et al., 2014). However, the same authors found that Mll2 has a unique role in establishing H3K4me3 at bivalent domains (Denissov et al., 2014). Therefore, while Mll2 seems not to be involved in bulk H3K4me3 establishment (Hu et al., 2013b; Lubitz et al., 2007), it might be important for transcription regulation of genes involved in further development. The here obtained data suggests that in zebrafish, *mll2* mutation does not result in embryonic lethality and failure of germline development. Furthermore, Mll2 seems not to be involved in global H3K4me3 establishment during embryonic development (Figure 3.6). However, the possible effect on specific genomic features, such as bivalent domains, was not addressed in this study. Further ChIP-Seq experiments addressing H3K4me3 distribution in *mll2* mutant zebrafish embryos will

show whether this unique role of Mll2 is also present in zebrafish development.
Despite little effects on global H3K4me3 levels, both *mll1* and *mll2* mutation resulted in increased mortality during larval stages. It is possible that molecular mechanisms other than H3K4me3 establishment are causal for this phenotype. In plants, most phenotypes caused by HMT mutations are associated with the absence of the respective protein, but not its role in H3K4me3 establishment (Fromm and Avramova, 2014). Furthermore, knock-out studies in mice suggested that Mll1 has specific molecular functions independent of its role in H3K4me3 establishment, as *mll1* mutations with a deletion of the SET-domain have a different phenotype than full *mll1* knock-out (Jude et al., 2007; Mishra et al., 2014; Terranova et al., 2006). Therefore it is possible that the observed phenotypes for *mll1* and *mll2* mutants are not driven by changes in H3K4me3 levels, but by the absence of Mll1 and Mll2 proteins.
While mutation of *mll1* or *mll2* alone did not have an effect on H3K4me3 establishment, double-mutants could address whether Mll1 and Mll2 together are the main actors to establish H3K4me3 in early embryogenesis. Generation of *mll1-mll2* double-mutants might allow to address this question. Another strategy to test whether Mll1/Mll2 establish bulk H3K4me3 is to genetically interfere with *menin1*, which is a shared essential complex component specific to Mll1/Mll2 (Agarwal and Jothi, 2012; Yokoyama et al., 2004). These genetic interference strategies could be used to study the effect of locus-specific H3K4me3 reduction during later developmental stages in a possible more repressive chromatin environment (section 6.4).

6.2 H3K4me3 reduction does not affect transcription initiation during genome activation

Global reduction of H3K4me3 could be achieved by overexpression of H3K4-to-M (K4M) histone proteins. Upon H3K4me3 reduction, I analyzed transcription activation of early activated genes. Because H3K4me3 has been shown to mark the majority of gene promoters (Lindeman et al., 2011; Vastenhouw et al., 2010) and to facilitate transcription factor recruitment (Flanagan et al., 2005; Vermeulen et al., 2007), I expected to observe a decreased transcriptional output of the majority of activated genes. In contrast, no transcriptional changes could be observed. During the course of this study I analyzed 98 early zygotic genes (without maternal contribution) over at least four consecutive stages by two different methods (section 4.6). I further analyzed 17 genes which are activated during gastrulation stages. For none of these genes, general tran-

script levels were changed upon H3K4me3 interference, and their onset of transcription was not different when compared to samples without H3K4me3 interference. These experiments suggest that H3K4me3 reduction does not affect transcription activation of genes during early stages of zebrafish embryonic development.

The here described method results in a reduction, but not a complete loss of H3K4me3. Furthermore, assaying genome-wide H3K4me3 levels relied on ChIP-Seq experiments, which represents an average picture over multiple embroys. Therefore it cannot be excluded that individual embryos, and even cells, might have different levels of H3K4me3 left at their promoters. Thus it cannot be excluded that remaining levels of H3K4me3 are sufficient to allow for transcription activation.

It has been shown previously that H3K4me3 might have little effects on transcription activation. Interference with H3K4me3 establishment upon rapid gene induction triggered by stimuli showed that only a subset of genes relies on H3K4me3 establishment for their activation (Clouaire et al., 2014). In mouse neural progenitor cells, knockdown of Wdr5, an essential COMPASS component (Wysocka et al., 2005), resulted in transcriptional reduction of only a small subset of genes (Benayoun et al., 2014). Furthermore, clonal depletion of histone lysine 4 in all histone genes resulted in no apparent changes of genes involved in *Drosophila* wing disc development (Hödl and Basler, 2012). These experiments have shown that H3K4me3 seems not to be important for transcriptional activity in differentiated systems. The here obtained results complement these observations with evidence that H3K4me3 seems not to be involved in general transcription activation of previously inactive genes in a developing embryo. However, it cannot be excluded that the here analyzed early activated genes are insensitive to H3K4me3 interference. While these findings suggest that H3K4me3 is not a general regulator of transcription activation, further experiments should focus on whole transcriptome analysis to address, whether a specific subset of genes might be sensitive to loss of H3K4me3.

H3K4me3 reduction did not result in transcriptional changes during early embryonic development (section 4.6). This raises the question why so many genes acquire H3K4me3 early during genome activation (section 5.1). In the next sections I will discuss potential explanations for why H3K4me3 is established during genome activation (section 6.3), and what its potential function might be (section 6.4).

6.3 The timing of H3K4me3 establishment might be determined by a permissive chromatin state

The establishment of H3K4me3 coincides with genome activation. Importantly, both actively transcribed as well as not transcribed genes gain H3K4me3 at their promoters (section 5.1). Genes that are transcribed most likely acquire H3K4me3 by transcription-coupled recruitment of Set1 COMPASS complexes (section 1.3). Indeed I could show that early zygotic genes become progressively overrepresented for H3K4me3-positive promoters when they are transcribed, while genes activated later than the here observed developmental stages for H3K4me3 establishment are not overrepresented for H3K4me3-positive promoters (Figure 5.1). However, because also transcriptionally inactive genes acquire H3K4me3, transcription-coupled H3K4me3 establishment cannot be the only mechanism.

Transcripts coding for all essential COMPASS components are present in zebrafish embryos before genome activation (Pauli et al., 2012; Sun et al., 2008). This raises the question why H3K4me3 cannot be established before genome activation. Recent observations in our laboratory suggest that histones act as general inhibitors for chromatin binding (Joseph et al., in preparation). Histones are thought to be present in excess during stages prior to genome activation. The excess of histones might compete with transcription factors and other proteins for binding to DNA. In this model, regulatory chromatin modifications might only be established coinciding with genome activation, when the excess of histones compared to zygotic DNA has decreased. Future experiments could address whether the increase of histone content by micromanipulation results in a delay of the establishment of H3K4me3.

These observations might explain why H3K4me3 is not established earlier, but do not address why H3K4me3 is established early, even though it is not involved in general transcription activation. It is believed that the genome of early embryos is in a permissive state, which facilitates DNA binding of regulatory proteins during early development (Bogdanovic et al., 2011; Schneider et al., 2011; Wen et al., 2009). Many post-translational modifications that allow for heterochromatin establishment are only established after genome activation (Fadloun et al., 2013; Lindeman et al., 2011; Santenard et al., 2010). Additionally, an embryo-specific linker histone H1M is present during development, that might allow for a less compacted chromatin state (Gao et al., 2004; Saeki et al., 2005). This linker histone is progressively lost and replaced by the more repressive somatic linker histone H1S (Steinbach et al., 1997; Yang et al., 2013).

Furthermore it was found that the regulatory potential of many epigenetic modifications, as for example DNA methylation, is temporarily muted during early embryonic stages, and complete transcription regulation potential is only acquired in later developmental stages (Schneider et al., 2011). Together these findings suggest that after genome activation, when competition with histones is overcome, embryonic chromatin is in a permissive state that favors DNA interaction of regulatory proteins. In the course of further embryonic development, the chromatin might gradually become more repressive, and transcription activation increasingly relies on transcription-favoring chromatin modifications. In this model (depicted in Figure 6.1), H3K4me3 establishment cannot be established at transcriptionally inactive genes during later developmental stages, when the chromatin is in a repressive state. Instead, H3K4me3 is established during early stages, when a permissive chromatin state allows for the recruitment of H3K4-specific HMTs to transcriptionally inactive loci. Therefore, H3K4me3 is a consequence of a permissive chromatin state during genome activation, but not involved in transcription activation at ZGA.

To test this hypothesis, transient interference by H3K4-specific inhibitors might be a valuable strategy. Inhibition of HMTs during genome activation, and a following depletion of the inhibitor would show, whether H3K4me3 can be established during later developmental stages. So far, no H3K4-specific HMT inhibitor has been reported, however recent advances in deciphhering the protein complex structure might allow the design of potent inhibitors in the near future (Mersman et al., 2012; Senisterra et al., 2012). Another approach to achieve temporal control might be to genetically modify essential HMT complex components to allow for their temporal mislocalization upon stimulus. By addition of specific dimerization domains to HMT complex proteins, and overexpression of a membrane-targeting domain, mislocalization could be achieved either by light-induced dimerization (Simmich et al., 2012) or by small-molecule treatment (Hathaway et al., 2012). Due to the rapid cell cycle within the zebrafish embryos, nuclear proteins are regularly distributed within the cytoplasm, thus making this strategy feasible. These experiments have the potential to reveal whether H3K4me3 is a consequence of a permissive chromatin state during genome activation.

6.4 H3K4me3 potentially gains importance during later developmental stages

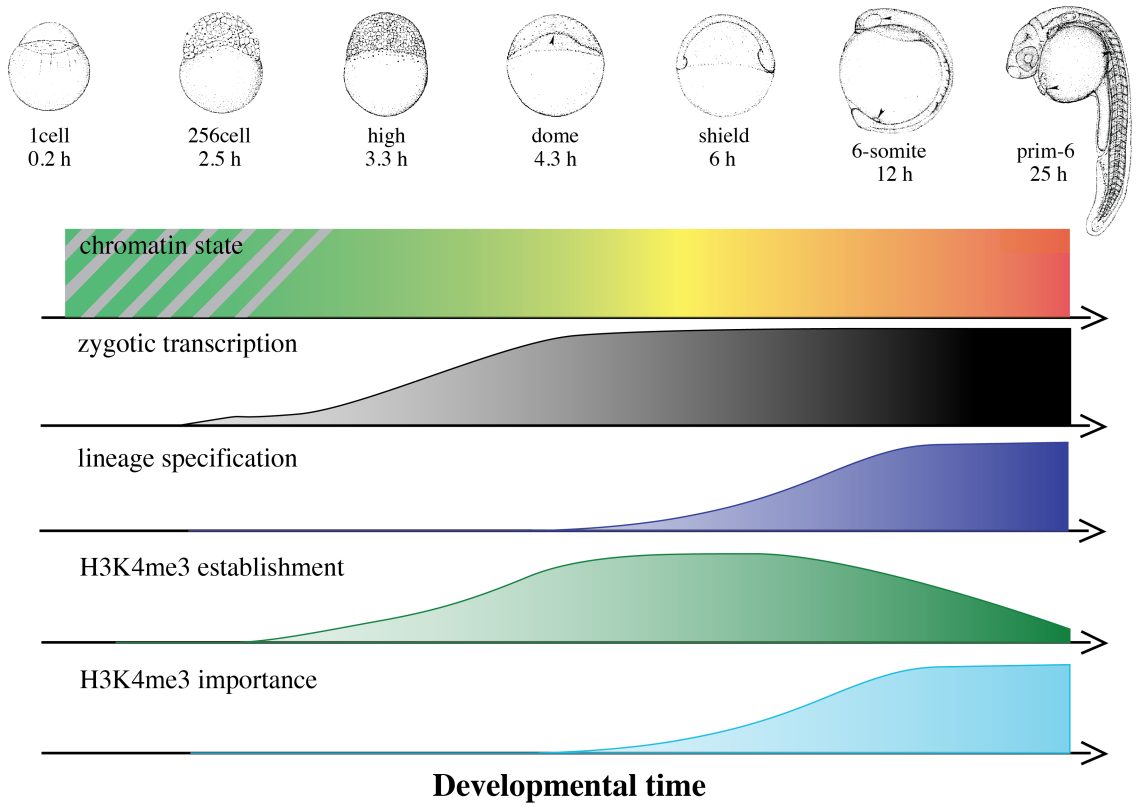


Figure 6.1: Model for the role of H3K4me3 in embryonic development Zebrafish zygotic transcription commences during early developmental stages (black), after general repression by histones (grey) is overcome. Zygotic transcripts drive differentiation, which begins during early gastrulation stages and leads to decision for specific cell fates during further embryonic development (dark blue). H3K4me3 establishment (green) coincides zygotic genome activation at CpG-rich promoters in a generally permissive chromatin environment. During early stages, reduction of H3K4me3 has no effect on transcription activation. However, promoter-associated H3K4me3 might gain in importance (light blue) by regulating cell lineage specification during later stages, when a repressive chromatin state prevails. Stages of embryonic development adapted from Kimmel et al. 1995.

The here obtained data suggests that H3K4me3 has no effect on genome activation. This is seemingly in contrast to the current understanding, that H3K4me3 facilitates transcription (Flanagan et al., 2005; Vermeulen et al., 2007), and is important for activation of developmental regulators (Bernstein et al., 2006; Lindeman et al., 2011; Vastenhouw

et al., 2010). It may be possible, however, that H3K4me3 is not required for robust transcription activation in the permissive chromatin environment of the early embryo, and only plays a role in later developmental stages, when the chromatin environment becomes increasingly repressive (Bogdanovic et al., 2011; Schneider et al., 2011; Wen et al., 2009). Evidence for this hypothesis comes from the observation that loss of H3K4me3 by replacement of wildtype histones with methylation-defective H3.3 in Xenopus eggs resulted in loss of transcription activation of genes involved in muscle differentiation (Ng and Gurdon, 2008).

To test whether H3K4me3 indeed plays a role in transcription later during development, one could monitor the activation of lineage-specific transcription profiles upon H3K4me3 interference during stages later than gastrulation. However, the here implementented approach to reduce H3K4me3 did not allow for studying transcription regulation in later developmental stages, because both overexpression of H3K4-to-M and control H3HA constructs was detrimental to embryonic development later than nine hours post fertilization. Therefore it is necessary to further elaborate on tools to interfere with H3K4me3. Reduction of H3K4me3 on a subset of genes would reduce possible deleterious effects to embryonic development, thus allowing to address the question whether activation of these genes is affected later during embryonic development. This may be achieved by genetic interference with COMPASS components that are involved in targeting H3K4me3 activity to specific genes, yet are not necessary for global H3K4me3 establishment (discussed in section 6.1). Further approaches could be based on transient H3K4me3 interference as discussed above (section 6.3). Together, these experiments could address whether H3K4me3 is established due to a specific window of opportunity, and whether it is involved in transcription activation later during development.

6.5 CpG-content and H2A.z enrichment might be predictive for H3K4me3 establishment during genome activation

In the previous section I hypothesized that H3K4me3 establishment is a consequence of the permissive chromatin state that allows for the recruitment of HMT complexes. In this study I provided evidence that CpG density, but not general GC content, predicts H3K4me3 establishment. Maternally provided genes have more CpG-rich promoters than other classes, which might explain their higher-than-average fraction of H3K4me3-positive promoters (Figure 5.5). Furthermore, genes that are not transcriptionally active

have to recruit H3K4me3 by transcription-independent mechanisms. I could show that non-transcribed, yet H3K4me3-positive genes are generally CpG-rich as well (Figure 5.4). These observations are in line with studies of genome-wide H3K4me3 distribution in different cell lines, where H3K4me3 was found to be enriched at CpG-rich DNA (Mikkelsen et al., 2007). Furthermore, experiments in mouse embryonic stem cells found that introduction of synthetic CpG-rich DNA elements resulted in *de novo* establishment of H3K4me3 and H3K27me3 at sites of integration (Wachter et al., 2014).

The number of CpG-dense promoters alone could not explain the large amount of H3K4me3-positive promoters in zebrafish embryos. Another candidate that could facilitate H3K4me3 establishment is the histone variant H2A.z. This histone variant has been shown to be actively incorporated into chromatin independent of DNA replication (Goldman et al., 2010), and to increase histone turnover by nucleosome destabilization at promoter elements (Jin and Felsenfeld, 2007; Jin et al., 2009; Weber et al., 2014). In the context of zebrafish development, H2A.z incorporation and the resulting higher histone turnover might allow for recruitment of HMT complexes, as was shown in mouse embryonic stem cells (Creyghton et al., 2008; Hu et al., 2013c). I have analyzed the genome-wide distribution of H2A.z, and its correlation with H3K4me3. I found that early established H3K4me3 is enriched at promoters high in H2A.z levels. Even though solely based on correlation I hypothesize that presence of H2A.z at promoters might favor H3K4me3 establishment.

High CpG content and high H2A.z levels together might be sufficient to explain H3K4me3 overrepresentation in maternally provided genes (Figure 5.5), and 94% of non-transcribed, yet H3K4me3-positive genes have either of the two features, or both (Figure 5.4). Therefore, CpG content and H2A.z enrichment together could be predictive for H3K4me3 establishment early during embryonic development.

However, while almost all genes with high H2A.z levels are H3K4me3-positive, not all CpG-rich promoters are marked by H3K4me3 (48% of non-transcribed promoters have high CpG-levels, Figure 5.4). One explanation for this discrepancy might be DNA-methylation. DNA methylation occurs on CpG dinucleotides, and promoter-associated CpG islands are differentially methylated during zebrafish embryonic development (Andersen et al., 2012; McGaughey et al., 2014). Methylated CpG islands have been shown to prevent both H2A.z and H3K4me3 establishment at promoters in human cell lines (Okitsu and Hsieh, 2007; Zilberman et al., 2008). Future analysis have to address whether methylation of CpG islands prevents H3K4me3 establishment during zebrafish

development.
Both H2A.z and H3K4me3 have been shown to be enriched at CpG elements (Mikkelsen et al., 2007; Obri et al., 2014). However, since it has been shown that H2A.z incorporation is not facilitated by H3K27me3 establishment (Illingworth et al., 2012), it is possible that CpG-rich promoter elements allow for H2A.z incorporation prior to H3K4me3 establishment. Once established in chromatin, both H3K4me3 as well as H2A.z were found to interfere with DNA methylation in human cell lines (Balasubramanian et al., 2012; Conerly et al., 2010). Therefore, H3K4me3 and H2A.z might prevent erroneous DNA methylation.
Unifying these results, a permissive chromatin state during early embryogenesis could allow for H2A.z and H3K4me3 establishment at unmethylated CpG-islands (Figure 6.1), whereas DNA methylation prevents H3K4me3 establishment. The presence of H2A.z and H3K4me3 could prevent excessive DNA methylation during later stages, thus might prevent silencing of genes important for cell specification when a more repressive chromatin state is established.
Future studies might shed light onto the order of events in this process. For example, it would be interesting to study DNA methylation patterns upon H3K4me3 interference by K4M overexpression as performed in this study. These experiments would address whether early H3K4me3 establishment indeed restricts spreading of DNA methylaton, and thus prepares local chromatin for faithful transcription activation during later developmental stages.
Furthermore, in the light of recent developments of gene editing (Hoshijima et al., 2016; Hwang et al., 2013) it might be possible to change specific promoter sequences by scrambling of CpG dinucleotides without affecting GC-content. Since GC-content is less predictive for H3K4me3 enrichment, this strategy might allow to dissect, which sequence preference allows for H3K4me3 establishment. Furthermore one could interfere with H2A.z establishment. Since H2A.z is actively incorporated into chromatin, interfering with H2A.z-specific chromatin remodeling by genetic ablation could help to understand the role of H2A.z in establishing an open chromatin conformation during early zebrafish embryogenesis. Alternatively, transient overexpression of ANP32E could be used to specifically evict H2A.z from chromatin during zebrafish development. These experiments would unravel whether CpG content or H2A.z enrichment are essential for H3K4me3 establishment during genome activation (Mao et al., 2014; Obri et al., 2014). More importantly, if this addition of CpG-rich elements in zebrafish embryos can be performed with temporal control - either by specific CRISPR localization or

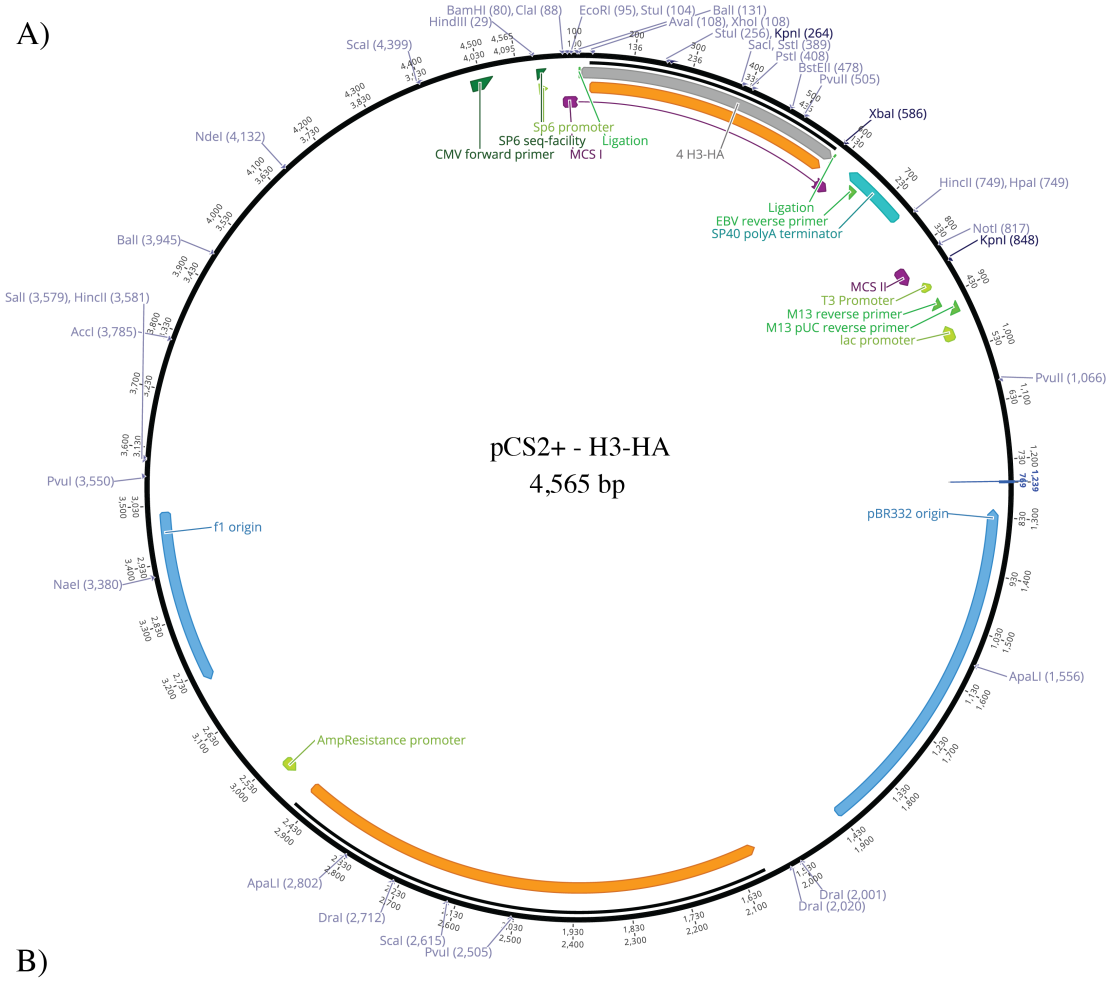
guide-RNA expression - this experimental set-up would unravel whether H3K4me3 establishment depends on a window of opportunity during zebrafish embryonic development. Together, these experiments would address the functional connection between DNA methylation and histone modifications, and their cooperative role in transcription regulation during lineage commitment.

6.6 Conclusion

In this study I could show that H3K4me3 reduction does not have an influence on zygotic transcription activation. H3K4me3 is preferentially established at CpG-rich promoters and correlates with H2A.z enrichment. Based on the here presented results it seems likely that early H3K4me3 establishment at CpG-rich, H2A.z-enriched promoters is a consequence of a generally permissive chromatin state during early embryogenesis. While H3K4me3 seems not to be involved in zygotic genome activation, it remains to be addressed in future experiments whether H3K4me3 is important for transcription activation during later stages, when a repressive chromatin environment prevails. Further studies will additionally address, whether H3K4me3 establishment is restricted to early developmental stages, and prevented from later stages.

APPENDIX

A)

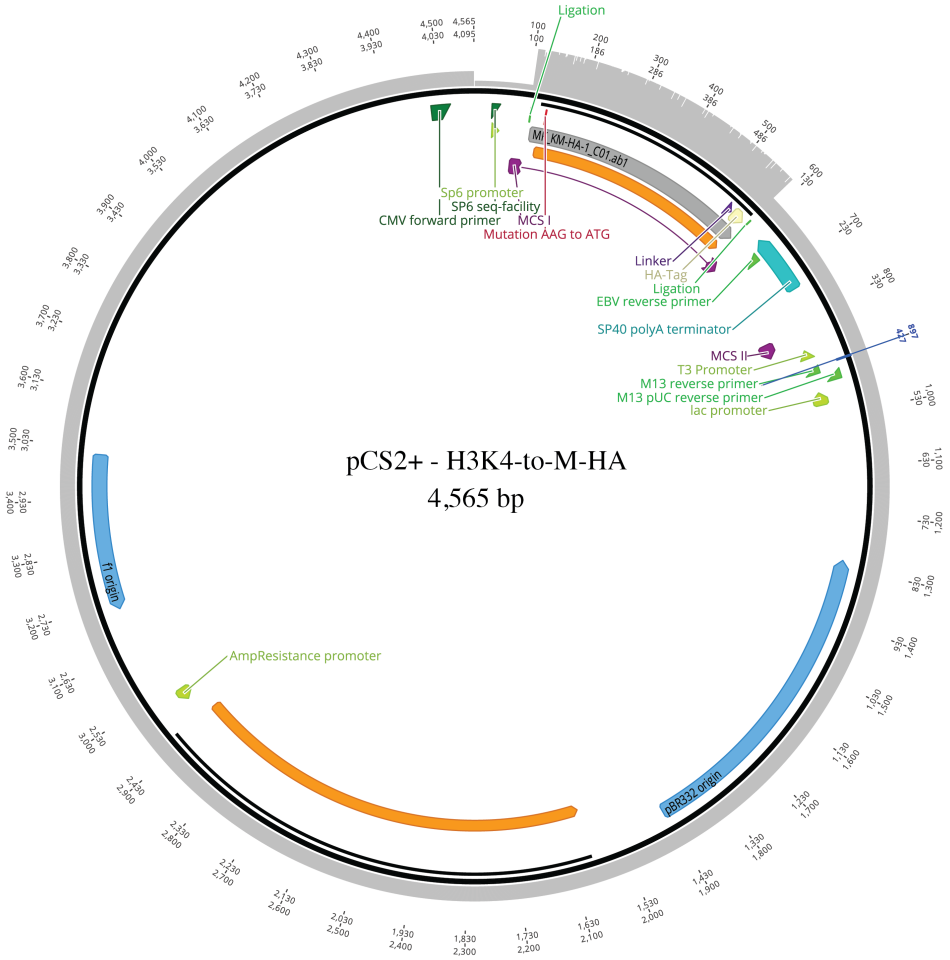


B)

ATGGCCAGAACCAAGCAGACCGCCCGTAAATCCACCGGAGGCAAAGCCCCCAGGAA-
GCAGCTCGCCACTAAGGCTGCCCGCAAGAGCGCCCCGGCTACCGGTGGGGTGAAAAAGCCTCACCGTTATAG
GCCTGGTACCGTGGCTCTTAGAGAGATCCGCCGCTACCAGAAGTCCACCGAGTTGCTGATCCGCAAGCTGCCT
TTCCAGCGTCTGGTGCGAGAAATCGCTCAGGACTTCAAGACTGACCTGCGCTTCCAGAGCTCCGCCGTCATGG
CCCTGCAGGAGTCCAGCGAGGCTTATCTGGTCGGCCTTTTCGAGGACACAAACCAGTGCGCCATCCACGCCAA
GAGGGTCACCATCATGCCCAAGGACATCCAGCTGGCCCGCCGCATCCGCGGAGAGCGCGCTGGATCAGGATC
AGGCTACCCCTACGACGTGCCAGACTACGCTTAA

Vector information pCS2+-H3HA A) Vector map of pCS2+ backbone including H3-HA sequence. Orange, ORF; blue, Origin of replication; green, Primer binding sites. B) Insert sequence coding for wildtype H3.1 with linker sequence (green) and HA-tag sequence (blue) at the 3' end of insert.

A)

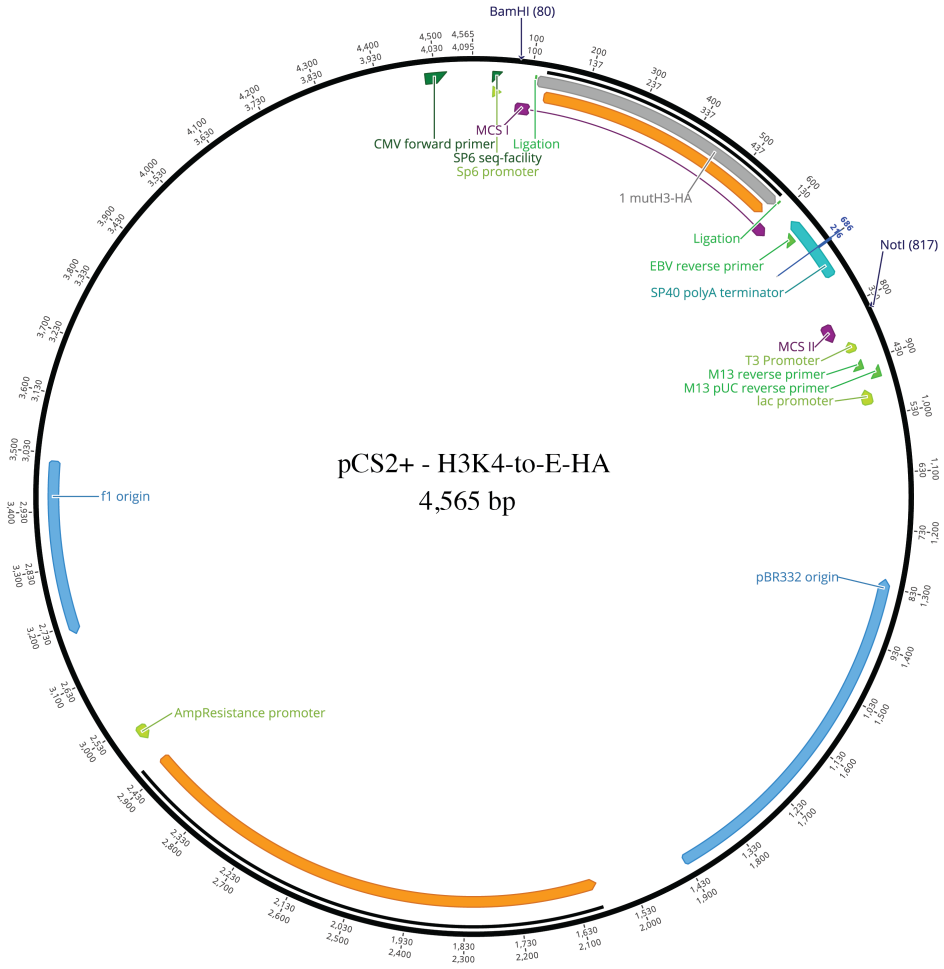


B)

ATGGCCAGAACCATGCAGACCGCCCGTAAATCCACCGGAGGCAAAGCCCCCAGGAA-
GCAGCTCGCCACTAAGGCTGCCCGCAAGAGCGCCCCGGCTACCGGTGGGGTGAAAAAGCCTCACCGTTATAG
GCCTGGTACCGTGGCTCTTAGAGAGATCCGCCGCTACCAGAAGTCCACCGAGTTGCTGATCCGCAAGCTGCCT
TTCCAGCGTCTGGTGCGAGAAATCGCTCAGGACTTCAAGACTGACCTGCGCTTCCAGAGCTCCGCCGTCATGG
CCCTGCAGGAGTCCAGCGAGGCTTATCTGGTCGGCCTTTTCGAGGACACAAACCAGTGCGCCATCCACGCCAA
GAGGGTCACCATCATGCCCAAGGACATCCAGCTGGCCCGCCGCATCCGCGGAGAGCGCGCTGGATCAGGATC
AGGCTACCCCTACGACGTGCCAGACTACGCTTAA

Vector information pCS2+-K4M A) Vector map of pCS2+ backbone including H3-HA sequence with basebair change to allow K4-to-M substitution. Orange, ORF; blue, Origin of replication; green, Primer binding sites. B) Insert sequence coding for wildtype H3.1 with AAG (K)-to-ATG (M) codon substitution at bp 13-16 (red), and linker sequence (green) and HA-tag sequence (blue) at the 3' end of insert.

A)

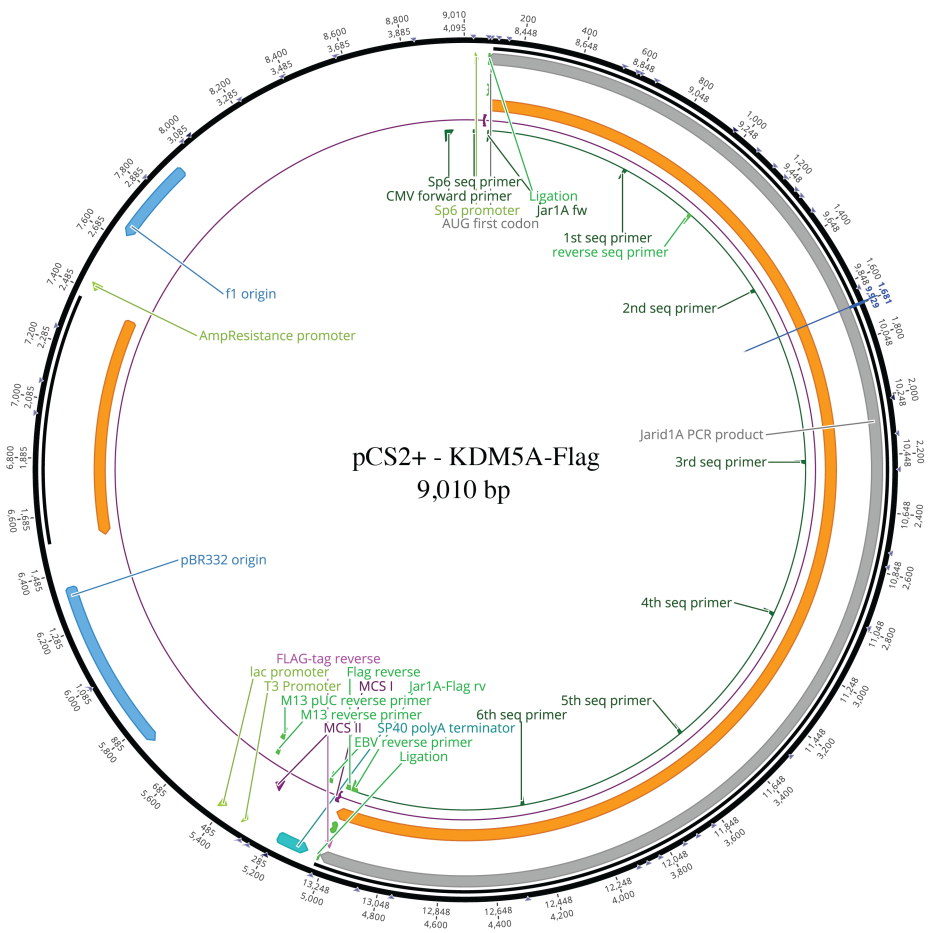


B)

ATGGCCAGAACCGAGCAGACCGCCCGTAAATCCACCGGAGGCAAAGCCCCCAGGAA-
GCAGCTCGCCACTAAGGCTGCCCGCAAGAGCGCCCCGGCTACCGGTGGGGTGAAAAAGCCTCACCGTTATAG
GCCTGGTACCGTGGCTCTTAGAGAGATCCGCCGCTACCAGAAGTCCACCGAGTTGCTGATCCGCAAGCTGCCT
TTCCAGCGTCTGGTGCGAGAAATCGCTCAGGACTTCAAGACTGACCTGCGCTTCCAGAGCTCCGCCGTCATGG
CCCTGCAGGAGTCCAGCGAGGCTTATCTGGTCGGCCTTTTCGAGGACACAAACCAGTGCGCCATCCACGCCAA
GAGGGTCACCATCATGCCCAAGGACATCCAGCTGGCCCGCCGCATCCGCGGAGAGCGCGCTGGATCAGGATC
AGGCTACCCCTACGACGTGCCAGACTACGCTTAA

Vector information pCS2+-K4E A) Vector map of pCS2+ backbone including H3-HA sequence with basebair change to allow K4-to-E substitution. Orange, ORF; blue, Origin of replication; green, Primer binding sites. B) Insert sequence coding for wildtype H3.1 with AAG (K)-to-GAG (E) codon substitution at bp 13-16 (red), and linker sequence (green) and HA-tag sequence (blue) at the 3' end of insert.

A)

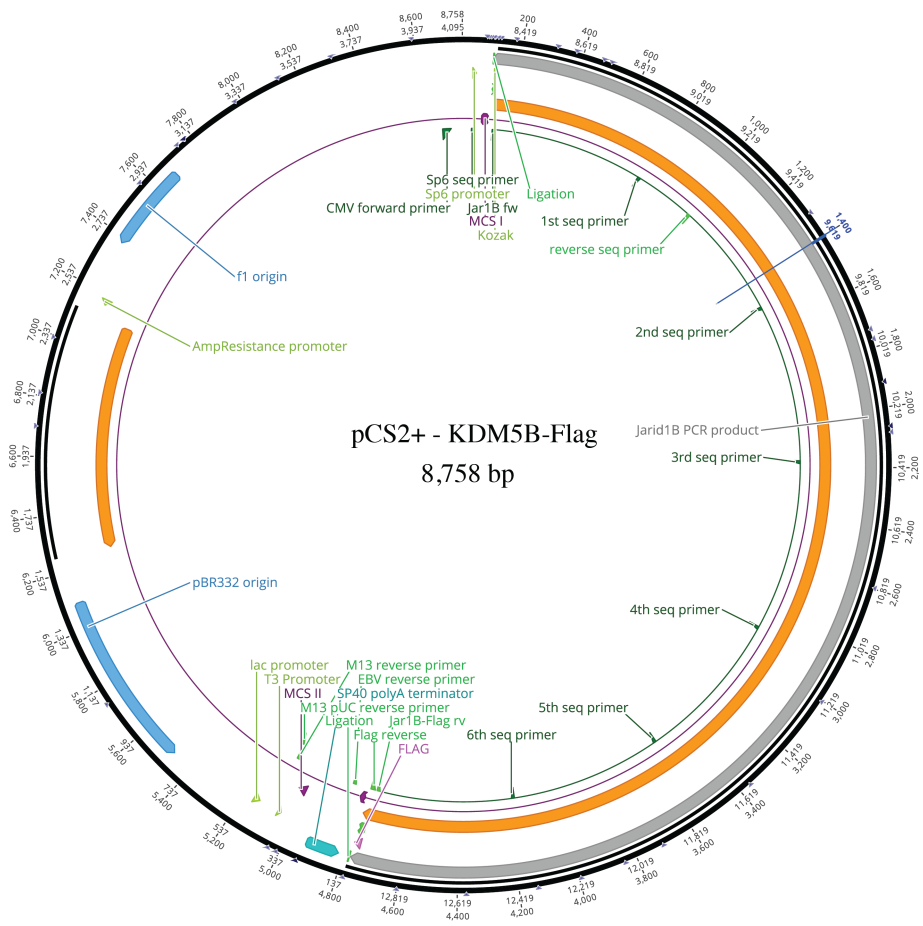


B)

```
ATGGCGTCCGTGGGCCCGGGGGGCTACGCGGCGGAGTTCGTGCCACCGCCCGAGTGCCCG-
GTCTTTGAGCCCAGTTGGGAGGAGTTCACAGACCCGCTGAGTTTTATCGGGCGCATCCGGCCCTTTGCGGAGAAAACGGGCATCTGCAAAATCCGGCCACCCAAGGATTGGCAACCTCCATTTGCCTGTGAAGTA
AAAACCTTCCGTTTCACTCCTAGAGTCCAGCGCCTGAATGAACTTGAGGCAATGACTAGAGTAAGATTGGACTTCTTGGATCAACTGGCAAAATTTTGGGAACTTCAAGGATCTACATTGAAGATCCCTGTGGTGG
AGAGAAAAATCCTGGATTTGTATGCTTTAAGCAAGATTGTTGCCAGCAAAGGAGGTTTTGAAATAGTCACCAAAGAGAAGAAATGGTCTAAAGTGGGTAGCCGCTTGGGATATCTGCCAGGAAAAGGAACTGGGT
CTCTTTTGAAGTCACACTATGAAAGAATTCTCTACCCATATGAACTTTTCCAATCTGGTGTCAGCCTTATGGGTGTACAAATGCCTGATTTAGATCTTAAAGAAAAAGTTGAGGCTGAGGTTCTCAGCACTGACATCC
AGCCTTCCCCAGAGCGGGGAACAAGAATGAATATTCCACCAAAGAGAACAAGACGAGTTAAGTCTCAGTCAGACTCTGGAGAAGTAAATAGAAACACAGAACTGAAGAAACTTCAGATTTTTGGGGCTGGACCC
AAAGTTGTGGGCCTGGCAGTAGGAGCAAAAGATAAAGAAGATGAAGTCACCCGAAGAAGAAAAGTTACCAACAGGTCAGATGCATTTAACATGCAAATGAGACAACGGAAAGGAACTCTCTCTGTTAACTTTGT
TGATCTCTATGTTTGCATGTTTTGTGGTCGAGGAAATAATGAGGATAAACTGCTTTTGTGTGATGGGTGCGATGACAGCTATCATACATTTTGTCTACTTCCACCACTACCTGATGTCCCCAAAGGAGACTGGAGGTG
TCCTAAATGTGTTGCTGAGGAATGTAACAAACCTCGAGAAGCCTTTGGATTTGAGCAAGCTGTAAGAGAATATACACTTCAAAGCTTTGGAGAGATGGCAGATAATTTTAAATCTGATTATTTCAATATGCCAGTCCA
TATGGTTCCTACAGAGCTAGTAGAAAAAGAATTCTGGCGGCTAGTAAGTAGCATTGAAGAAGATGTTATTGTAGAGTATGGAGCTGATATCTCTTCAAAAGATTTTGGAAGTGGATTTCCTAAAAAGGATGGTCAAA
GAAAGATGTTGCCAGAAGAAGAGGAATATGCACTTTCTGGTTGGAATTTGAATAACATGCCTGTTCTGGAGCAGTCCGTTCTTGCACATATTAATGTGGACATTTCTGGTATGAAGGTGCCATGGCTCTATGTGGGAA
TGTGCTTTTCTTCTTTTTGCTGGCATATTGAAGATCACTGGAGTTATTCCATTAACTACTTGCATTGGGGCGAGCCAAAGACATGGTATGGTGTACCATCTCATGCTGCAGAACAACTGGAGGAGGGTGATGAGGGAG
CTGGCCCCTGAGTTATTTGAATCCCAGCCTGATCTCCTACATCAGTTAGTAACCATCATGAACCCAAATGTGCTCATGGAGCATGGTGTGCCTGTGTACAGGACTAATCAGTGTGCTGGAGAGTTTGTGGTGACGTTT
CCTCGTGCCTATCACTCTGGATTTAACCAGGGATACAACTTTGCCGAAGCGGTGAACTTCTGTACTGCTGACTGGTTGCCCATTGGACGTCAGTGTGTAAATCACTATCGACGGCTCAGGCGCCATTGTGTCTTTTC
ACATGAGGAACTCATTTTCAAGATGGCAGCAGATCCAGAATGCTTAGATGTGGGGCTGGCTGCCATGGTATGCAAAGAGTTGACTCTCATGACTGAAGAAGAAACACGATTGAGAGAATCAGTTGTTCAAATGGGT
GTCGTGATGTCAGAAGAAGAAGTCTTTGAACTTGTTCCTGATGATGAGCGGCAGTGTTCAGCATGCAGAACTACATGTTTTCTCTCAGCTCTCACATGCTCCTGTAATCCGGAGCGACTTGTATGTCTGTATCATCCA
ACTGATTTGTGCTCTTGCCCCATGCAGAACAAATGTCTTAGGTACCGCTATCCACTAGAAGACCTTCCATCTCTGCTTTATGGTGTCAAAGTGAGAGCACAATCCTATGACACTTGGGTCAATCGTGTTACAGAAGCA
TTGTCAGCTAGCTTCAACCATAAGAAAGATTTGATTGAATTGAGAGTAATGCTGGAAGATGCTGAGGATAGGAAATACCCAGAGAATGATCTCTTTCGAAAGCTCAGGGACGCGGTGAAAGAAGCTGAGACCTGT
GGTTCTGTGGCTCAGCTGCTTCTGAGCAAAAAGCAAAAACACAGGCAGAGCTCAGATAGTGGGAAGACTCGGACCAAACTGACAGTAGAAGAATTGAAGGCCTTTGTCCAACAACTTGTTAGTCTTCCGTGTGT
CATCAGCCAAACTCGACAAGTAAAGAATCTGCTGGATGATGTAGAAGAATTTCATGAACGTGCCCAGGAGGCCATGATGGATGAAACCCCAGATTCTTCCAAACTCCAAATGTTAATAGACATGGGCTCTAGTCTCT
ATGTGGAGCTACCTGAATTGCCCCGACTAAAGCAAGAGCTCCAGCAGGCTCGATGGCTGGATGAAGTAAGGCTAACTCTATCAGATCCACAACAAGTCACTTTGGATGTCATGAAGAAATTGATTGACTCTGGCGT
AGGCTTGGCACCTCACCATGCTGTAGAAAAAGCAATGGCCGAACTACAGGAACTCCTTACAGTCTCTGAACGATGGGAAGAGAAGGCTAAGGTCTGCCTGCAAGCAAGACCTCGGCACAGTATGGCAAATTTAG
AAAATATTGTGAATGAAGCCAAGAACATTCCAGCCTTTCTACCCAATGTGTTGTCCTTAAAAGAAGCCTTACAAAAGGCTCGTGAATGGACAGCTAAAGTGGAAGCTATTCAGAGTGGCAACAACTATGCTTACTT
GGAACAGCTTGAGAGCCTATCTGCTAAAGGACGCCCTATCCCTGTGCGTCTTGATGCCTTGCCTCAAGTGGAGTCACAGGTAGCAGCAGCACGGGCCTGGAGAGAACGTACTGGGCGGACCTTCCTTAAAAAGAA
TTCCAGCCATACATTGTTACAGGTGCTAAGTCCACGAACTGACATTGGTGTATATGGGAGTGGTAAAAACAGGAGGAAAAAGGTAAAAGAAATAATAGAAAAAGAAAAGGGAGAAGGACCTCGACCTTGAGCCTTT
GAGTGACTTAGAGGAAGGACTGGAGGAGTCAAGAGATACAGCCATGGTGGTGGCAGTTTTCAAAGAACGAGAGCAAAAGGAGATTGAAGCCATGCATTCCCTCAGAGCAGCCAACCTAGCCAAGATGACAATTG
TGGAGCGCATAGAAGAAGTGAAATTTTGTATTTGCCGCAAGACAGCCAGTGGGTTTATGCTACAGTGTGAGCTCTGCAAAGACTGGTTCCATAACAGCTGTGTTCCCCTTCCTAAATCAAGTTCCCAGAAAAAGGG
ATCCAGCTGGCAGGCTAAAGATGTAAAATTCCTTTGCCCTCTTTGCATGAGATCTCGGAGGCCAAGGCTAGAGACTATTCTGTCACTCCTGGTATCCCTTCAGAAGTTGCCTGTACGGTTGCCGGAAGGAGAAGCA
CTGCAGTGTTTGACAGAACGTGCCATGAGTTGGCAAGATAAAGCACGCCAAGCTTTAGCCACAGATGAACTGTCCTCTGCCCTGGCCAAGCTCTCTGTGCTAAGCCAGCGTATGGTGGAACAAGCAGCTAGAGAA
AAAACTGAAAAGATTATCAGTGCAGAACTCCAAAAAGCAGCTGCCAATCCAGATTTACAGGGACACTTACCTAGTTTCCAGCAGTCTGCTTTTAACCGGGTGGTGAGCAGTGTATCATCTTCCCCCCACCAAACAA
TGGACTATGATGATGAAGAAACAGATTCTGATGAAGATATTCGGGAGACATACGGCTATGACATGAAGGATACAGCCAGCGTGAAGTCCTCTAGTAGTCTGGAACCCAATCTCTTTTGTGATGAAGAGATTCCCATC
AAGTCTGAGGAAGTGGTGACTCACATGTGGACGGCACCGTCATTCTGTGCAGAGCATGCATATTCTTCTGCTTCAAAGAGTTGTTCTCAAGGTTCTAGCACTCCAAGAAAGCAGCCTCGGAAGAGCCCTTTGGTTC
CCAGAAGTTTGGAGCCTCCAGTATTGGAATTGTCTCCTGGAGCAAAAGCCCAGTTGGAAGAACTCATGATGGTTGGAGACCTCCTAGAAGTGTCTCTGGATGAGACACAACACATTTGGCGGATTTTGCAGGCCA
CACATCCCACCCTCAGAGGACAGATTCTTACATATCATGGAGGATGATAGCATAGAAGAGAAAACCATTAAAAATGAAAGGAAAGGACTCTTCAGAGAAGAAACGGAAACGGAAGCTAGAAAAGGTAGAACAGCTT
TTTGGAGAAGGAAAACAGAAGACCAAGGAGCTAAAGAAAATAGATAAACCTAAAAAGAAGAAATTAAAATTAAATGTGGACAAATCAAAAGAGCTGAATAAATTGGCCAAGAAGTTAGCTAAAGAAGAAGAGC
GAAAGAAAAAGAAAGAAAAGGCTGCTGCCGCCAAAGTGGAGCTTGTGAAAGAGAGCACGGAGAAGAAACGGGAGAGGAAGGTGTTGGATATCCCCTCAAAGTATGACTGGTCGGGGGCAGAGGAGTCTGATG
ATGAGAATGCTGTGTGTGCAGCCCAAAACTGCCAAAGGCCGTGTAAGGACAAGGGAGTTGTATTTGTAACGGAAGAAGAGAAGAATAAAAAAGATTACAAGGATGACGATGACAAGTAA
```

Vector information pCS2+-KDM5A A) Vector map of pCS2+ backbone including mouse KDM5A exonic sequence. Orange, ORF; blue, Origin of replication; green, Primer binding sites. B) Insert sequence coding for full-length mouse KDM5A with single FLAG tag sequence at the 3' end of insert.

A)

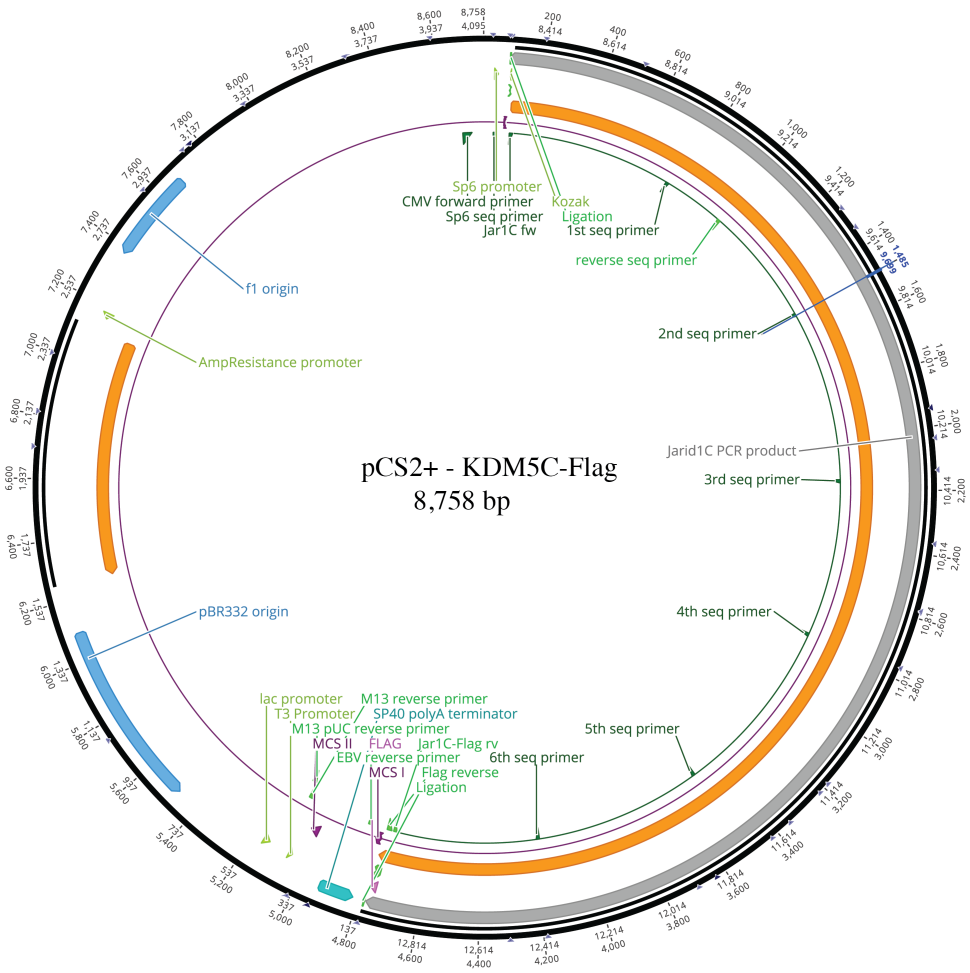


B)

ATGGAGCCGGCCACCACGCTGCCCCCAGGCCCGCGCCCGGCGCTGCCCCTCGGGGGCC-
CGGGCCCGCTGGGAGAGTTCCTGCCTCCCCCCCGAGTGTCCGGTCTTCGAGCCGAGCTGGGAAGAGTTCGCGGACCCCTTCGCTTTCATCCACAAGATCCGGCCCATAGCCGAGCAGACTGGGATCTGTAAGGTGC
GGCCGCCGCCGGATTGGCAGCCACCATTCGCTTGTGATGTCGATAAACTTCATTTCACCCCGCGGATCCAGAGGCTGAATGAGCTGGAGGCCCAAACACGTGTAAAATTGAATTTCCTGGACCAGATTGCAAAGTA
TTGGGAGCTACAAGGAAGTACTCTGAAGATTCCCCACGTGGAGAGAAAGATCTTAGACTTGTTCCAGCTAAATAAGTTAGTCGCAGAAGAAGGTGGGTTTGCAGTTGTTTGCAAGGATAGAAAATGGACCAAAAT
TGCCACCAAGATGGGTTTTGCTCCTGGTAAAGCTGTGGGCTCACATATCAGGGGACATTATGAGCGAATACTCAATCCATACAATTTGTTCCTGTCTGGAGACAGTCTGAGGTGTCTTCAGAAGCCTAACCTGACCT
CCGACACAAAGGACAAGGAGTACAAACCCCATGATATTCCCCAGAGGCAGTCTGTGCAGCCCGCAGAAACCTGTCCTCCGGCCCGCCGAGCAAAGCGCATGAGAGCAGAGGCTATGAATATTAAAATTGAGCCA
GAAGAGGCAACAGAAGCCAGGACTCATAATCTGAGACGCCGAATGGGTTGCACAACTCCAAAATGGGAAAATGAGAAAGAAATGAAAAGCACCATCAAGCAAGAGCCCACTGAGAAGAAAGACTGCGAGTTA
GAAAGTGAGAAGGAGAAGCCCAAGAGCCGGGCCAAGAAAACGGCCACTGCGGTGGACCTGTATGTGTGCCTCTTATGTGGCAGTGGCAATGACGAGGACCGGCTCCTACTGTGTGACGGCTGTGATGACAGTTA
CCACACCTTCTGCCTGGTTCCCCCTCTCCATGACGTTCCCAAGGGAGACTGGAGGTGTCCTAAGTGTTTGGCTCAAGAATGTAATAAGCCACAAGAAGCTTTTGGCTTTGAACAAGCAGCAAGGGACTATACACTT
CGTACGTTTGGAGAAATGGCAGATGCATTCAAATCTGATTACTTCAACATGCCAGTCCATATGGTTCCCACAGAGCTGGTGGAGAAGGAGTTTTGGCGGCTCGTAAGCACTATTGAAGAAGACGTCACGGTGGAGT
ATGGAGCTGATATTGCTTCAAAAGAGTTTGGCAGTGGCTTTCCTGTTCGAGATGGGAAAATCAAGATTTCTCCTGAGGAAGAGGAATACCTGGATAGTGGCTGGAATTTGAACAACATGCCAGTGATGGAGCAGTC
TGTCCTTGCTCATATTACTGCTGATATATGTGGCATGAAACTCCCCTGGTTGTATGTGGGCATGTGCTTTTCTTCATTCTGTTGGCACATTGAAGACCACTGGAGCTATTCAATCAACTACCTACACTGGGGTGAGCCA
AAAACCTGGTATGGAGTTCCAGGCTACGCTGCTGAGCAGCTAGAAAACGTGATGAAGAAGCTCGCTCCTGAGCTCTTTGTGTCCCAGCCGGACCTGCTCCACCAGCTGGTGACCATCATGAACCCCAACACTCTG
ATGACTCATGAAGTGCCTGTTTATCGAACGAATCAGTGTGCTGGGGAGTTTGTAATTACGTTTCCAAGAGCCTACCACAGTGGCTTCAACCAAGGTTTTAATTTTGCCGAGGCTGTTAACTTCTGCACTGTTGATTG
GCTGCCATTGGGCCGCCAGTGTGTGGAGCATTATCGCTTGCTGCACCGTTATTGTGTCTTTTCCCACGATGAAATGATATGCAAAATGGCTTCTAAGGCTGATGTATTAGACGTTGTGGTAGCGTCAACTGTTCAGAA
GGACATGGCCATCATGATCGAGGACGAGAAAGCTTTAAGGGAAACTGTTCGGAAATTGGGAGTAATTGATTCTGAAAGAATGGATTTTGAGCTGTTGCCAGATGATGAGCGGCAGTGTATAAAATGCAAAACTACA
TGCTTCATGTCTGCCATCTCCTGCTCTTGTAAACCTGGCCTACTTGTTTGCCTACATCATGTGAAAGAATTGTGTTCTTGCCCCCCGTATAAGTATAACCTGCGGTATAGATACACTCTGGATGATCTCTATCCAATGAT
GAATGCATTAAAGCTTCGAGCAGAGTCTTACAATGAATGGGCCTTAAATGTGAATGAAGCCTTGGAAGCAAAGATCAATAAGAAGAAAAGTCTTGTCAGCTTTAAAGCTTTGATTGAAGAATCAGAAATGAAGAA
ATTCCCAGACAATGATCTTTTGCGTCACCTTCGACTAGTCACGCAGGATGCAGAAAAGTGTGCCTCTGTTGCCCAGCAGTTGCTTAATGGCAAAAGGCAAACAAGGTATCGATCTGGTGGAGGAAAGTCTCAGAAT
CAGCTGACGGTGAATGAACTTAGACAGTTTGTGACTCAGCTGTATGCTCTGCCGTGCGTCCTTAGTCAGACACCGTTGCTAAAGGATCTCTTGAATCGTGTGGAAGATTTCCAACAGCAAAGCCAAAAACTACTGT
CTGAGGAAATGCCTAGCGCTGCCGAGCTACAGGAGTTGCTGGATGTCAGCTTTGAATTTGATGTTGAGCTTCCACAGCTCACTGAGATGCGCATTCGCTTAGAGCAGGCCCGCTGGCTGGAAGAGGTTCAGCAGG
CTTGCCTGGACTCCAGCTCCCTTTCTTTGGATGATATGAGACGACTCATAGACTTAGGGGTGGGGCTGGCTCCATATTCTGCTGTGGAGAAAGCTATGGCCCGCCTGCAGGAGCTGCTAACAGTGTCAGAACACTG
GGACGACAAAGCCAAGAGTCTCCTCAGAGCAAGACCACGGCATTCTTTGAGTAGCCTTGCTACAGCAGTCAAAGAGATGGAGGAGATCCCCGCCTACCTGCCCAATGGTACAGTCCTGAAGGATTCTGTGCAAA
GAGCCAGAGACTGGGTTCAGGATGTGGATGCATTGCAGGCAGGAGGACGTGTGCCAGTGTTAGAGACACTCATTGAACTTGTTGCAAGAGGCCGATCTATCCCAGTACATCTCAATTCTTTACCAAGATTGGAAAT
GTTAGTAGCTGAAGTTCATGCATGGAAAGAATGTGCTGCTAAAACATTCTTGCCTGAGAATTCTACATATTCTCTGTTAGAGGTTCTTTGCCCTCGGTGTGACATTGGCCTTTTGGGGCTGAAAAGGAAACAGAGAA
AGTTGAAGGAGCCCTTGCCGAGTGGAAAGAAGAAGACCACCAAGTTAGAGAGCCTGAGCGATCTGGAGAGGGCTCTGATGGAGAGCAAAGAGACTGCGCAGCTATGGCAACTCTTGGGGAAGCTCGCCTACG
AGAGATGGAAGCTTTGCAGTCTCTCAGATTTGCCAATGAAGAGAAATTGCTGTCACCTGTCCAAGACTTAGAGATGAAGTCTGCTTGTGTCAGAAGACTCCAGCCACTCCCATGATTCAGTGTGAACTCTGCAGG
GATGCTTTCCATACTAGTTGTGTGGCAGCACCCAGTATTCACAAAGCTCCCGAATCTGGCTTTGTCCTCACTGTCCGGAGGTCAGAGAAGCCTCCATTAGAGAAAATTCTGCCTCTGCTGGCCTCCCTACAGCGTAT
ACGAGTTCGTCTTCCTGAAGGAGATGCACTCCGATACATGATTGAAGAAACCGTGAACTGGCAACATAGAGCCCAGCAGCTGCTTTCTTCAGGGAATCTAAAACTTGTGCAAGACCAAGTGGGTTCTGGACTGTT
GTCTAGCAGATGGCCAGCCTCAGCAGGACAGGCATCAGCAACAGACAAGGTGTCTCAGCCTCCTGGCACAACATCCTTTTCATTGCCCGATGACTGGGACAACAGAACCTCATATTTACACTCTCCCTTCTCTACT
GGACAGAGTTGTCTCCCCCTCCATGGTTTGAGTCCAGAAGTGGCCTTAAATGTGAATGAAGCCCAGCTGCTCCAGGTATCCCTTCCTGAAATTCAGGAGCTTTATCAGACTCTACTTACTAAGCCAAGCTCTGTTCA
GCAAGCTGACCGAAGTTCACCAGTTAGATCCAGCAGTGAGAAGAACGATTGCCTTCGAGGGAAACGAGATGCAATCAACAGTCCTGAGAGGAAACTGAAAAGGCGCCCAGAAAGAGAAGGCCTTCCTAGCGAG
CGGTGGGATCGAGTTAGACACATGCGGACCCCCCAAAAGAAGAAAATCAAACTGAGTCACCCCAAGGACATGGACAGTTTCAAGTTAGAAAGAGAGCGTAGCTATGACTTAGTGCGTAATGCTGAAACTCATTCC
CTGCCCTCAGACACATCCTACTCCGAACAGGAGGACTCTGAGGATGAAGATGCCATCTGTCCAGCTGTGAGCTGCCTGCAGCCAGAAGGAGATGAGGTGGACTGGGTCCAGTGTGATGGCAGCTGCAATCAGTG
GTTTCACCAGGTCTGTGTTGGTGTCTCTCCAGAGATGGCAGAGAAGGAAGACTACATCTGTGTGCGCTGTACTGGGAAGGACGCACCAAGCCGAAAGGATTACAAGGATGACGATGACAAGTAA

Vector information pCS2+-KDM5B A) Vector map of pCS2+ backbone including mouse KDM5B exonic sequence. Orange, ORF; blue, Origin of replication; green, Primer binding sites. B) Insert sequence coding for full-length mouse KDM5B with single FLAG tag sequence at the 3' end of insert.

A)

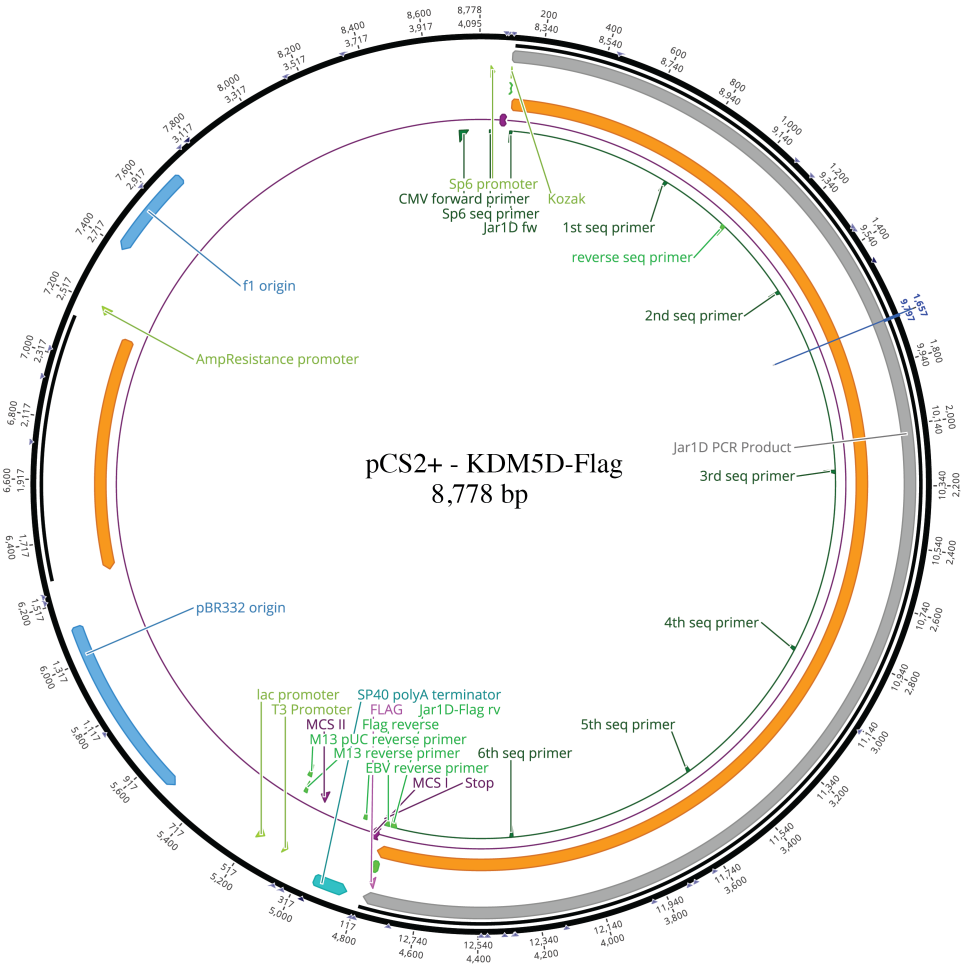


B)

ATGGAGCTGGGGTCCGACGATTTCCTACCGCCGCCGGAGTGTCCGGTGTTCGAGC-
CTAGCTGGGCCGAGTTCCGAGACCCTCTTGGCTACATCGCGAAAATCAGACCCATCGCCGAGAAGTCCGGCATTTGCAAGATCCGCCCACCCGCGGACTGGCAGCCACCCTTTGCCGTGGAAGTGGATAACTTCA
GGTTTACTCCCCGAATCCAGAGGTTGAATGAGCTGGAGGCCCAGACAAGAGTGAAACTGAACTATTTGGACCAGATTGCCAAATTCTGGGAGATCCAGGGCTCATCCTTAAAGATTCCCAATGTAGAACGGCGGAT
CTTGGACCTCTATAGCCTCAGCAAAATTGTGGTGGAAGAAGGTGGTTATGAAACTATCTGCAAGGACCGTCGATGGGCCCGTGTGGCCCAGCGTCTCAACTACCCACCTGGCAAAAACATTGGTTCCTTGCTACGC
TCTCACTATGAACGCATTGTTTATCCCTATGAAATGTACCAGTCTGGAGCCAACCTTGTGCAGTGTAACACACGCCCATTTGATAATGAGGAAAAGGACAAGGAATATAAACCCCACAGCATCCCCCTTCGACAGTC
TGTGCAGCCTTCCAAGTTCAACAGTTATGGACGACGGGCAAAGAGACTACAGCCTGATCCGGAACCCACAGAGGAAGACATTGAAAAGAATCCAGAACTGAAGAAATTACAGATATATGGAGCAGGCCCCAAGA
TGATGGGGCTGGGCCTAATGGCCAAAGATAAGACTCTGCGGAAAAAAGATAAAGAAGGGCCTGAGTGTCCCCCCACTGTGGTGGTAAAGGAGGAATTAGGTGGGGATGTCAAGATGGAATCAACCTCACCCAAG
ACCTTCCTAGAAGGCAAGGAGGAACTGAGTCATAGCCCTGAGCCCTGCACCAAGATGACCATGAGGCTTCGAAGGAACCACAGCAATGCCCAGTTTATTGAGTCATACGTATGCCGAATGTGTTCCCGAGGAGAT
GAAGATGACAAACTCTTACTGTGTGATGGCTGTGATGACAATTACCACATCTTTTGCCTGCTGCCTCCTTTGCCTGAAATCCCGAAAGGTGTCTGGAGGTGTCCAAAGTGTGTCATGGCGGAGTGTAAGCGGCCCC
CTGAAGCCTTTGGCTTTGAGCAAGCTACCCGGGAATATACTCTGCAGAGCTTTGGAGAGATGGCTGACTCCTTTAAAGCTGACTACTTCAACATGCCTGTGCATATGGTACCCACAGAACTTGTAGAGAAGGAGTT
TTGGCGGCTGGTGAATAGCATCGAGGAAGACGTAACTGTGGAGTATGGGGCTGACATCCATTCCAAAGAATTTGGCAGCGGTTTCCCTGTCAGTGACAGTAAGCGGCACCTAACCCCAGAGGAGGAGGAGTATGC
TACAAGTGGTTGGAACCTGAATGTGATGCCTGTATTGGAGCAGTCTGTACTGTGCCACATCAATGCAGATATATCTGGCATGAAGGTACCGTGGCTGTATGTGGGAATGGTCTTCTCAGCCTTTTGCTGGCATATTGA
AGATCACTGGAGTTACTCTATTAACTACCTCCACTGGGGTGAGCCAAAAACCTGGTATGGAGTCCCATCGCTTGCAGCGGAACACTTGGAAGAAGTGATGAAGAAGCTGACACCTGAGCTTTTTGATAGCCAGCCT
GATCTCCTTCACCAACTCGTCACCCTCATGAATCCTAACACGCTCATGTCCCATGGTGTGCCGGTTGTCCGCACAAACCAGTGTGCAGGAGAATTTGTCATCACTTTTCCTCGTGCTTACCACAGTGGTTTCAACCA
AGGCTACAACTTTGCTGAGGCAGTCAACTTTTGCACTGCTGACTGGCTGCCTGCTGGGCGCCAATGCATTGAGCACTACCGCCGACTACGAAGATACTGTGTCTTTTCTCATGAGGAGCTCATTTGCAAGATGGCT
GCCTGCCCTGAGAAACTGGACCTGAACCTGGCAGCAGCTGTACATAAGGAGATGTTCATTATGGTGCAAGAGGAGCGGCGTCTACGAAAGGCTCTGTTGGAAAAAGGTATCACCGAGGCTGAGCGAGAGGCTTT
TGAGCTGCTTCCAGATGATGAACGTCAGTGCATAAAGTGCAAGACCACATGTTTCCTATCTGCCCTGGCTTGCTATGACTGCCCAGATGGCCTTGTCTGCCTTTCTCACATCAATGACCTCTGCAAGTGCTCCAGTAG
CCGCCAGTACCTTCGGTATCGGTACACCTTGGATGAGCTCCCTGCCATGCTGCATAAACTGAAAGTTAGAGCTGAGTCCTTTGACACCTGGGCTAACAAAGTACGAGTGGCCCTGGAGGTGGAAGATGGGCGGAA
GCGCAGCCTGGAAGAACTGAGAGCACTAGAATCTGAGGCTCGTGAGCGAAGGTTTCCTAACAGTGAGCTTCTACAGCGACTGAAAAACTGCCTGAGTGAAGCAGAGGCTTGTGTGTCCCGGGCTCTGGGATTGG
TCAGTGGCCAGGAAGCTGGCCCTGACAGGGTGGCTGGCCTACAGATGACCCTGGCTGAGCTACGGGATTTTCTGGGCCAGATGAACAACCTGCCCTGTGCCATGCACCAGATTGGGGATGTCAAGGGTATTCTGG
AACAGGTGGAAGCCTACCAGACTGAGGCCCGTGAGGCCCTGGTCTCACAGCCTTCCAGTCCAGGGCTGCTGCAGTCCTTGTTGGAGAGAGGACAACAGCTGGGTGTGGAGGTACCTGAAGCCCAGCAGCTCCA
AAGGCAGGTGGAACAGGCGCGGTGGCTGGATGAGGTAAAACGCACGCTGGCTCCCTCTGCCCGAAGGGGTACCCTGGCCATCATGCGGGGACTGTTGGTTGCGGGCGCCAGTGTAGCCCCTAGCCCTGCTGTGG
ATAAGGCCCAGGCAGAGCTTCAGGAGCTGCTGACCATCGCTGAACGCTGGGAAGAGAAAGCCCACCTCTGCTTAGAGGCCAGGCAGAAGCATCCACCAGCCACACTTGAGGCCATAATCCATGAAGCAGAAAAC
ATCCCTGTTCACTTGCCCAATATCCAGTCTCTCAAGGAGGCTCTTGCTAAGGCCCGGGCATGGATTGCTGATGTGGATGAGATCCAAAATGGTGACCAACTACCCTTGCTTGGATGACTTGGAGGGCCTCGTAGCTGT
GGGTCGGGACCTCCCTGTGGGGCTGGAGGAACTGAGACAGCTAGAGCTACAGGTATTGACAGCACACTCCTGGAGAGAAAAGGCCTCCAAGACCTTCCTCAAGAAGAACTCCTGCTATACATTATTGGAGGTGCT
CTGCCCGTGTGCAGACGCCGGCTCAGACAGCACCAAACGCAGCCGGTGGATGGAGAAGGAGCTGGGGTTGTACAAATCTGACACAGAGCTGTTGGGGCTGTCTGCGCAGGACCTCAGGGACCCAGGCTCTGTG
ATCGTGGCCTTCAAAGAGGGGGAGCAGAAGGAGAAGGAAGGGATACTGCAGCTTCGTCGTACCAACTCAGCCAAGCCAAGTCCACTGGCACTGTTGACAACAGCCTCCTCTACAGCTTCTATCTGCGTGTGTGG
GCAAGTGCCAGCTGGGGTGGGAGCTCTGCAGTGTGACCTGTGTCAGGACTGGTTCCATGGGCGCTGTGTGACAGTACCCCGCCTCCTCAGCTCCCAGAGGTCCAGTCTTCCCTCATCCCCACTGCTGGCCTGGTG
GGAATGGGACACCAAATTCTTGTGCCCTCTGTGTATGCGCTCACGGCGCCCACGCTTGGAAACCATCCTGGCACTGCTGGTAGCCCTACAGAGACTGCCTGTGCGGCTGCCTGAGGGTGAGGCTCTACAGTGCCTC
ACAGAGAGAGCCATCAGCTGGCAAGGCCGTGCCAGACAGGTTTTGGCCTCTGAGGAAGTGACTGCTCTGTTGGGACGGCTGGCTGAACTGCGCCAACGGCTACAGGCTGAATCCAAGCCTGAGGAGTCCCTGGC
TTACCCTTCAGATGGTGGAGAGGGCACTGGCAATATGCCGAAGGTCCAGGGGCTGTTGGAGAATGGGGACAGTGTGACCAGTCCTGAAAAGGTAGCCACAGAGGAGGGCTCAGATCTGGAGCTGCTATCCTCAAT
ATTACCACAGTTGTCTGGCCCTGTGTTGGAACTGCCTGAGGCAACCCGAGCCCCACTGGAGGAACTTATGATGGAAGGGGATCTGCTTGAGGTGACCCTAGATGAGAATCATAGCATCTGGCAGCTGCTGCAGGCT
GGGCAGCCTCCAGACTTGAAGAGGGTCCAGACACTTCTGGAGCTGGAGAAGGCAGAGCGCCATGGGAGTCGGACACGAGGCCGAGCCCTAGAGAGGAGACGACGGAGGAAGGTGGATCGGGGTGGGGAGCC
GGATGACCCGGCACGAGAGGAGCTAGAGCCAAAGAGGGTACGAAGCTCAGGACCAGAAGCCGAGGAAGTCCAGGAGGAAGAGGAGCTGGAGGAGGAGACTGGGGGTGAGGTCCCTCCTGTACCCTTCCCCAA
CAGTGGCAGCCCCAGCATCCAGGAGGACCAGGATGGCTTAGAACCAGTGTTAGAGGCTGGTTCTGACACCTCAGCCCCTTTCTCTACTCTGACTTCCCGGCTGTTGATGTCCTGCCCACAGCAGCCATCTCTGCAA
CAATTGGATTACAAGGATGACGATGACAAGTAA

Vector information pCS2+-KDM5C A) Vector map of pCS2+ backbone including mouse KDM5C exonic sequence. Orange, ORF; blue, Origin of replication; green, Primer binding sites. B) Insert sequence coding for full-length mouse KDM5C with single FLAG tag sequence at the 3' end of insert.

A)

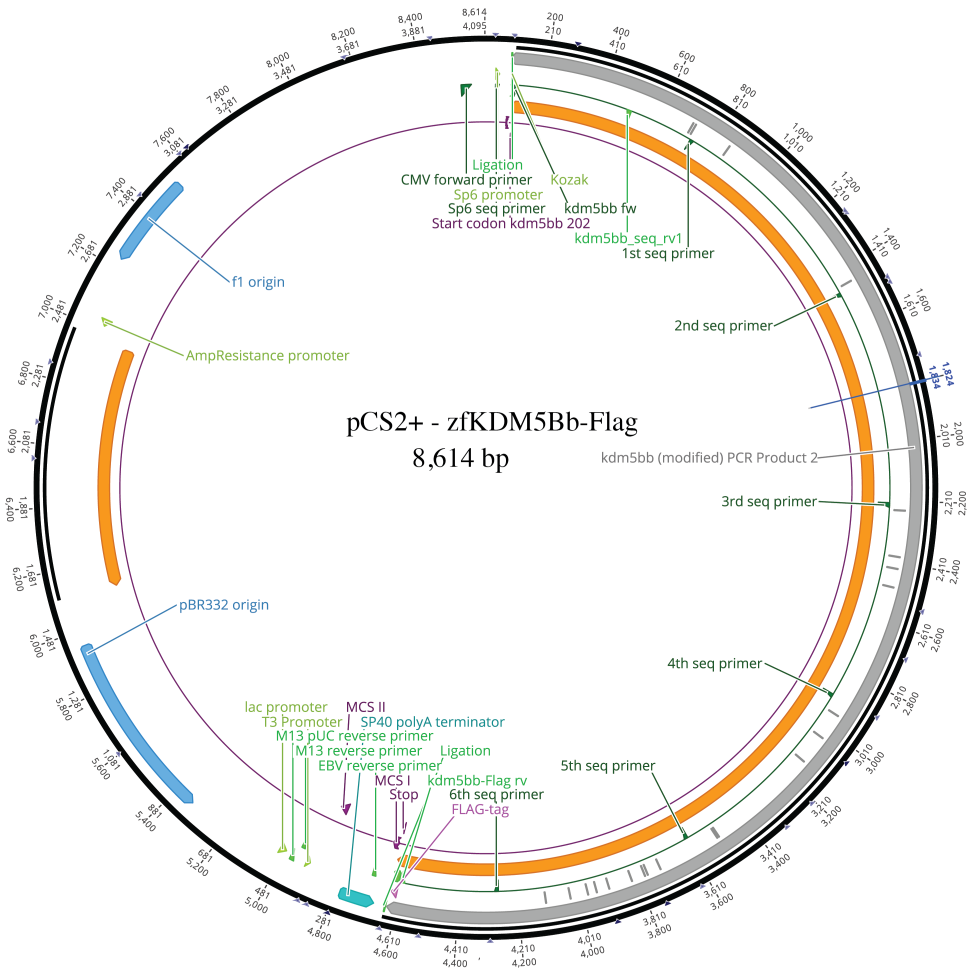


B)

ATGAAGCCAGGATCTGACGACTTTCTACCGCCGCCTGAGTGCCCTGTGTTTGAGCC-
CAGTTGGGCGGAATTCCGCGATCCTCTTGGCTATATAGCTAAAATCAGACCCATTGCGGAGAAATCAGGCATTTGCAAGATACGCCCACCTGCGGATTGGCAACCTCCTTTTGCTGTAGAAGTTGATAACTTCAGATT
TACTCCTCGAATTCAGAGGCTAAATGAACTGGAGGCCCAGACCAGAGTAAAACTGAACTATTTGGATCAGATTGCAAAGTTCTGGGAAATTCAAGGTTCTTCCTTAAAGATACCTAATGTGGAGAGGAAGATCTTG
GACCTTTACAGCCTTAATAAGATTGTGATGGAGGAAGGTGGATATGAAGCAATCTGCAAGGACCGTCGGTGGGCCCGTGTTGCCCAACGCCTTAATTACCCATCAGGGAAAAACATTGGTTCCCTGCTTCGGTCCC
ACTACGAACGGATCATTTACCCTTATGAAATATTTCAGTCTGGAGCTAATCTTGTGCAATGTAACACAGATCCATTTGACAGTGAAGAAAGAGACAAGGAATACAAACCTCATAGCATTCCTCTTAGACAGTCTGTTC
AGCCATCAAAATTCAGTTGCTACAGTAGACGAGGAAAAAGGCTACAGCCTGAGCCAGAACCCACAGAAGAAGACATTGAGAAAAACCCAGAGCTAAAGAAACTACAGATATATGGGGCAGGCCCCAAGATGATA
GGCTTGGGCCTCAAGGCTAAGGAGAAAACTCTGAGAAAGAAAGATAGCAAACAACCAGATAAAGAAGAAGTGACATGCCCTGCAACTATTGTAGTGAAGGGGGAAGCCAGTGAGTTTGGAAAGGTGACATCAG
CATTTTCAGACAAGAACTTAAATCACAGTTTTGAACCTTGTATGAAGATGACTATGCAGCTACGGAACAATCACAGCAGTACCCAGTTTATGAATTCATATGTTTGCCGAATATGCTCTCGTGGGGATGAAGTCGATA
AGTTTCTGCTCTGTGATGGCTGCAGTGATAATTACCACATCTTTTGCCTATTGCCACCGCTTTCTGAAGTCCCCAAGGGAGTATGGAGATGCCCGAAGTGTATCTTAGCAGAGTGTAAGAGTCCCCCTGAAGCTTTTG
GCTTTGAGCAGGCTACACAGGAGTATACTTTGCAGAGTTTTGGTGAGATGGCTGACTCCTTCAAGGCTGACTACTTCAACATGCCTGTGCATATGGTACCTACAGAAGTTGTGGAAAAGGAATTCTGGAGGCTGGT
GAGCAGCATTGAGGAGGATGTGACAGTTGAATATGGAGCAGACATTCACTCCAAAGAATTTGGCAGTGGGTTTCCTGTCAACAATAGCAAATGGGACTTATCTCCTGAAGAAAAGGAATATGCTGCTTGTGGTTGG
AATCTCAATGTGATGCCAGTGCTGGATCAGTCTGTTCTCTGCCACATCAATGCAGACATCTCAGGCATGAAAGTGCCCTGGTTATATGTGGGCATGGTGTTTTCAGCATTTTGTTGGCATATTGAGGATCACTGGAGT
TATTCCATTAACTACCTGCACTGGGGTGAACCAAAGACCTGGTATGGAGTACCTTCGCTAGCAGCAGAACACTTAGAGGATGTAATGAAGAGACTTACACCAGAGTTGTTTGACAGCCAACCTGACCTCCTGCACC
AACTTGTCACTCTGATGAATCCTAACACTCTCATGTCACATGGAGTGCCAGTTGTTCGTACAAACCAGTGTGCAGGGGAATTTGTCATTACTTTTCCCCGTGCTTACCACAGTGGCTTTAACCAAGGCTATAACTTTG
CTGAGGCAGTGAACTTTTGCACTGCCGATTGGCTACCTGTTGGACGCCAGTGTATTGAACACTACCGCCGACTCCGGCGCTATTGTGTCTTCTCCCATGAAGAACTTATCTGCAAGATGGCTGCATTTCCAGAGAAG
TTGGATCTAAATCTAGCAGTGGCAGTTCACAAGGAGATGTTCATTATTGGTACAGGAGGAGCGACGCCTACGAAAGACCTTGTTGGAGAAGGGCATCACGGAAGCTGAGCGAGAGGCTTTTGAGCTACTCCCAGAT
GATGAGCGTCAGTGCATCAAGTGTAAGACCACGTGTTTCCTATCAGCCTTGGCCTGTTATGACTGTCCAGATAGTCTTGTCTGCCTTTCCCATATCAATGATCTGTGCAAGTGCTCAAGGAATCGACAGTACTTACGG
TACAGGTACACATTGGATGAACTGCCTGCTATGCTACAAAAGCTGAAGATTCGAGCTGAGTCATTTGACAACTGGGCCAACAAAGTACAAGCAGCCCTAGAGGTGGAGGATGGTCGGAAACGGAGTTTTGAAGA
GCTAAGAGCATTGGAATCTGAGGCTCGTGATAGACGATTTCCTAATAGTGAGCTGCTTCAGCGATTGAAGAAATGCCTAACTGAAGCAGAGGCTTGTATTTCCCAAGTTCTTGGACTGATCAGTAACTCAGAGGAC
AGGTTACAAACACCGCAAATCACCCTTACTGAACTACAGCTTCTTCTCAAACAGATGGGCACTCTGCCCTGTACCATGCATCAGATTGATGAAGTTAAGGATGTTCTTCAACAGGTGGAATCATACCAAATTGAGAC
CCGTGAGGCTCTAACTTCACTGCCTTACAGTTTAGAAATACTTCAGTCCCTGATGGAAAAGGGTCAGCAGCTTCGTGTGGAAGTGCCTGAAGCTCATCAACTTGAAGAGCTGTTGGAGCAGGCACAATGGCTAGAT
CAAGTGAAGCAGGCCCTGGCTCCCTCAGGCCAGAGGCACTCTTTGGTTATCATGAAGAAGCTTTTAGTTATGGGAACCAAGGTAGCCTCCAGCCCTTCTGTGAACAAAGCTCGAGCAGAGTTACAAGAACTGCTG
ACTATTGCAGAATGTTGGGAGGAAAAGGCCCATTTTTGCTTAAAAGCTAGCCAAAAGCATTCTCCAGCCACTTTGGAAGTTATAATACGTGAGGCAGAAAACATCCCTGTCTACCTACCTAACATCCAGTCTCTGAA
GGAAGCTCTGACTAAGGCACAGGCTTGGATTGCTGATGTGAATGAGATTCAAAATGGTGATCACTACCCCTGTTTGGATGATTTGGAGGGCCTTGTAGCTGTTGGTCGGGATTTACCTGTGGAGCTAGAGGAATTGA
GACAGCTGGAGAATCAGGTACTAACAGCACATTCCTGGAAAGAGAAAGCTTCTAAGACTTTTCTTAAGAAGAATTCTTGCTACACATTGTTGGAGGTTCTCTGCCCATGTGCAGATGCTGGTTCAGTCAGCACCAA
GCGCAGTCGGTGGATTGAAAAGGAGATGGGATTATACAAATATGACACAGAATTGCTGGGGCTGTCTGCACAGGACCTCAGGGACCCAGGGTCTGTGATCATGGCCTTTAAGGAAGGGGAAGAGAAAGAAAAGG
AAGGTATCTTGCACTTACGTCACATCAACTCAGCCAAGCCCAGTCCAATGTCCTCATCGATGAATGCCTCTGCAACCTCCATCTGTATATGTGGGCAGGTCTGTGCAGGAGTGGAATCTCTGCAGTGTGATCTTTGTC
ATGACTGGTTCCATGGACAATGTGTGACTGTACCCCACCTACTCAGCTCAGTAAGGGCCAGTCATACCTCATCTCAACTGCTAGCTTGGTGGGAATGGGACACCAAGTTCCTGTGCCCACTGTGCATGCGATCAAG
GCGTCCACGCCTGGAGACCATCTTGTCATTGCTGGTAGGACTACAGAGACTGTCTGTAAGGCTACCTGAGGGTGAGGCCCTACAGTGCCTCACAGAGAGAGCCATTGGTTGGCAAGGCCGTGCCAGGCAGGCTCT
AGCTTCTGAGGATGTGACTGCTTTGTTGAAACAGCTGGAAAAATCCCGCCAACAACTACAGGATGAATTGAGACATAAGAAGCCCCCTACCTTGCCTTCAGGCTTTGCCTTTGACTGTCTCACAGAGAATAGTGGC
AAGGATATCCTTAAGGAAGAGGAGGAGTTGGTACTAAATGAAGAAAGGATTAAAAGTTCTGAGAAGATTGTTCCAAAGGAAAGTTCATGTAAAGGAGATAAGGAACTGCTGCCTTCACTTTTATCCCAATTAACTG
GCCCAGTACTGGAACTGCCTGAAGCCACCCGGGCACCCCTAGAGGAGCTTATGATGGAAGGAGATCTACTTGAAGTTACCCTAGATGAGAACTACAGCATTTGGCAGCTGCTGCAGGCTGGGCAGAATCCCAATCT
AGAGCGCATTCATACACTTCTGGAGCTTGAGAAGCCTGAAAACCCAGGAAATTGGTCAGAGGAACAGACACCAGAGAGGCGACGGCAGAGGAGGCAAAAGGTGGTTCTGAGTAGAAAGGGTGAGGATTTTACT
CAAAAAGAGCTAGAATCAAAAAGAGTCAAGAGTTCAAGGATTAAGCCTAAGGAAGAAAAGTTTCAAAAGCCAATACTTGGGGATAATGTGCTGTACACACACCATACAGAACATACAACATCTTAAAAGAGCA
CATAAACAGTGTACAAGGAAAGGATCCAAGCCCTTCATCTTCTTTTCCTTCCTTAACACCTTTACTACACCTATCTTATTTTCATCAGCAAAAATTAGATTACAAGGATGACGATGACAAGTAA

Vector information pCS2+-KDM5D A) Vector map of pCS2+ backbone including mouse KDM5D exonic sequence. Orange, ORF; blue, Origin of replication; green, Primer binding sites. B) Insert sequence coding for full-length mouse KDM5D with single FLAG tag sequence at the 3' end of insert.

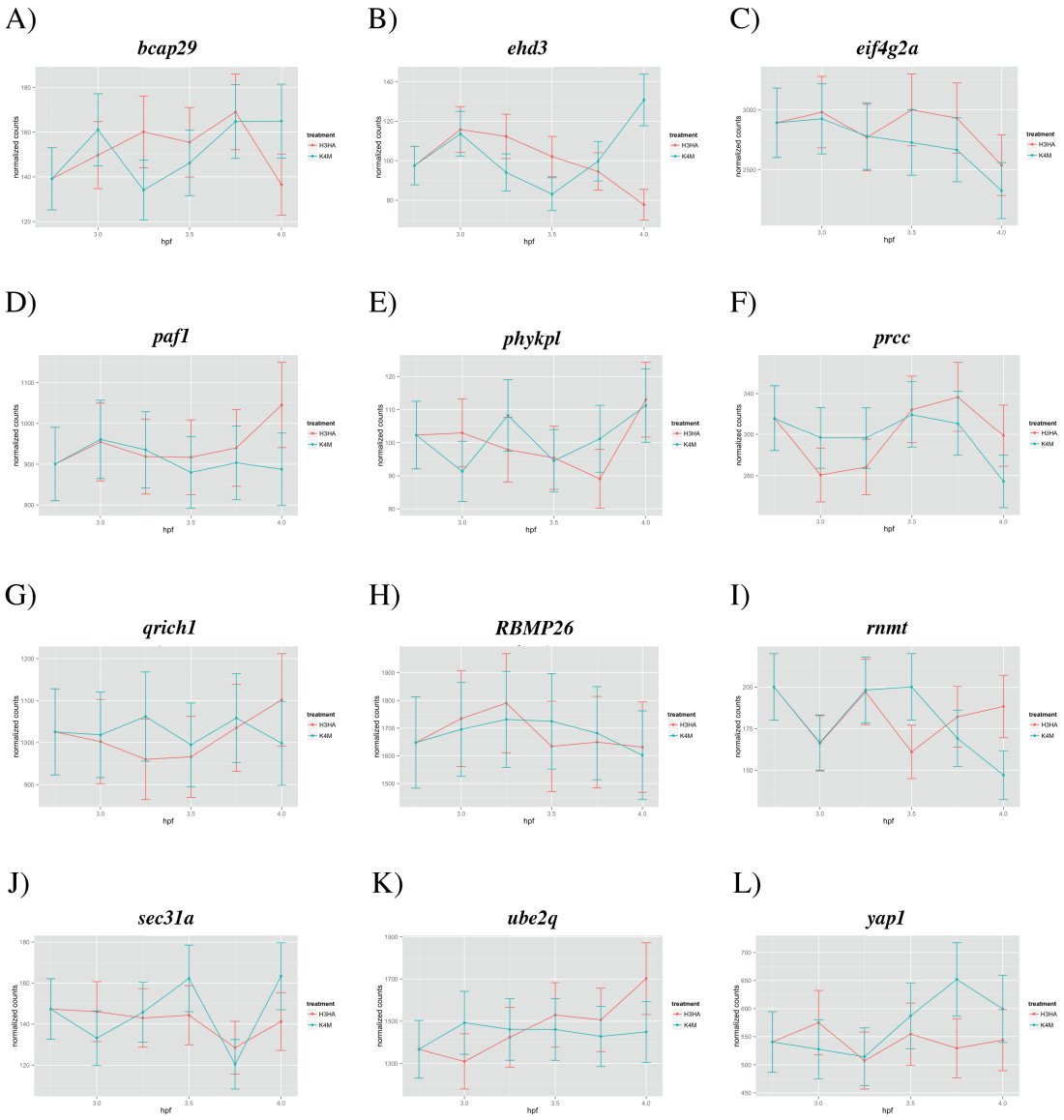
A)



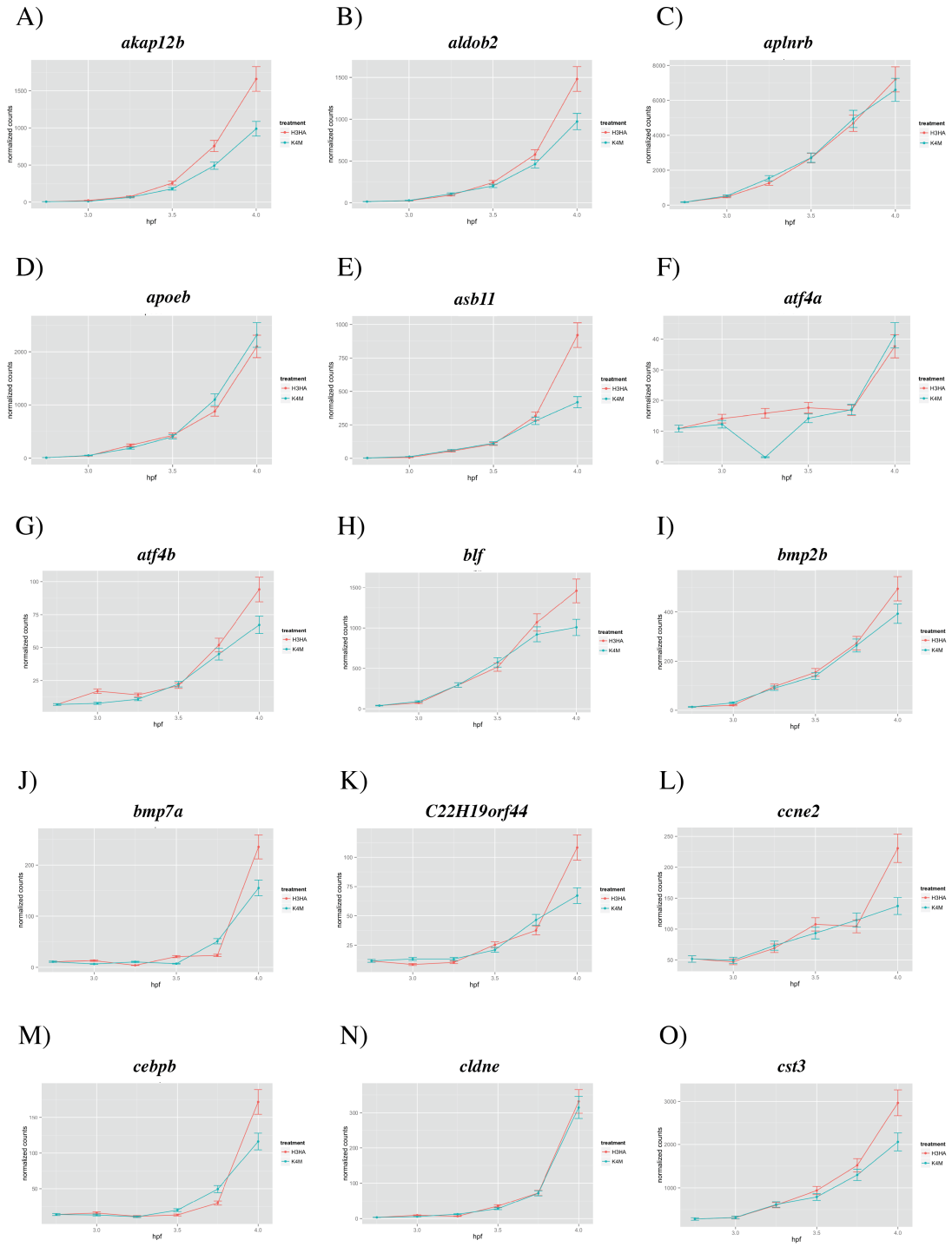
B)

ATGACCCAGCAGGGCCCGGCCGAGTTTACCCCACCGCCGGAGTGCCCGGTGTTCGAGCC-
CAGTTGGGAGGAATTTAAAGACCCGTTTGCGTTTATCAATAAAATCCGACCTATCGCCGAGAAAACCGGGATCTGCAAGGTTCGACCGCCTCCGGACTGGCAGCCCCCATTCGCATGTGACGTCGACAGACTTCAT
TTTACTCCACGGATCCAGAGGCTAAATGAGCTAGAGGCTCAAACCAGAGTCAAACTCAACTTCTTGGATCAGATCGCCAAATTTTGGGATCTACAAGGATGCACGCTCAAAATTCCCCACGTGGAGAGGAAAATCC
TCGATTTGTACCAGCTCAACAAATTGGTTGCAGATGAGGGTGGGTTCGATTTAGTTTGTAGAGAGAGGAGATGGACCAAGATCGCCATGACGATGGGTTTTGCCCCAGGCAAAGCTGTGGGCTCCCACCTGAGAG
CCCATTACGAGAGGATCCTGTACCCGTACAACCTCTTTCAGTCTGGGACAAATCTGCTGTGTGTCTGCAGAAGCCTGGGGATGAGGTGAATGACATTGACAAAGAGTATAAACCCCACGACCTACTGCAGAGGCAAA
ATGTTCCACTTCAGCCCAGCAACACAAGCGCACCGGCTAGACGAGCCAAGAGGATGAAAACAGAGTCTGGCTGTATTAAGTCAGAGGAAGGTGAGGGGGTTGAGAACAAGCCAAATCTCAGGAGGAGGATGGG
TTCTTTCGTTGTCAAGACTGAACCAAAAAAGGAGATCCCTATACAGGTGAAAGAGGAACCTGTGGAGATAAAAGAGCTGAATCCTGAACCGGAGAAAAGCAAACCGAAGAAGAAAAACATCCCCCCTCCTCCAG
TCAGCATGGTGGACCTCTATGTGTGTTTAGTGTGTGGAAAGGGAAACGACGAGGATCGGTTGTTGCTATGCGACGGTTGTGATGACAGCTACCACACATTCTGCTTAATCCCGCCCCTAACCGATGTTCCGAAAGGT
GATTGGCGGTGCCCTAAATGTCTGACTCAGGAGTGCTGTAAACCTCAGGAGGCATTTGGCTTCGAGCAAGCACACAGAGACTATACTCTCAAAGCCTTTGGGGAAATGGCAGACTCCTTCAAGTCAGATTATTTCA
ACATGCCTGTTCATATGGTACCCACTGAATTAGTGGAAAAAGAGTTCTGGAGGTTGGTCGGCACCATTCAGGAGGATGTGACAGTGGAATATGGCGCTGACATTGCCTCAAAAGAGTTCGGAAGTGGGTTTCCAAT
CAAAGGTGGAAGATTTAAGATTGCACCGCACGATGAGAAATATCTTCAGTGTGGTTGGAACCTGAACAATATGGCCATGATGACACCATCAGTGCTGACACACGTCACTGCTGATATCTGTGGAATGACTTTGCCAT
GGCTGTATGTGGGCATGTGCTTCTCATCGTTCTGCTGGCACATTGAGGACCACTGGAGCTATTCCATCAACTACCTGCACTGGGGAGAACCTAAGACCTGGTATGGTGCTCCGGGTTTTGCTGCTGAGCAGTTGGAA
GCCGTCATGAAGAAACTGGCCCCAGAACTATTTGATTCTCAGCCCGACCTTCTACACCAGCTTGTCACCAATCATGAATCCTAATACACTCATGGCCCATGGAGTACCAATTTACAGAACCAATCAGTGCGCAGGAGA
GTTTGTCATTACTTTTCCGCGATCCTATCACAGTGGCTTCAACCAGGGCTTCAACTTTGCAGAGGCTGTGAACTTTTGCACAGTGGACTGGATGCCACTGGGGCGTCAGTGTGTGGATCACTACCGCCAGCTGCATC
GCTACTGTGTGTTTTCTCATGATGAGATGGTGTGTAATATGGCCATGAAGGCTGACTGTCTTGATGTGGTTCTGGCGTCAGCCGTACAGAAGGACATGCAGCTCATGATCAAAGAAGAGAGAGAGCTGCGGGAGAA
AGTCCGTAAGATGGGGGTGGCCCAGTGCGAGTTGTTTCAATATGACTTGCTGGCAGATGATGAGAGACAGTGTGTGAAGTGTCGCACCACCTGTTACCTGTCTGCTCTCACCTGTCCCTGCAGACCAGGAGTCCAG
GTGTGTCTGTACCACACCCACGACCTCTGCTCCTGCCCCATCTCCAACTACACTCTCAATTATCGTTTCACACTGGATGACCTTTACCCGATGATGAACGCTGTGCGGCAGAGGGCAGAGTATTACGATGATTGGGC
CTCAAGAGTGACAGAGGTCATGGAAGCAAAGCTAGACAAGAAACGCAATGTCACCGTGTTTCGCACACTTCTGGAGGAGTCCAATGAGCAGTCTTTCCCAGAGAATGACCTGTTACGTCAGCTGAGGCTGGTTA
CTCAAGATGCTGAGAAATGCTCATCTGTTGCCCAGCAACTGCTAAATGGCAAGCGGCAGACCAGGTACCGCACAGGCAAGGCTAAATCTCCTAATCAGCTCACAGTGGAGGAAATGAGATCATTCGTCCGCCAGC
TTTATAACCTCCCCTGCAGCCTGACACAAGCACCGCTTCTCAAGGAGCTGCTTAACAGCATAGAGGATTTCCAGCAGCACAGTGAGAAGCTTCTATCTGATGAAGTGAGCGCAGATGCAGTGAGCGAGATCGAGT
CTTTGCTGGAGGAAGGCTCTCAGTTCGATGTGTTTCTTCCGGAGCTGCCGCTGCTACGGGAGCGTCTGGAACAGGCTCGATGGCTCACCGGAGTTCACCAAGCAGAGGATCCCGTAGCAAACCCCTGCGGTCTCT
CTCTGGAGAGCATGCGGAGGCTGATCGACCGAGGGGTTGGACTGACGCCTCATCCCTCCATAGAACGCATGATGGCTCGGTTGCAGGAGCTGCTTACCGTCTCAGAGGAGCTGGAGGAGAATGCGCAGGCTCTGC
TGAAGGCCAGACCTCCAGAGAGTTTAGAGACGCTCTGCTCCATGTTGACTCAGGTGGAAGGTGTACCTGCATACCTACCAAACTGCCTCCTACTACAGGACACAGTGAACAGGGCTAAAGAATGGCTACAAGAAG
CAGAGAGTTTACAGGTGGGAGGTCAGGTGCCGGTTTTCAGCTCGCTCTCTGATATGGTTCTGAGAGCTCATTCTATTCCTGTGCGACTGGAGCCCCTGGACCAGCTGGAGGTTCAGGTATCAGAGGTGCAGTCCTG
GAAGGAGACGGCAGCTAAAACCTTCCTAATCAAGAGCTCCCCCTTCACGCTGCTGGAGGTCTTATGTCCTCGATGGGAAATGGGCAGCTCTTTGAAAAAGAAGAAGATGAGAAAGATGAAGGGCGACTGTGTGT
CATCAGGGAAGAAGAAGTTTGTGAAGCTGGACAGTATGAGTGACGTCGAGAGAGCACTTTCAGACAGTAAAGACTCTGCATCTGCTATGTTTACTCTGGCCGAGGTGCGTTTAAAAGAGCTGGAGTCACTCTGCT
CTCTGCGTGCATCGAATGAATCCAAGCTCCTGCCCACAGCCGACTGCGCCTCGCTCAAAGTGTGTGTGTGTCAGAAACCAGCCATGGGCGCCATGCTGCAGTGTGAGCTGTGCCGGGACGCTTTCCACAGCGTGT
GTGTGCGGGGTCCCTCAGACCCCCTGGACCCTGAAGCGTGGCTCTGTCCATTATGCCTGAGGTCGACCAAACCGCCTCTGGATAAGATAAGGAGCCTGCTCTCTTCACTGCAGCGGATCCGCGTGCGCTTACCTGA
AGGCGATGCTCTACGTTATATGATCGAGCGCAGTGTGAGCTGGCAGCGGAGGGTGCGAGAGGTTATAGACTCTTATGGGCTCTCGTCGACCTCTCAGGATGGCAGAGGAGCTTCACCTTCACAGAACCTTCACAGG
AGCACTGGCCACAGCAGGAAGCAGCAGCTGTGTGGCTCTCCATCTCGGTGTCAGGACTGGGCAGTTTCAGGGCAGGAGCAGACAGTGTTTACACAGAACAGCGCTGTATTCCTCTCCAGGGTCTCAGTCCAGA
GCTGGAGGAGCTGATGGTGGAGGGTTTGGTTCTGCAGGTTTCTTTGCCTGAGACCCAGCAGCTCTACAGGCTTCTGCTGTCTGGCCCGCCAACCACAAACACGTCCCACGCAGAACACAAGTCCTACCTCACACC
ACCACAGACAGAGACAGATGCAATCACTTCAGCAGAAAAGAAAGCCAAGCGGCGAATGAACCGAGACGAGGTGGATATCAGGGAAAGAGGGACGAAGCTGAAAAGCAAGAAGCAGAGGATGATGGGTGTGGA
GAAGCGGAGAGAGGAAAGCAGCGTCAGTCTCTGCGTCAGACATGTCTCAGTCAGAAGACTCAGAGGAGGACATGACTCTGTGTCCTGCAGAGAGCTGTCTCCAGCCTGAGGGAGAAGAGGTGGACTGGGTG
CAGTGTGATTGCTGTAACCGCTGGTTTCATATGATTTGTGTGGGAGTCTCAGCAGAACTGGCAGCCGAGGAGGATTACATGTGTGTGAGCTGCAGCACAAGTCACATGGACCGCCGGAAGGATTACAAGGATGAC
GATGACAAGTAA

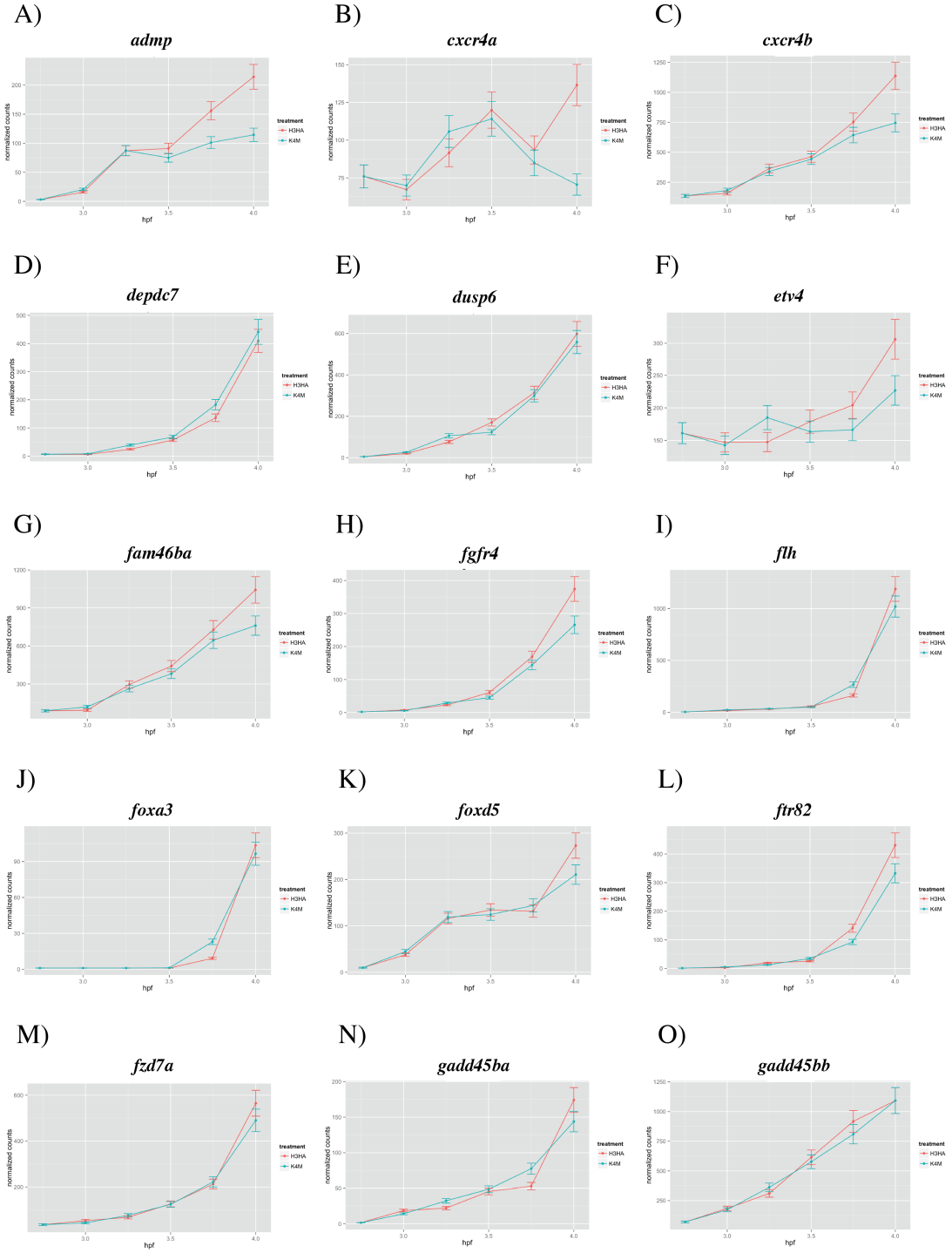
Vector information pCS2+-zfKDM5Bb A) Vector map of pCS2+ backbone including zebrafish KDM5Bb exonic sequence. Orange, ORF; blue, Origin of replication; green, Primer binding sites. B) Insert sequence coding for full-length zebrafish KDM5Bb with single FLAG tag sequence at the 3' end of insert.



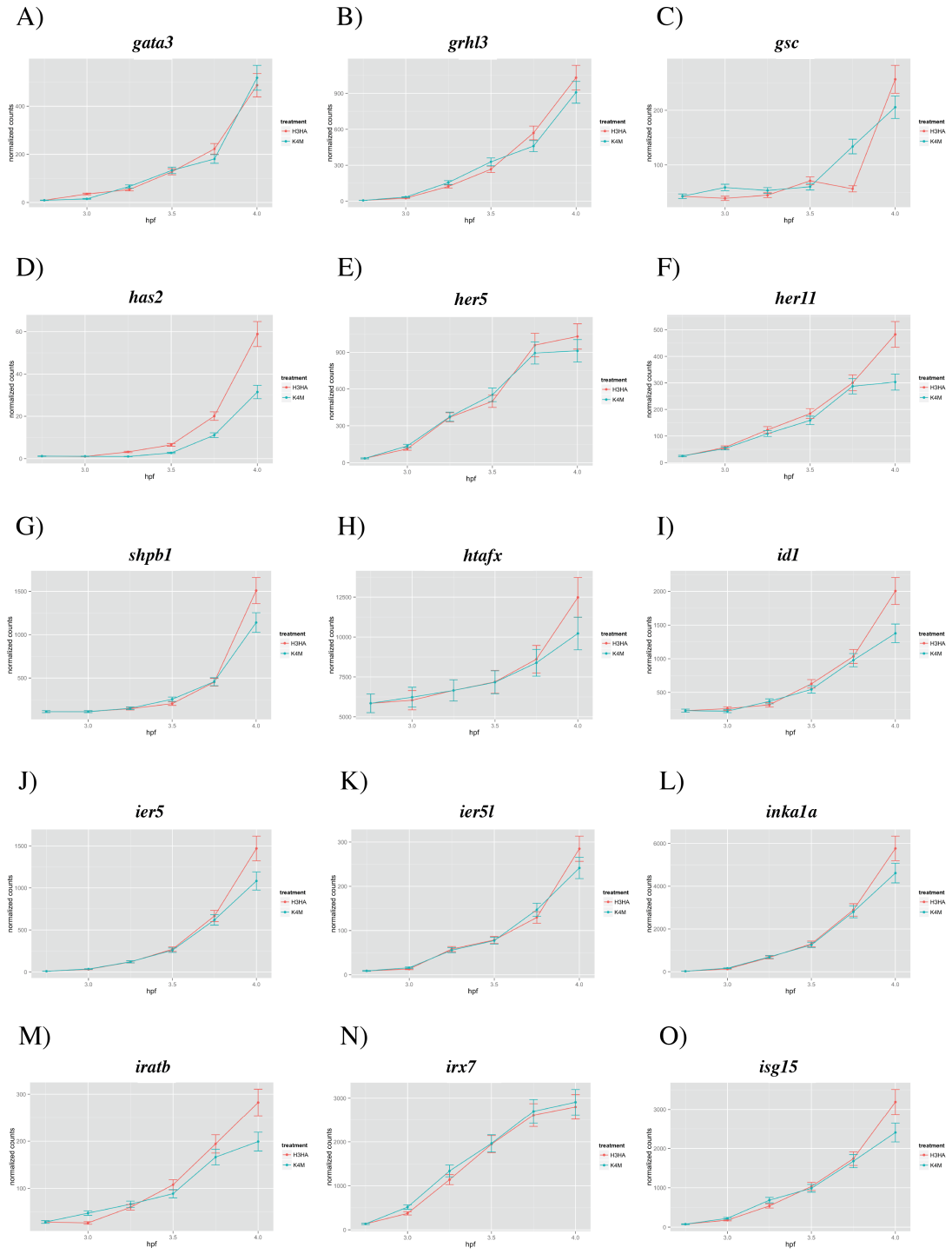
NanoString timeseries Normalizers Shown are 12 genes chosen to be normalizers for RNA levels over early embryonic development. Plots represent their normalized RNA count values measured by NanoString assay over developmental time from RNA samples of H3HA (red) or K4M (blue) introduction. Note the changed color code compared to figures in results. Error bars indicate technical variation of 10% of respective count value. This was estimated from variation in assay-internal positive controls as well as normalizer gene variation.



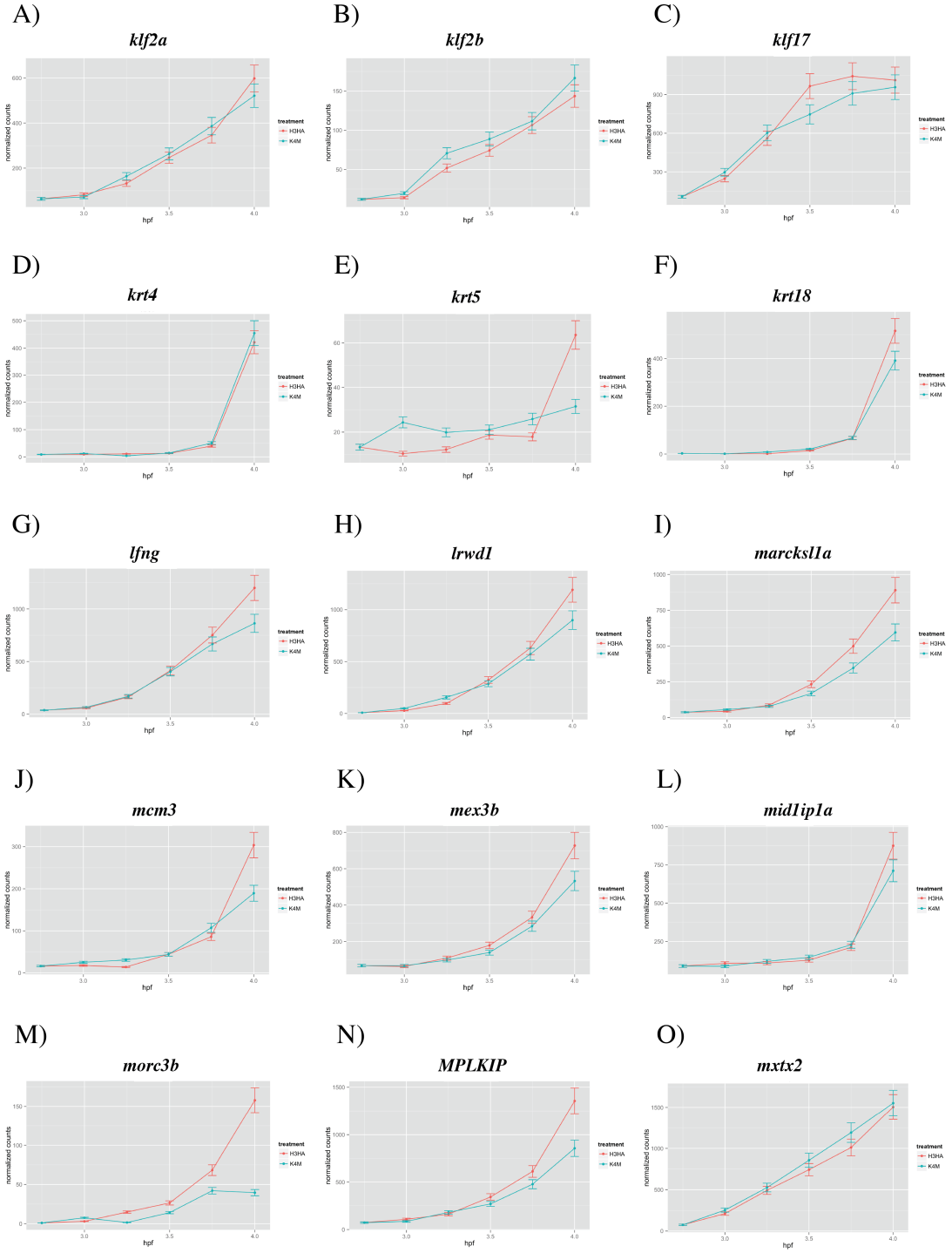
NanoString timeseries K4M introduction I Depicted are normalized RNA counts of zygotic activated genes over six stages at zygotic genome activation as measured by NanoString (I of VI). H3HA introduction (red) is compared to K4M introduction (blue). Note the changed color code compared to main matter figures. Error bars indicate technical variation of 10% of respective count value. This was estimated from variation in assay-internal positive controls as well as normalizer gene variation.



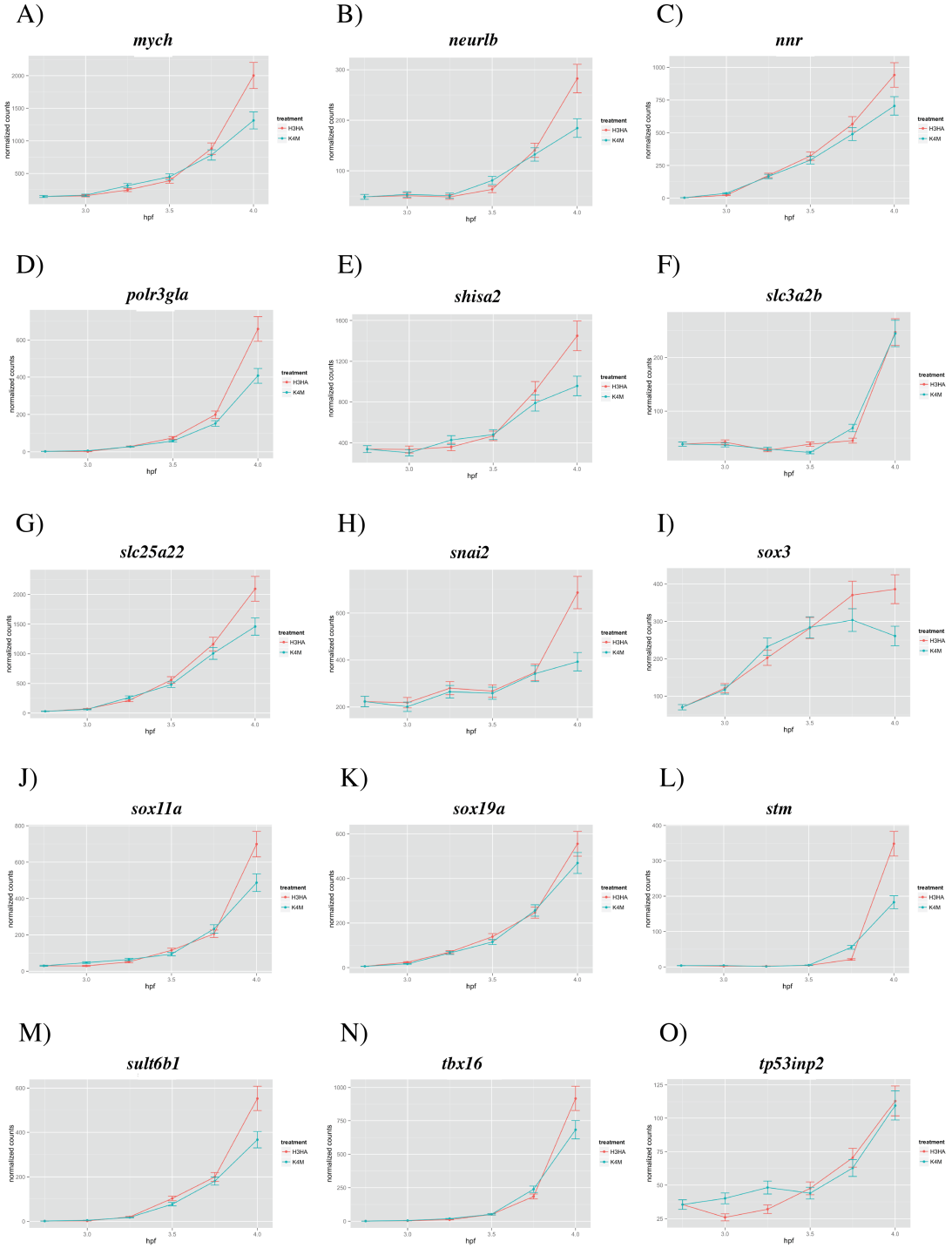
NanoString timeseries K4M introduction II Depicted are normalized RNA counts of zygotic activated genes over six stages at zygotic genome activation as measured by NanoString (II of VI). H3HA introduction (red) is compared to K4M introduction (blue). Note the changed color code compared to main matter figures. Error bars indicate technical variation of 10% of respective count value. This was estimated from variation in assay-internal positive controls as well as normalizer gene variation.



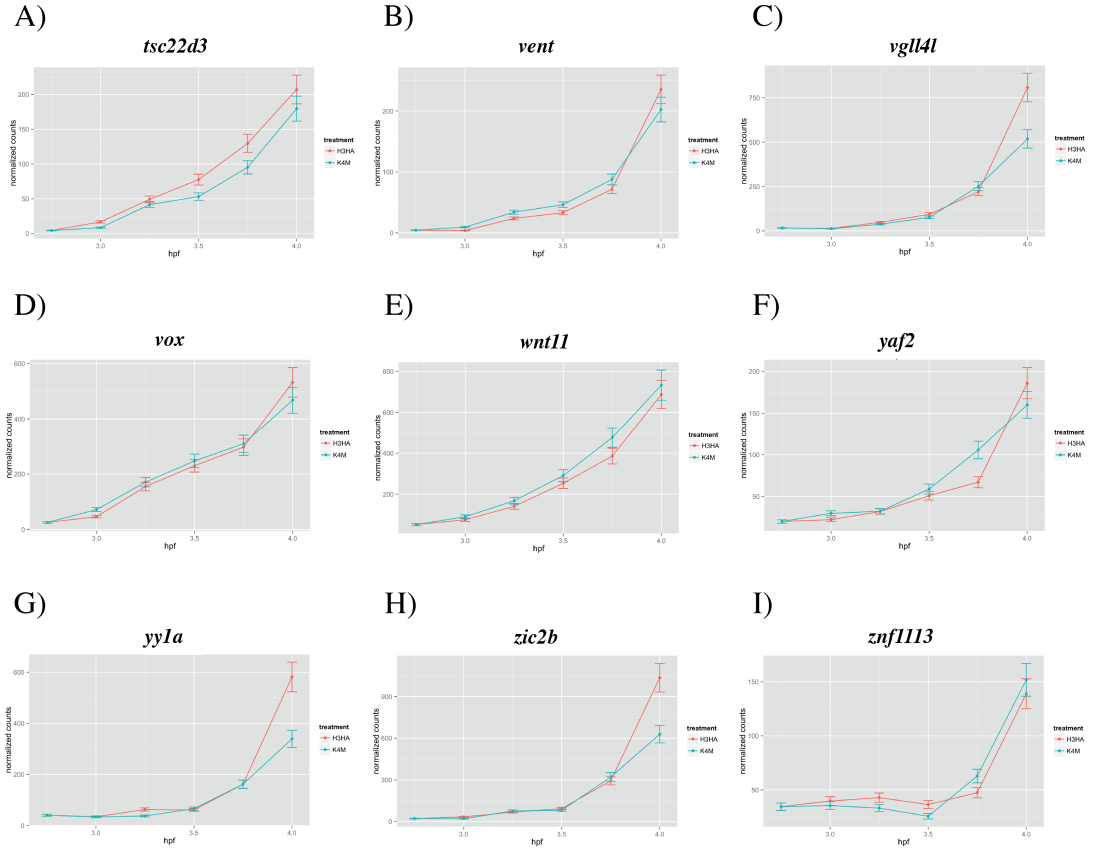
NanoString timeseries K4M introduction III Depicted are normalized RNA counts of zygotic activated genes over six stages at zygotic genome activation as measured by NanoString (III of VI). H3HA introduction (red) is compared to K4M introduction (blue). Note the changed color code compared to main matter figures. Error bars indicate technical variation of 10% of respective count value. This was estimated from variation in assay-internal positive controls as well as normalizer gene variation.



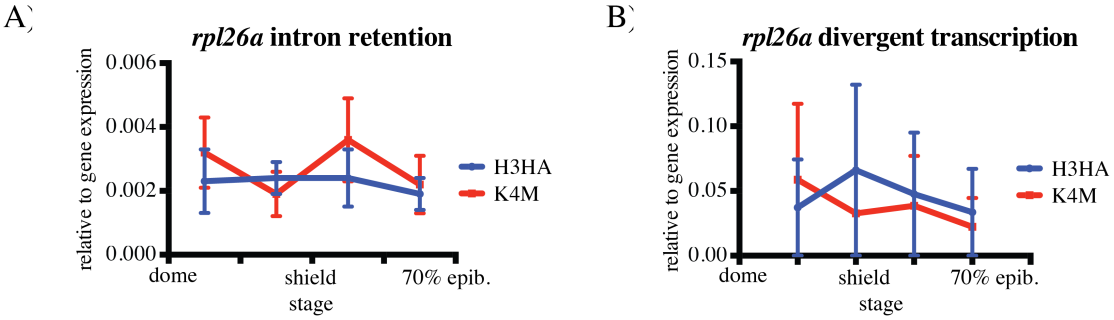
NanoString timeseries K4M introduction IV Depicted are normalized RNA counts of zygotic activated genes over six stages at zygotic genome activation as measured by NanoString (IV of VI). H3HA introduction (red) is compared to K4M introduction (blue). Note the changed color code compared to main matter figures. Error bars indicate technical variation of 10% of respective count value. This was estimated from variation in assay-internal positive controls as well as normalizer gene variation.



NanoString timeseries K4M introduction V Depicted are normalized RNA counts of zygotic activated genes over six stages at zygotic genome activation as measured by NanoString (V of VI). H3HA introduction (red) is compared to K4M introduction (blue). Note the changed color code compared to main matter figures. Error bars indicate technical variation of 10% of respective count value. This was estimated from variation in assay-internal positive controls as well as normalizer gene variation.



NanoString timeseries K4M introduction VI Depicted are normalized RNA counts of zygotic activated genes over six stages at zygotic genome activation as measured by NanoString (VI of VI). H3HA introduction (red) is compared to K4M introduction (blue). Note the changed color code compared to main matter figures. Error bars indicate technical variation of 10% of respective count value. This was estimated from variation in assay-internal positive controls as well as normalizer gene variation.



Splicing and divergent transcription are not changed upon H3K4me3 reduction
A) RT-qPCR experiments showing the relative expression of *rpl26a* containing the first intron compared to *rpl26a* mRNA level at stages dome to 70% epiboly. B) RT-qPCR experiments showing the relative expression of upstream regions of *rpl26a* locus compared to *rpl26a* mRNA level at stages dome to 70% epiboly. H3HA overexpression samples are depicted in blue, K4M overexpression samples in red. n = 3, error bars = s.d.

ChIP-Seq libraries

ChIP-Seq experiment and library statistics

genotype/ treatment	experiment	stage	Total library	unique mappable reads
H3HA	H3K4me3 ChIP	high	52,064,470	6,165,525
H3HA	ChIP Input	high	60,570,494	9,059,513
K4M	H3K4me3 ChIP	high	50,811,576	6,410,505
K4M	ChIP Input	high	52,685,461	8,616,120
H3HA	H3K4me3 ChIP	oblong	49,730,067	3,927,642
H3HA	ChIP Input	oblong	47,304,317	4,700,641
K4M	H3K4me3 ChIP	oblong	51,423,111	5,050,954
K4M	ChIP Input	oblong	50,462,787	6,456,628

H3HA	H3K4me3 ChIP	oblong2	62,136,515	13,954,143
K4M	H3K4me3 ChIP	oblong2	73,118,381	6,910,026
H3HA	H3K4me3 ChIP	sphere	84,226,793	11,770,538
K4M	H3K4me3 ChIP	sphere	71,015,396	12,482,032
H3HA	H3K4me3 ChIP	dome	84,432,377	12,193,878
K4M	H3K4me3 ChIP	dome	70,543,169	9,257,326
wildtype	H3K4me3 ChIP	oblong	47,513,081	1,462,206
wildtype	ChIP Input	oblong	50,966,520	5,470,004
mll1 mutant	H3K4me3 ChIP	oblong	34,115,775	1,351,487
mll1 mutant	ChIP Input	oblong	63,551,908	10,012,384
wildtype	H3K4me3 ChIP	dome	30,297,685	6,102,726
wildtype	ChIP Input	dome	97,860,279	22,383,030
mll1 mutant	H3K4me3 ChIP	dome	53,001,118	11,304,856
mll1 mutant	ChIP Input	dome	75,892,432	19,262,817
wildtype	H2A.z ChIP	dome	87,807,652	37,528,879

Extended experimental discussion

Increased fraction of reads observed in MNase-ChIP-Seq experiments

When performing ChIP-Seq experiments, I obtained Sequencing data comprising of 30 to 50 million reads of 75bp length per sample (see Table 1). Surprisingly, of these reads, only a fraction of total reads could be aligned to a single position in the genome (around 3% for oblong stage and 20% for dome stage), while a much larger fraction of reads was assigned to multiple regions with the same probability. Since this phenomenon was also observed in Input chromatin samples taken before immunoprecipitation procedures (Table 1), it is unlikely to be caused by an antibody bias towards specific genomic features. Additionally, these results cannot be explained by poor quality of the obtained sequencing reads, as more than 95% of the reads showed good sequencing quality throughout their length of 75 basepairs (data not shown).
To address whether these results could be an artifact from aligning the reads to zebrafish genome, I repeated alignment with two different bioinformatics algorithms (bwa and bowtie). Additionally I used two different options for mapping by bowtie. The default setting uses the whole read for alignment, and adds a penalty for each mismatch. The 'seeding-method' uses a subregion of the read for initial perfect matching, and expands the matching in both directions as far as possible. These three different mapping strategies were performed on all samples. However, none of these attempts increased the fraction of uniquely mappable reads significantly. I therefore concluded that the high fraction of multi-mappers does not result from errors within the analysis or chromatin immunoprecipitation. It might rather result from a bias in chromatin preparation, such as a bias in MNase restriction, or an influence of chromatin architecture on DNA accessibility in early zebrafish embryos. Since no attempt improved data quality, I proceeded with the results obtained by the experimental procedures described above.

Published bioinformatics tools for quantitative ChIP-Seq are not applicable to analyze H3K4me3 reduction in mll1 mutant embryos

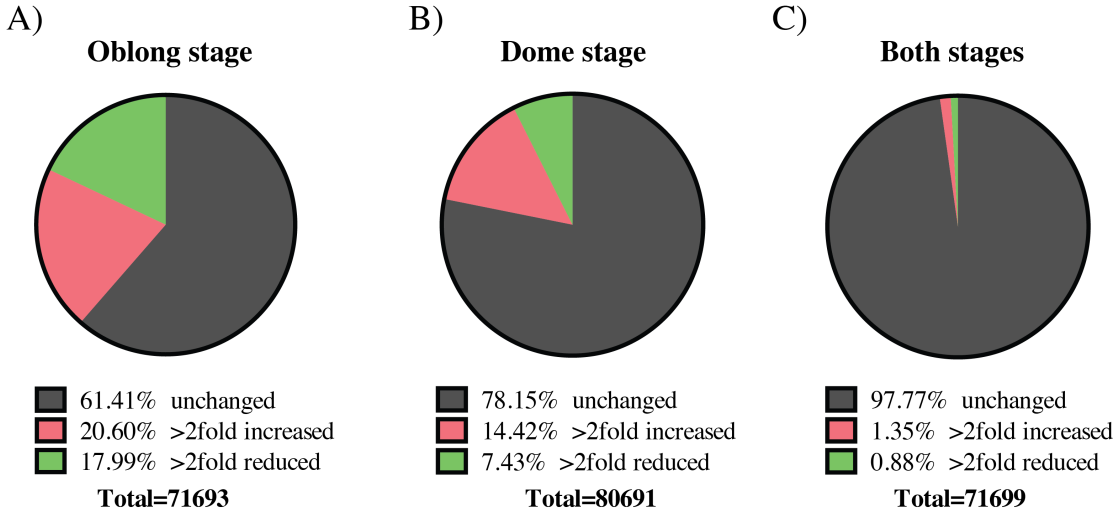
In order to determine whether H3K4me3 levels at specific promoters are affected upon H3K4me3 interference, I aimed to compared the ChIP-Seq data from unaffected and globally reduced H3K4me3 samples using algortithms that analyze local enrichment between different samples. I aimed to implement previously established algorithms that

analyze local enrichment between different samples. While a number of algorithms are reported for differential ChIP peak analysis, not all are applicable for the experimental setup presented here (Steinhauser et al., 2016).

Specific tools were designed to identify differential transcription factor binding (e.g. ChIPComp, MMDiff, DBChip, Homer, ODIN). These rely on motif binding information and defined local enrichment, which does not apply to histone modification data. Other algorithms (ChIPDiff with edgeR option, diffreps, RSEG, histone HMM) rely on providing biological replicates for each timepoint, to assess biological variation over which differentially enriched regions could be identified. However, since H3K4me3 is established increasingly over time, different time-points show too strong intrinsic variation to allow for estimation of biological variation. Therefore, our time-point data cannot be used as replicate samples.

Two different bioinformatics algorithms are applicable to the experimental setup here: DiffBind and MA-Norm. Both estimate biological variation from within one dataset and therefore do not rely on providing replicate data for each timepoint. However, Diffbind was unable to identify significant differences between *mll1* mutants and wildtype samples. This can be explained by too low numerical counts for analyzed promoter regions. The resulting estimated biological noise within the samples is too high to call significant differentially bound regions between ChIP samples. Diffbind uses DESeq2 as an algorithm to estimate biological variation, which is designed for transcript analyses and thus relies on a higher number of reads bound to a region considered positive (e.g. over the whole distance of gene bodies). Therefore I decided to use MA-Norm to identify differentially bound regions. MA-Norm compares two samples by assuming that most peaks do not change in intensity. This allows for normalization of peaks by average difference, and identification of significantly different regions after normalization.

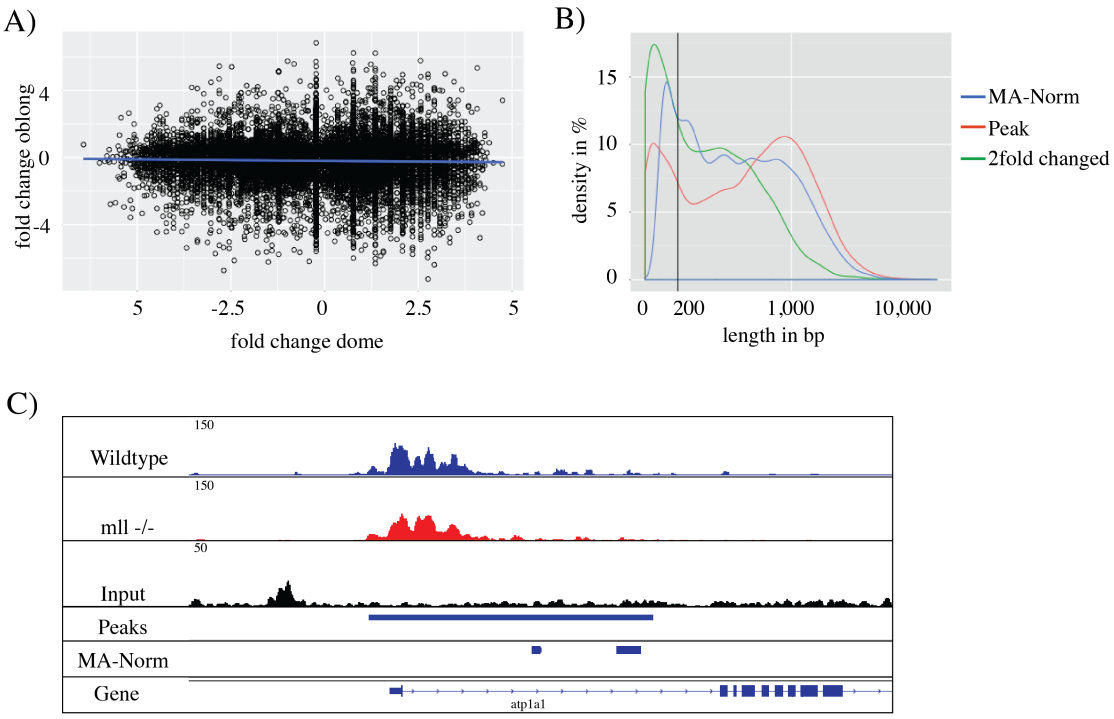
I performed MA-Norm analysis comparing *mll1* mutant samples with wildtype samples at oblong and dome stages. At oblong stage, MA-Norm identifies 20.60% that are more than 2fold increased, and 17.99% that are more than 2fold decreased in H3K4me3 reduction in *mll1* mutants (Figure **??** A). At dome stage, 14.42% are 2fold increased, and 7.43% are decreased in H3K4me3 enrichment (Figure **??** B). This corroborates the observation that H3K4me3 levels in*mll1* mutants are rather increased than decreased. However, only 1.35% of regions identified by MA-Norm are increased for H3K4me3 enrichment in both stages, and only 0.88% are reduced (Figure **??** C). This suggests, that the results obtained by MA-Norm analysis do not correlate between different de-



Identification of H3K4me3 level changes by MA-Norm Depicted are the fractions of MA-Norm-identified regions that are unchanged (grey), more than 2fold increased (red) or more than 2fold decreased (green) in *mll1* mutant embryos at oblong (A) and dome stage (B), as well as the fractions changed in both stages (C).

velopmental stages. Indeed, no correlation can be seen when fold changes at common regions were compared to each other (Figure ?? A).

MA-Norm bases the differential binding analysis on peak regions identified by other algorithms. In my case, MACS2 was used to identify significantly enriched regions, and MA-Norm to identify differentially bound regions. To understand whether the algorithm correctly reflects changes within the sample, I analyzed the size of identified regions by the MA-Norm algorithm. Interestingly, while MACS2 identifies a minimal peak size of 180 basepairs, regions that show differential binding by MA-Norm analysis are as small as 80 basepairs. This suggests that MA-Norm subdivides peak regions and thus does not reflect changes over the whole peak region. Indeed, as shown in Figure ?? C), MA-Norm identified regions that were smaller than identified peaks identified by MACS2. Furthermore, regions identified as significantly changed were enriched for regions of that smaller size. This suggests that MA-Norm subdivides identified peak regions, and favors smaller sized, less-intense peaks, which most likely have lower biological significance (Figure ?? B). Thus I find that MA-Norm is not applicable on the dataset analysed here, as too many regions of low intensity are overrepresented, and significantly called regions are not reproducible between stages.



MA-Norm does not correctly reflect changes in H3K4me3 levels A) The fold change as calculated by MA-Norm do not correlate for oblong and dome stage. B) Peak length distribution was plotted for MA-Norm analyzed regions (blue), identified peak regions by MACS2 (red) and regions identified as 2fold changed between wildtype and *mll1* mutants. Total MA-Norm regions and significantly changed regions are enriched for small-sized regions. C) Representative genomic region illustrating that MA-Norm focuses on small-sized regions subdivided from peak regions as identified by MACS2.

BIBLIOGRAPHY

Aanes, H., C. L. Winata, C. H. Lin, J. P. Chen, K. G. Srinivasan, S. G. P. Lee, A. Y. M. Lim, H. S. Hajan, P. Collas, G. Bourque, Z. Gong, V. Korzh, P. Aleström, and S. Mathavan
2011. Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome Research*, 21(8):1328–1338.

Abdellah, Z., A. Abbott-Ozersky, E. S. Lander, A. Abrajano, C. Q. Adams, Y. Sakaki, R. Kaul, R. Heilig, N. Shimizu, J. Battles, B. Birditt, M. Platzer, J. Bao, R. Subrak, H. Bloecker, E. Chen, R. Baertsch, and International Human Genome Sequencing Consortium
2004. Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011):931–945.

Agarwal, S. K. and R. Jothi
2012. Genome-wide characterization of menin-dependent H3K4me3 reveals a specific role for menin in the regulation of genes implicated in MEN1-like tumors. *PLoS ONE*, 7(5):e37952.

Ahmad, K. and S. Henikoff
2002. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Molecular Cell*, 9(6):1191–1200.

Allan, J., G. J. Cowling, N. Harborne, P. Cattini, R. Craigie, and H. Gould
1981. Regulation of the higher-order structure of chromatin by histones H1 and H5. *The Journal of cell biology*, 90(2):279–288.

Allfrey, V. G., R. Faulkner, and A. E. Mirsky
1964. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proceedings of the National Academy of Sciences*, 51:786–794.

Amodeo, A. A., D. Jukam, A. F. Straight, and J. M. Skotheim
2015. Histone titration against the genome sets the DNA-to-cytoplasm threshold for the Xenopus midblastula transition. *Proceedings of the National Academy of Sciences of the United States of America*, 112(10):E1086–95.

Andersen, I. S., A. H. Reiner, H. Aanes, P. Aleström, and P. Collas
2012. Developmental features of DNA methylation during activation of the embryonic zebrafish genome. *Genome Biology*, 13(7):R65.

Andreu-Vieyra, C. V., R. Chen, J. E. Agno, S. Glaser, K. Anastassiadis, A. F. Stewart, and M. M. Matzuk
2010. MLL2 Is Required in Oocytes for Bulk Histone 3 Lysine 4 Trimethylation and Transcriptional Silencing. *PLoS Biology*, 8(8):e1000453.

Ang, Y.-S., S.-Y. Tsai, D.-F. Lee, J. Monk, J. Su, K. Ratnakumar, J. Ding, Y. Ge, H. Darr, B. Chang, J. Wang, M. Rendl, E. Bernstein, C. Schaniel, and I. R. Lemischka
2011. Wdr5 Mediates Self-Renewal and Reprogramming via the Embryonic Stem Cell Core Transcriptional Network. *Cell*, 145(2):183–197.

Ausio, J., F. Dong, and K. E. van Holde
1989. Use of selectively trypsinized nucleosome core particles to analyze the role of the histone "tails" in the stabilization of the nucleosome. *Journal of molecular biology*, 206(3):451–463.

Bach, C., D. Mueller, S. Buhl, M. P. Garcia-Cuellar, and R. K. Slany
2009. Alterations of the CxxC domain preclude oncogenic activation of mixed-lineage leukemia 2. *Oncogene*, 28(6):815–823.

Balasubramanian, D., B. Akhtar-Zaidi, L. Song, C. F. Bartels, M. Veigl, L. Beard, L. Myeroff, K. Guda, J. Lutterbaugh, J. Willis, G. E. Crawford, S. D. Markowitz, and P. C. Scacheri
2012. H3K4me3 inversely correlates with DNA methylation at a large class of non-CpG-island-containing start sites. *Genome medicine*, 4(5):47.

Bannister, A. J. and T. Kouzarides
2011. Regulation of chromatin by histone modifications. *Cell research*, 21(3):381–395.

Bannister, A. J., R. Schneider, F. A. Myers, A. W. Thorne, C. Crane-Robinson, and T. Kouzarides
2005. Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *The Journal of biological chemistry*, 280(18):17732–17736.

Barski, A., S. Cuddapah, K. Cui, T.-Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao
2007. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*, 129(4):823–837.

Bartke, T., M. Vermeulen, B. Xhemalce, S. C. Robson, M. Mann, and T. Kouzarides
2010. Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell*, 143(3):470–484.

Benayoun, B. A., E. A. Pollina, D. Ucar, S. Mahmoudi, K. Karra, E. D. Wong, K. Devarajan, A. C. Daugherty, A. B. Kundaje, E. Mancini, B. C. Hitz, R. Gupta, T. A. Rando, J. C. Baker, M. P. Snyder, J. M. Cherry, and A. Brunet
2014. H3K4me3 Breadth Is Linked to Cell Identity and Transcriptional Consistency. *Nature*, 158(3):673–688.

Bernstein, B. E., M. Kamal, K. Lindblad-Toh, S. Bekiranov, D. K. Bailey, D. J. Huebert, S. McMahon, E. K. Karlsson, E. J. Kulbokas, T. R. Gingeras, S. L. Schreiber, and E. S. Lander
2005. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, 120(2):169–181.

Bernstein, B. E., T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber, and E. S. Lander
2006. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell*, 125(2):315–326.

Birke, M., S. Schreiner, M.-P. García-Cuéllar, K. Mahr, F. Titgemeyer, and R. K. Slany
2002. The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. *Nucleic acids research*, 30(4):958–965.

Blackwood, E. M. and J. T. Kadonaga
1998. Going the distance: a current view of enhancer action. *Science*, 281(5373):60–63.

Bledau, A. S., K. Schmidt, K. Neumann, U. Hill, G. Ciotta, A. Gupta, D. C. Torres, J. Fu, A. Kranz, A. F. Stewart, and K. Anastassiadis
2014. The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation. *Development (Cambridge, England)*, 141(5):1022–1035.

Bogdanovic, O., A. Fernandez-Minan, J. J. Tena, E. de la Calle-Mustienes, C. Hidalgo, I. van Kruysbergen, S. J. van Heeringen, G. J. C. Veenstra, and J. L. Gomez-Skarmeta
2012. Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Research*, 22(10):2043–2053.

Bogdanovic, O., S. W. Long, S. J. van Heeringen, A. B. Brinkman, J. L. Gómez-Skarmeta, H. G. Stunnenberg, P. L. Jones, and G. J. C. Veenstra
2011. Temporal uncoupling of the DNA methylome and transcriptional repression during embryogenesis. *Genome Research*, 21(8):1313–1327.

Bonhoure, N., G. Bounova, D. Bernasconi, V. Praz, F. Lammers, D. Canella, I. M. Willis, W. Herr, N. Hernandez, M. Delorenzi, and C. Consortium
2014. Quantifying ChIP-seq data: a spiking method providing an internal reference for sample-to-sample normalization. *Genome Research*, 24(7):1157–1168.

Burke, T. W. and J. T. Kadonaga
1997. The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. *Genes & Development*, 11(22):3020–3031.

Byvoet, P., G. R. Shepherd, J. M. Hardin, and B. J. Noland
1972. The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells. *Archives of biochemistry and biophysics*, 148(2):558–567.

Cao, F., Y. Chen, T. Cierpicki, Y. Liu, V. Basrur, M. Lei, and Y. Dou
2010. An Ash2L/RbBP5 heterodimer stimulates the MLL1 methyltransferase activity through coordinated substrate interactions with the MLL1 SET domain. *PLoS ONE*, 5(11):e14102.

Cao, F., E. C. Townsend, H. Karatas, J. Xu, L. Li, S. Lee, L. Liu, Y. Chen, P. Ouillette, J. Zhu, J. L. Hess, P. Atadja, M. Lei, Z. S. Qin, S. Malek, S. Wang, and Y. Dou
2013. Targeting MLL1 H3K4 Methyltransferase Activity in Mixed-Lineage Leukemia. *Molecular Cell*, Pp. 1–15.

Carrozza, M. J., B. Li, L. Florens, T. Suganuma, S. K. Swanson, K. K. Lee, W.-J. Shia, S. Anderson, J. Yates, M. P. Washburn, and J. L. Workman
2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell*, 123(4):581–592.

Chan, K.-M., D. Fang, H. Gan, R. Hashizume, C. Yu, M. Schroeder, N. Gupta, S. Mueller, C. D. James, R. Jenkins, J. Sarkaria, and Z. Zhang
2013. The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. *Genes & Development*, 27(9):985–990.

Chan, T.-M., W. Longabaugh, H. Bolouri, H.-L. Chen, W.-F. Tseng, C.-H. Chao, T.-H. Jang, Y.-I. Lin, S.-C. Hung, H.-D. Wang, and C.-H. Yuh
2009. Developmental gene regulatory networks in the zebrafish embryo. *BBA - Gene Regulatory Mechanisms*, 1789(4):279–298.

Chang, Y.-F., J. S. Imam, and M. F. Wilkinson
2007. The nonsense-mediated decay RNA surveillance pathway. *Annual Review of Biochemistry*, 76(1):51–74.

Chen, K., Z. Hu, Z. Xia, D. Zhao, W. Li, and J. K. Tyler
2016. The Overlooked Fact: Fundamental Need for Spike-In Control for Virtually All Genome-Wide Analyses. *Molecular and Cellular Biology*, 36(5):662–667.

Chen, Y., F. Cao, B. Wan, Y. Dou, and M. Lei
2012. Structure of the SPRY domain of human Ash2L and its interactions with RbBP5 and DPY30. *Cell research*, 22(3):598–602.

Cho, Y. W., T. Hong, S. Hong, H. Guo, H. Yu, D. Kim, T. Guszczynski, G. R. Dressler, T. D. Copeland, M. Kalkum, and K. Ge
2007. PTIP Associates with MLL3- and MLL4-containing Histone H3 Lysine 4 Methyltransferase Complex. *Journal of Biological Chemistry*, 282(28):20395–20406.

Churchman, L. S. and J. S. Weissman
2011. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*, 469(7330):368–373.

Ciruna, B., G. Weidinger, H. Knaut, B. Thisse, C. Thisse, E. Raz, and A. F. Schier
2002. Production of maternal-zygotic mutant zebrafish by germ-line replacement. *Proceedings of the*

National Academy of Sciences of the United States of America, 99(23):14919–14924.

Clarkson, M. J., J. R. Wells, F. Gibson, R. Saint, and D. J. Tremethick
1999. Regions of variant histone His2AvD required for Drosophila development. *Nature*, 399(6737):694–697.

Clouaire, T., S. Webb, and A. Bird
2014. Cfp1 is required for gene expression-dependent H3K4 trimethylation and H3K9 acetylation in embryonic stem cells. *Genome Biology*, 15(9):451.

Clouaire, T., S. Webb, P. Skene, R. Illingworth, A. Kerr, R. Andrews, J. H. Lee, D. Skalnik, and A. Bird
2012. Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes & Development*, 26(15):1714–1728.

Comb, M. and H. M. Goodman
1990. CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2. *Nucleic acids research*, 18(13):3975–3982.

Conerly, M. L., S. S. Teves, D. Diolaiti, M. Ulrich, R. N. Eisenman, and S. Henikoff
2010. Changes in H2A.Z occupancy and DNA methylation during B-cell lymphomagenesis. *Genome Research*, 20(10):1383–1390.

Creyghton, M. P., S. Markoulaki, S. S. Levine, J. Hanna, M. A. Lodato, K. Sha, R. A. Young, R. Jaenisch, and L. A. Boyer
2008. H2AZ Is Enriched at Polycomb Complex Target Genes in ES Cells and Is Necessary for Lineage Commitment. *Cell*, 135(4):649–661.

Denissov, S., H. Hofemeister, H. Marks, A. Kranz, G. Ciotta, S. Singh, K. Anastassiadis, H. G. Stunnenberg, and A. F. Stewart
2014. Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development (Cambridge, England)*, 141(3):526–537.

Dhar, S. S., S.-H. Lee, K. Chen, G. Zhu, W. Oh, K. Allton, O. Gafni, Y. Z. Kim, A. S. Tomoiga, M. C. Barton, J. H. Hanna, Z. Wang, W. Li, and M. G. Lee
2016. An essential role for UTX in resolution and activation of bivalent promoters. *Nucleic acids research*, 44(8):3659–3674.

Dou, Y., T. A. Milne, A. J. Ruthenburg, S. Lee, J. W. Lee, G. L. Verdine, C. D. Allis, and R. G. Roeder
2006. Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nature Structural & Molecular Biology*, 13(8):713–719.

Duerre, J. A. and C. T. Lee
1974. In vivo methylation and turnover of rat brain histones. *Journal of neurochemistry*, 23(3):541–547.

Ernst, P., M. Mabon, A. J. Davidson, L. I. Zon, and S. J. Korsmeyer
2004. An Mll-dependent Hox program drives hematopoietic progenitor expansion. *Current Biology*, 14(22):2063–2069.

Etchegaray, J.-P. and R. Mostoslavsky
2016. Interplay between Metabolism and Epigenetics: A Nuclear Adaptation to Environmental Changes. *Molecular Cell*, 62(5):695–711.

Fadloun, A., A. Eid, and M.-E. Torres-Padilla
2013. Mechanisms and dynamics of heterochromatin formation during mammalian development: closed paths and open questions. *Current topics in developmental biology*, 104:1–45.

Fan, J. Y., D. Rangasamy, K. Luger, and D. J. Tremethick
2004. H2A.Z alters the nucleosome surface to promote HP1alpha-mediated chromatin fiber folding. *Molecular Cell*, 16(4):655–661.

Fang, D., H. Gan, J.-H. Lee, J. Han, Z. Wang, S. M. Riester, L. Jin, J. Chen, H. Zhou, J. Wang, H. Zhang, N. Yang, E. W. Bradley, T. H. Ho, B. P. Rubin, J. A. Bridge, S. N. Thibodeau, T. Ordog, Y. Chen, A. J. van Wijnen, A. M. Oliveira, R.-M. Xu, J. J. Westendorf, and Z. Zhang
2016a. The histone H3.3K36M mutation reprograms the epigenome of chondroblastomas. *Science*, 352(6291):1344–1348.

Fang, L., J. Zhang, H. Zhang, X. Yang, X. Jin, L. Zhang, D. G. Skalnik, Y. Jin, Y. Zhang, X. Huang, J. Li, and J. Wong
2016b. H3K4 Methyltransferase Set1a Is A Key Oct4 Coactivator Essential for Generation of Oct4 Positive Inner Cell Mass. *Stem cells (Dayton, Ohio)*, 34(3):565–580.

Fang, R., F. Chen, Z. Dong, D. Hu, A. J. Barbera, E. A. Clark, J. Fang, Y. Yang, P. Mei, M. Rutenberg, Z. Li, Y. Zhang, Y. Xu, H. Yang, P. Wang, M. D. Simon, Q. Zhou, J. Li, M. P. Marynick, X. Li, H. Lu, U. B. Kaiser, R. E. Kingston, Y. Xu, and Y. G. Shi
2013. LSD2/KDM1B and Its Cofactor NPAC/GLYR1 Endow a Structural and Molecular Model for Regulation of H3K4 Demethylation. *Molecular Cell*, 49(3):558–570.

Flanagan, J. F., L.-Z. Mi, M. Chruszcz, M. Cymborowski, K. L. Clines, Y. Kim, W. Minor, F. Rastinejad, and S. Khorasanizadeh
2005. Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature*, 438(7071):1181–1185.

Foe, V. E. and B. M. Alberts
1983. Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in Drosophila embryogenesis. *Journal of Cell Science*, 61:31–70.

Fromm, M. and Z. Avramova
2014. ATX1/AtCOMPASS and the H3K4me3 marks: how do they activate Arabidopsis genes? *Current opinion in plant biology*, 21:75–82.

Gao, S., Y. G. Chung, M. H. Parseghian, G. J. King, E. Y. Adashi, and K. E. Latham
2004. Rapid H1 linker histone transitions following fertilization or somatic cell nuclear transfer: evidence for a uniform developmental program in mice. *Developmental biology*, 266(1):62–75.

Geiss, G. K., R. E. Bumgarner, B. Birditt, T. Dahl, N. Dowidar, D. L. Dunaway, H. P. Fell, S. Ferree, R. D. George, T. Grogan, J. J. James, M. Maysuria, J. D. Mitton, P. Oliveri, J. L. Osborn, T. Peng, A. L. Ratcliffe, P. J. Webster, E. H. Davidson, and L. Hood
2008. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature Biotechnology*, 26(3):317–325.

Glaser, S., S. Lubitz, K. L. Loveland, K. Ohbo, L. Robb, F. Schwenk, J. Seibler, D. Roellig, A. Kranz, K. Anastassiadis, and A. F. Stewart
2009. The histone 3 lysine 4 methyltransferase, Mll2, is only required briefly in development and spermatogenesis. *Epigenetics & chromatin*, 2(1).

Glaser, S., J. Schaft, S. Lubitz, K. Vintersten, F. van der Hoeven, K. R. Tufteland, R. Aasland, K. Anastassiadis, S.-L. Ang, and A. F. Stewart
2006. Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development (Cambridge, England)*, 133(8):1423–1432.

Goldman, J. A., J. D. Garlick, and R. E. Kingston
2010. Chromatin remodeling by imitation switch (ISWI) class ATP-dependent remodelers is stimulated by histone variant H2A.Z. *Journal of Biological Chemistry*, 285(7):4645–4651.

Grebien, F., M. Vedadi, M. Getlik, R. Giambruno, A. Grover, R. Avellino, A. Skucha, S. Vittori, E. Kuznetsova, D. Smil, D. Barsyte-Lovejoy, F. Li, G. Poda, M. Schapira, H. Wu, A. Dong, G. Senisterra, A. Stukalov, K. V. M. Huber, A. Schönegger, R. Marcellus, M. Bilban, C. Bock, P. J. Brown, J. Zuber, K. L. Bennett, R. Al-awar, R. Delwel, C. Nerlov, C. H. Arrowsmith, and G. Superti-

Furga
2015. Pharmacological targeting of the Wdr5-MLL interaction in C/EBPα N-terminal leukemia. *Nature Chemical Biology*, 11(8):571–578.

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

Grove, G. W. and A. Zweidler
1984. Regulation of nucleosomal core histone variant levels in differentiating murine erythroleukemia cells. *Biochemistry*, 23(19):4436–4443.

Grunstein, M.
1997. Histone acetylation in chromatin structure and transcription. *Nature*, 389(6649):349–352.

Grunwald, D. J. and J. S. Eisen
2002. Headwaters of the zebrafish – emergence of a new model vertebrate. *Nature Reviews Genetics*, 3(9):717–724.

Guenther, M. G., R. G. Jenner, B. Chevalier, T. Nakamura, C. M. Croce, E. Canaani, and R. A. Young
2005. Global and Hox-specific roles for the MLL1 methyltransferase. *Proceedings of the National Academy of Sciences of the United States of America*, 102(24):8603–8608.

Hahn, S., E. T. Hoar, and L. Guarente
1985. Each of three "TATA elements" specifies a subset of the transcription initiation sites at the CYC-1 promoter of Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*, 82(24):8562–8566.

Hallson, G., R. E. Hollebakken, T. Li, M. Syrzycka, I. Kim, S. Cotsworth, K. A. Fitzpatrick, D. A. R. Sinclair, and B. M. Honda
2012. dSet1 is the main H3K4 di- and tri-methyltransferase throughout Drosophila development. *Genetics*, 190(1):91–100.

Harvey, S. A., I. Sealy, R. Kettleborough, F. Fenyes, R. White, D. Stemple, and J. C. Smith
2013. Identification of the zebrafish maternal and paternal transcriptomes. *Development (Cambridge, England)*.

Hathaway, N. A., O. Bell, C. Hodges, E. L. Miller, D. S. Neel, and G. R. Crabtree
2012. Dynamics and Memory of Heterochromatin in Living Cells. *Cell*, Pp. 1–14.

Heintzman, N. D., G. C. Hon, R. D. Hawkins, P. Kheradpour, A. Stark, L. F. Harp, Z. Ye, L. K. Lee, R. K. Stuart, C. W. Ching, K. A. Ching, J. E. Antosiewicz-Bourget, H. Liu, X. Zhang, R. D. Green, V. V. Lobanenkov, R. Stewart, J. A. Thomson, G. E. Crawford, M. Kellis, and B. Ren
2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, 459(7243):108–112.

Heintzman, N. D., R. K. Stuart, G. Hon, Y. Fu, C. W. Ching, R. D. Hawkins, L. O. Barrera, S. Van Calcar, C. Qu, K. A. Ching, W. Wang, Z. Weng, R. D. Green, G. E. Crawford, and B. Ren
2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics*, 39(3):311–318.

Herz, H. M., M. Mohan, A. S. Garruss, K. Liang, Y. h. Takahashi, K. Mickey, O. Voets, C. P. Verrijzer, and A. Shilatifard
2012. Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. *Genes & Development*, 26(23):2604–2620.

Herz, H.-M., M. Morgan, X. Gao, J. Jackson, R. Rickels, S. K. Swanson, L. Florens, M. P. Washburn, J. C. Eissenberg, and A. Shilatifard
2014. Histone H3 lysine-to-methionine mutants as a paradigm to study chromatin signaling. *Science*, 345(6200):1065–1070.

Hess, J. L.
2004. Mechanisms of transformation by MLL. *Critical reviews in eukaryotic gene expression*, 14(4):235–254.

Heyn, P., M. Kircher, A. Dahl, J. Kelso, P. Tomancak, A. T. Kalinka, and K. M. Neugebauer
2014. The Earliest Transcribed Zygotic Genes Are Short, Newly Evolved, and Different across Species. *Cell Reports*, Pp. 1–8.

Hödl, M. and K. Basler
2012. Transcription in the absence of histone H3.2 and H3K4 methylation. *Current biology : CB*, 22(23):2253–2257.

Hoshijima, K., M. J. Jurynec, and D. J. Grunwald
2016. Precise Editing of the Zebrafish Genome Made Simple and Efficient. *Developmental Cell*, 36(6):654–667.

Hu, D., X. Gao, M. A. Morgan, H. M. Herz, E. R. Smith, and A. Shilatifard
2013a. The MLL3/MLL4 Branches of the COMPASS Family Function as Major Histone H3K4 Monomethylases at Enhancers. *Molecular and Cellular Biology*, 33(23):4745–4754.

Hu, D., A. S. Garruss, X. Gao, M. A. Morgan, M. Cook, E. R. Smith, and A. Shilatifard
2013b. The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. *Nature Structural & Molecular Biology*, 20(9):1093–1097.

Hu, G., K. Cui, D. Northrup, C. Liu, C. Wang, Q. Tang, K. Ge, D. Levens, C. Crane-Robinson, and K. Zhao
2013c. H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. *Cell stem cell*, 12(2):180–192.

Huang, Y.-C., H.-Y. Shih, S.-J. Lin, C.-C. Chiu, T.-L. Ma, T.-H. Yeh, and Y.-C. Cheng
2014. The epigenetic factor Kmt2a/Mll1 regulates neural progenitor proliferation and neuronal and glial differentiation. *Developmental neurobiology*, 75(5):452–462.

Hung, T., O. Binda, K. S. Champagne, A. J. Kuo, K. Johnson, H. Y. Chang, M. D. Simon, T. G. Kutateladze, and O. Gozani
2009. ING4 mediates crosstalk between histone H3 K4 trimethylation and H3 acetylation to attenuate cellular transformation. *Molecular Cell*, 33(2):248–256.

Huntsman, D. G., S. F. Chin, M. Muleris, S. J. Batley, V. P. Collins, L. M. Wiedemann, S. Aparicio, and C. Caldas
1999. MLL2, the second human homolog of the Drosophila trithorax gene, maps to 19q13.1 and is amplified in solid tumor cell lines. *Oncogene*, 18(56):7975–7984.

Hwang, W. Y., Y. Fu, D. Reyon, M. L. Maeder, S. Q. Tsai, J. D. Sander, R. T. Peterson, J.-R. J. Yeh, and J. K. Joung
2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology*, 31(3):227–229.

Illingworth, R. S., C. H. Botting, G. R. Grimes, W. A. Bickmore, and R. Eskeland
2012. PRC1 and PRC2 Are Not Required for Targeting of H2A.Z to Developmental Genes in Embryonic Stem Cells. *PLoS ONE*, 7(4):e34848.

Ingham, P. W.
1998. trithorax and the regulation of homeotic gene expression in Drosophila: a historical perspective. *The International Journal of Developmental Biology*, 42(3):423–429.

Iwase, S., F. Lan, P. Bayliss, L. de la Torre-Ubieta, M. Huarte, H. H. Qi, J. R. Whetstine, A. Bonni, T. M. Roberts, and Y. Shi
2007. The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. *Cell*, 128(6):1077–1088.

Jenuwein, T., G. Laible, R. Dorn, and G. Reuter
1998. SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cellular and Molecular Life Sciences*, 54(1):80–93.

Jeong, K. W., K. Kim, A. J. Situ, T. S. Ulmer, W. An, and M. R. Stallcup
2011. Recognition of enhancer element-specific histone methylation by TIP60 in transcriptional activation. *Nature Structural & Molecular Biology*, 18(12):1358–1365.

Jiang, H., A. Shukla, X. Wang, W.-y. Chen, B. E. Bernstein, and R. G. Roeder
2011. Role for Dpy-30 in ES Cell-Fate Specification by Regulation of H3K4 Methylation within Bivalent Domains. *Cell*, 144(4):513–525.

Jin, C. and G. Felsenfeld
2007. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes & Development*, 21(12):1519–1529.

Jin, C., C. Zang, G. Wei, K. Cui, W. Peng, K. Zhao, and G. Felsenfeld
2009. H3.3/H2A.Z double variant–containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nature Publishing Group*, 41(8):941–945.

Jones, P. L., G. J. Veenstra, P. A. Wade, D. Vermaak, S. U. Kass, N. Landsberger, J. Strouboulis, and A. P. Wolffe
1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature genetics*, 19(2):187–191.

Joshi, A. A. and K. Struhl
2005. Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Molecular Cell*, 20(6):971–978.

Jude, C. D., L. Climer, D. Xu, E. Artinger, J. K. Fisher, and P. Ernst
2007. Unique and independent roles for MLL in adult hematopoietic stem cells and progenitors. *Cell stem cell*, 1(3):324–337.

Jung, Y. L., L. J. Luquette, J. W. K. Ho, F. Ferrari, M. Tolstorukov, A. Minoda, R. Issner, C. B. Epstein, G. H. Karpen, M. I. Kuroda, and P. J. Park
2014. Impact of sequencing depth in ChIP-seq experiments. *Nucleic acids research*, 42(9):e74–e74.

K, N., C. D. Allis, and S. I. Grewal
2001. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science*, 293(5532):1150–1155.

Kalhan, S. C. and S. E. Marczewski
2012. Methionine, homocysteine, one carbon metabolism and fetal growth. *Reviews in endocrine & metabolic disorders*, 13(2):109–119.

Kamakaka, R. T. and S. Biggins
2005. Histone variants: deviants? *Genes & Development*, 19(3):295–310.

Kane, D. A., M. Hammerschmidt, M. C. Mullins, H. M. Maischein, M. Brand, F. J. van Eeden, M. Furutani-Seiki, M. Granato, P. Haffter, C. P. Heisenberg, Y. J. Jiang, R. N. Kelsh, J. Odenthal, R. M. Warga, and C. Nüsslein-Volhard
1996. The zebrafish epiboly mutants. *Development (Cambridge, England)*, 123:47–55.

Kane, D. A. and C. B. Kimmel
1993. The zebrafish midblastula transition. *Development (Cambridge, England)*, 119(2):447–456.

Kettleborough, R. N. W., E. M. Busch-Nentwich, S. A. Harvey, C. M. Dooley, E. de Bruijn, F. van Eeden, I. Sealy, R. J. White, C. Herd, I. J. Nijman, F. Fenyes, S. Mehroke, C. Scahill, R. Gibbons, N. Wali, S. Carruthers, A. Hall, J. Yen, E. Cuppen, and D. L. Stemple
2013. A systematic genome-wide analysis of zebrafish protein-coding gene function. *Nature*, 496(7446):494–497.

Kim, J.-H., A. Sharma, S. S. Dhar, S.-H. Lee, B. Gu, C.-H. Chan, H.-K. Lin, and M. G. Lee
2014. UTX and MLL4 coordinately regulate transcriptional programs for cell proliferation and invasiveness in breast cancer cells. *Cancer research*, 74(6):1705–1717.

Kim, T. and S. Buratowski
2009. Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. *Cell*, 137(2):259–272.

Kimmel, C. B., W. W. Ballard, S. R. Kimmel, B. Ullmann, and T. F. Schilling
1995. Stages of embryonic development of the zebrafish. *Developmental dynamics : an official publication of the American Association of Anatomists*, 203(3):253–310.

Kobor, M. S., S. Venkatasubrahmanyam, M. D. Meneghini, J. W. Gin, J. L. Jennings, A. J. Link, H. D. Madhani, and J. Rine
2004. A protein complex containing the conserved Swi2/Snf2-related ATPase Swr1p deposits histone variant H2A.Z into euchromatin. *PLoS Biology*, 2(5):E131.

Kok, F. O., M. Shin, C.-W. Ni, A. Gupta, A. S. Grosse, A. van Impel, B. C. Kirchmaier, J. Peterson-Maduro, G. Kourkoulis, I. Male, D. F. DeSantis, S. Sheppard-Tindell, L. Ebarasi, C. Betsholtz, S. Schulte-Merker, S. A. Wolfe, and N. D. Lawson
2014. Reverse Genetic Screening Reveals Poor Correlation between Morpholino-Induced and Mutant Phenotypes in Zebrafish. *Developmental Cell*, Pp. 1–12.

Kornberg, R. D.
1974. Chromatin structure: a repeating unit of histones and DNA. *Science*, 184(4139):868–871.

Krogan, N. J., J. Dover, A. Wood, J. Schneider, J. Heidt, M. A. Boateng, K. Dean, O. W. Ryan, A. Golshani, M. Johnston, J. F. Greenblatt, and A. Shilatifard
2003. The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Molecular Cell*, 11(3):721–729.

Kutach, A. K. and J. T. Kadonaga
2000. The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. *Molecular and Cellular Biology*, 20(13):4754–4764.

Kutyavin, I. V., I. A. Afonina, A. Mills, V. V. Gorn, E. A. Lukhtanov, E. S. Belousov, M. J. Singer, D. K. Walburger, S. G. Lokhov, A. A. Gall, R. Dempcy, M. W. Reed, R. B. Meyer, and J. Hedgpeth
2000. 3'-minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. *Nucleic acids research*, 28(2):655–661.

Lagrange, T., A. N. Kapanidis, H. Tang, D. Reinberg, and R. H. Ebright
1998. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes & Development*, 12(1):34–44.

LaJeunesse, D. and A. Shearn
1995. Trans-regulation of thoracic homeotic selector genes of the Antennapedia and bithorax complexes by the trithorax group genes: absent, small, and homeotic discs 1 and 2. *Mechanisms of development*, 53(1):123–139.

Lan, F., R. E. Collins, R. De Cegli, R. Alpatov, J. R. Horton, X. Shi, O. Gozani, X. Cheng, and Y. Shi
2007. Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature*, 448(7154):718–722.

Lander, E. S., L. M. Linton, B. Birren, and International Human Genome Sequencing Consortium
2001. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921.

Laubichler, M. D. and E. H. Davidson
2008. Boveri's long experiment: sea urchin merogones and the establishment of the role of nuclear chromosomes in development. School of Life Sciences, Center for Biology and Society, Arizona State University, Tempe, AZ 85287-4501, USA.

Lee, J.-E., C. Wang, S. Xu, Y.-W. Cho, L. Wang, X. Feng, A. Baldridge, V. Sartorelli, L. Zhuang, W. Peng, and K. Ge
2013a. H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *eLife*, 2(0):e01503.

Lee, J.-H. and D. G. Skalnik
2005. CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *The Journal of biological chemistry*, 280(50):41725–41731.

Lee, J. H. and D. G. Skalnik
2008. Wdr82 Is a C-Terminal Domain-Binding Protein That Recruits the Setd1A Histone H3-Lys4 Methyltransferase Complex to Transcription Start Sites of Transcribed Human Genes. *Molecular and Cellular Biology*, 28(2):609–618.

Lee, J.-H., C. M. Tate, J.-S. You, and D. G. Skalnik
2007. Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex. *The Journal of biological chemistry*, 282(18):13419–13428.

Lee, M. T., A. R. Bonneau, C. M. Takacs, A. A. Bazzini, K. R. DiVito, E. S. Fleming, and A. J. Giraldez
2013b. Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature*, Pp. 1–24.

Lee, S., J. Lee, S.-K. Lee, and J. W. Lee
2008. Activating signal cointegrator-2 is an essential adaptor to recruit histone H3 lysine 4 methyltransferases MLL3 and MLL4 to the liver X receptors. *Molecular endocrinology (Baltimore, Md.)*, 22(6):1312–1319.

Lehnertz, B., Y. Ueda, A. A. H. A. Derijck, U. Braunschweig, L. Perez-Burgos, S. Kubicek, T. Chen, E. Li, T. Jenuwein, and A. H. F. M. Peters
2003. Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Current Biology*, 13(14):1192–1200.

Leichsenring, M., J. Maes, R. Mossner, W. Driever, and D. Onichtchouk
2013. Pou5f1 Transcription Factor Controls Zygotic Gene Activation In Vertebrates. *Science*, 341(6149):1005–1009.

Lewis, E. B.
1978. A gene complex controlling segmentation in Drosophila. *Nature*, 276(5688):565–570.

Lewis, P. W., M. M. Muller, M. S. Koletsky, F. Cordero, S. Lin, L. A. Banaszynski, B. A. Garcia, T. W. Muir, O. J. Becher, and C. D. Allis
2013. Inhibition of PRC2 Activity by a Gain-of-Function H3 Mutation Found in Pediatric Glioblastoma. *Science*, 340(6134):857–861.

Li, W.-D., Q.-R. Li, S.-N. Xu, F.-J. Wei, Z.-J. Ye, J.-K. Cheng, and J.-P. Chen
2013. Exome sequencing identifies an MLL3 gene germ line mutation in a pedigree of colorectal cancer and acute myeloid leukemia. *Blood*, 121(8):1478–1479.

Li, X. and M. Noll
1994. Compatibility between enhancers and promoters determines the transcriptional specificity of gooseberry and gooseberry neuro in the Drosophila embryo. *The EMBO Journal*, 13(2):400–406.

Liang, H.-L., C.-Y. Nien, H.-Y. Liu, M. M. Metzstein, N. Kirov, and C. Rushlow
2008. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. *Nature*, 456(7220):400–403.

Lifton, R. P., M. L. Goldberg, R. W. Karp, and D. S. Hogness
1978. The organization of the histone genes in Drosophila melanogaster: functional and evolutionary implications. *Cold Spring Harbor symposia on quantitative biology*, 42 Pt 2:1047–1051.

Lindeman, L. C., I. S. Andersen, A. H. Reiner, N. Li, H. Aanes, O. Østrup, C. Winata, S. Mathavan, F. Müller, P. Aleström, and P. Collas
2011. Prepatterning of developmental gene expression by modified histones before zygotic genome activation. *Developmental Cell*, 21(6):993–1004.

Link, V., A. Shevchenko, and C.-P. Heisenberg
2006. Proteomics of early zebrafish embryos. *BMC developmental biology*, 6(1):1.

Liu, X. and J. Secombe
2015. The Histone Demethylase KDM5 Activates Gene Expression by Recognizing Chromatin Context through Its PHD Reader Motif. *Cell Reports*, 13(10):2219–2231.

Lubitz, S., S. Glaser, J. Schaft, A. F. Stewart, and K. Anastassiadis
2007. Increased apoptosis and skewed differentiation in mouse embryonic stem cells lacking the histone methyltransferase Mll2. *Molecular biology of the cell*, 18(6):2356–2366.

Luger, K., A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond
1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, 389(6648):251–260.

Malik, H. S. and S. Henikoff
2003. Phylogenomics of the nucleosome. *Nature structural biology*, 10(11):882–891.

Mansour, A. A., O. Gafni, L. Weinberger, A. Zviran, M. Ayyash, Y. Rais, V. Krupalnik, M. Zerbib, D. Amann-Zalcenstein, I. Maza, S. Geula, S. Viukov, L. Holtzman, A. Pribluda, E. Canaani, S. Horn-Saban, I. Amit, N. Novershtern, and J. H. Hanna
2012. The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. *Nature*, Pp. 1–7.

Mao, Z., L. Pan, W. Wang, J. Sun, S. Shan, Q. Dong, X. Liang, L. Dai, X. Ding, S. Chen, Z. Zhang, B. Zhu, and Z. Zhou
2014. Anp32e, a higher eukaryotic histone chaperone directs preferential recognition for H2A.Z. *Cell research*, Pp. 1–11.

Margueron, R. and D. Reinberg
2011. The Polycomb complex PRC2 and its mark in life. *Nature*, 469(7330):343–349.

Matkar, S., A. Thiel, and X. Hua
2013. Menin: a scaffold protein that controls gene expression and cell signaling. *Trends in Biochemical Sciences*, 38(8):394–402.

McGaughey, D. M., H. Ozel Abaan, R. M. Miller, P. A. Kropp, and L. C. Brody
2014. Genomics of CpG Methylation in Developing and Developed Zebrafish. *G3: Genes|Genomes|Genetics*, P. g3.113.009514.

McKittrick, E., P. R. Gafken, K. Ahmad, and S. Henikoff
2004. Histone H3.3 is enriched in covalent modifications associated with active chromatin. *Proceedings of the National Academy of Sciences of the United States of America*, 101(6):1525–1530.

Merli, C., D. E. Bergstrom, J. A. Cygan, and R. K. Blackman
1996. Promoter specificity mediates the independent regulation of neighboring genes. *Genes & Development*, 10(10):1260–1270.

Mersman, D. P., H.-N. Du, I. M. Fingerman, P. F. South, and S. D. Briggs
2012. Charge-based interaction conserved within histone H3 lysine 4 (H3K4) methyltransferase complexes is needed for protein stability, histone methylation, and gene expression. *Journal of Biological Chemistry*, 287(4):2652–2665.

Mikkelsen, T. S., M. Ku, D. B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T.-K. Kim, R. P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie,

A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E. S. Lander, and B. E. Bernstein
2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 448(7153):553–560.

Miller, T., N. J. Krogan, J. Dover, H. Erdjument-Bromage, P. Tempst, M. Johnston, J. F. Greenblatt, and A. Shilatifard
2001. COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proceedings of the National Academy of Sciences of the United States of America*, 98(23):12902–12907.

Milne, T. A., S. D. Briggs, H. W. Brock, M. E. Martin, D. Gibbs, C. D. Allis, and J. L. Hess
2002. MLL targets SET domain methyltransferase activity to Hox gene promoters. *Molecular Cell*, 10(5):1107–1117.

Mishra, B. P., K. M. Zaffuto, E. L. Artinger, T. Org, H. K. A. Mikkola, C. Cheng, M. Djabali, and P. Ernst
2014. The histone methyltransferase activity of MLL1 is dispensable for hematopoiesis and leukemogenesis. *Cell Reports*, 7(4):1239–1247.

Mizuguchi, G., X. Shen, J. Landry, W.-H. Wu, S. Sen, and C. Wu
2004. ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science*, 303(5656):343–348.

Mohan, M., H.-M. Herz, E. R. Smith, Y. Zhang, J. Jackson, M. P. Washburn, L. Florens, J. C. Eissenberg, and A. Shilatifard
2011. The COMPASS family of H3K4 methylases in Drosophila. *Molecular and Cellular Biology*, 31(21):4310–4318.

Morgan, M. A. and A. Shilatifard
2013. Drosophila SETs Its Sights on Cancer: Trr/MLL3/4 COMPASS-Like Complexes in Development and Disease. *Molecular and Cellular Biology*, 33(9):1698–1701.

Morris, N. R.
1976. A comparison of the structure of chicken erythrocyte and chicken liver chromatin. *Cell*, 9(4 Pt 1):627–632.

Moussaieff, A., N. M. Kogan, and D. Aberdam
2015. Concise Review: Energy Metabolites: Key Mediators of the Epigenetic State of Pluripotency. *Stem cells (Dayton, Ohio)*, 33(8):2374–2380.

Myer, V. E. and R. A. Young
1998. RNA polymerase II holoenzymes and subcomplexes. *The Journal of biological chemistry*, 273(43):27757–27760.

Nady, N., A. Lemak, J. R. Walker, G. V. Avvakumov, M. S. Kareta, M. Achour, S. Xue, S. Duan, A. Allali Hassani, X. Zuo, Y.-X. Wang, C. Bronner, F. Chédin, C. H. Arrowsmith, and S. Dhe-Paganon
2011. Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein. *Journal of Biological Chemistry*, 286(27):24300–24311.

Nagawa, F. and G. R. Fink
1985. The relationship between the "TATA" sequence and transcription initiation sites at the HIS4 gene of Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*, 82(24):8557–8561.

Narlikar, G. J., R. Sundaramoorthy, and T. Owen-Hughes
2013. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell*, 154(3):490–503.

Newport, J. and M. Kirschner
1982. A major developmental transition in early Xenopus embryos: II. Control of the onset of transcription. *Cell*, 30(3):687–696.

Ng, H.-H., F. Robert, R. A. Young, and K. Struhl
2003. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Molecular Cell*, 11(3):709–719.

Ng, R. K. and J. B. Gurdon
2008. Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription. *Nature cell biology*, 10(1):102–109.

Nishino, Y., M. Eltsov, Y. Joti, K. Ito, H. Takata, Y. Takahashi, S. Hihara, A. S. Frangakis, N. Imamoto, T. Ishikawa, and K. Maeshima
2012. Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *The EMBO Journal*, 31(7):1644–1653.

Obri, A., K. Ouararhni, C. Papin, M.-L. Diebold, K. Padmanabhan, M. Marek, I. Stoll, L. Roy, P. T. Reilly, T. W. Mak, S. Dimitrov, C. Romier, and A. Hamiche
2014. ANP32E is a histone chaperone that removes H2A.Z from chromatin. *Nature*, Pp. 1–18.

Okitsu, C. Y. and C.-L. Hsieh
2007. DNA methylation dictates histone H3K4 methylation. *Molecular and Cellular Biology*, 27(7):2746–2757.

O'Meara, E., D. Stack, S. Phelan, N. McDonagh, L. Kelly, R. Sciot, M. Debiec-Rychter, T. Morris, D. Cochrane, P. Sorensen, and M. J. O'Sullivan
2014. Identification of an MLL4-GPS2 fusion as an oncogenic driver of undifferentiated spindle cell sarcoma in a child. *Genes, chromosomes & cancer*, 53(12):991–998.

Orlando, D. A., M. W. Chen, V. E. Brown, S. Solanki, Y. J. Choi, E. R. Olson, C. C. Fritz, J. E. Bradner, and M. G. Guenther
2014. Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Reports*, 9(3):1163–1170.

Oron, E. and N. Ivanova
2012. Cell fate regulation in early mammalian development. *Physical biology*, 9(4):045002.

Ortega-Molina, A., I. W. Boss, A. Canela, H. Pan, Y. Jiang, C. Zhao, M. Jiang, D. Hu, X. Agirre, I. Niesvizky, J.-E. Lee, H.-T. Chen, D. Ennishi, D. W. Scott, A. Mottok, C. Hother, S. Liu, X.-J. Cao, W. Tam, R. Shaknovich, B. A. Garcia, R. D. Gascoyne, K. Ge, A. Shilatifard, O. Elemento, A. Nussenzweig, A. M. Melnick, and H.-G. Wendel
2015. The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. *Nature medicine*, 21(10):1199–1208.

Park, Y.-J., P. N. Dyer, D. J. Tremethick, and K. Luger
2004. A new fluorescence resonance energy transfer approach demonstrates that the histone variant H2AZ stabilizes the histone octamer within the nucleosome. *The Journal of biological chemistry*, 279(23):24274–24282.

Passarge, E.
1979. *Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago.*, volume 31.

Pataskar, A., J. Jung, P. Smialowski, F. Noack, F. Calegari, T. Straub, and V. K. Tiwari
2016. NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program. *The EMBO Journal*, 35(1):24–45.

Pauli, A., E. Valen, M. F. Lin, M. Garber, N. L. Vastenhouw, J. Z. Levin, L. Fan, A. Sandelin, J. L. Rinn, A. Regev, and A. F. Schier
2012. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Research*, 22(3):577–591.

Perrimon, N., L. Engstrom, and A. P. Mahowald
1984. The effects of zygotic lethal mutations on female germ-line functions in Drosophila. *Developmental biology*, 105(2):404–414.

Pinskaya, M. and A. Morillon
2009. Histone H3 lysine 4 di-methylation: a novel mark for transcriptional fidelity? *Epigenetics : official journal of the DNA Methylation Society*, 4(5):302–306.

Postlethwait, J. H., I. G. Woods, P. Ngo-Hazelett, Y. L. Yan, P. D. Kelly, F. Chu, H. Huang, A. Hill-Force, and W. S. Talbot
2000. Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Research*, 10(12):1890–1902.

Prendergast, G. C., D. Lawe, and E. B. Ziff
1991. Association of Myn, the murine homolog of max, with c-Myc stimulates methylation-sensitive DNA binding and ras cotransformation. *Cell*, 65(3):395–407.

Prioleau, M. N., J. Huet, A. Sentenac, and M. Méchali
1994. Competition between chromatin and transcription complex assembly regulates gene expression during early development. *Cell*, 77(3):439–449.

Reinberg, D., M. Horikoshi, and R. G. Roeder
1987. Factors involved in specific transcription in mammalian RNA polymerase II. Functional analysis of initiation factors IIA and IID and identification of a new factor operating at sequences downstream of the initiation site. *The Journal of biological chemistry*, 262(7):3322–3330.

Roguev, A., D. Schaft, A. Shevchenko, W. W. Pijnappel, M. Wilm, R. Aasland, and A. F. Stewart
2001. The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *The EMBO Journal*, 20(24):7137–7148.

Rugg-Gunn, P. J., B. J. Cox, A. Ralston, and J. Rossant
2010. Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. *Proceedings of the National Academy of Sciences of the United States of America*, 107(24):10783–10790.

Saeki, H., K. Ohsumi, H. Aihara, T. Ito, S. Hirose, K. Ura, and Y. Kaneda
2005. Linker histone variants control chromatin dynamics during early embryogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 102(16):5697–5702.

Sansom, O. J., K. Maddison, and A. R. Clarke
2007. Mechanisms of disease: methyl-binding domain proteins as potential therapeutic targets in cancer. *Nature clinical practice. Oncology*, 4(5):305–315.

Santenard, A., C. Ziegler-Birling, M. Koch, L. Tora, A. J. Bannister, and M.-E. Torres-Padilla
2010. Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. *Nature cell biology*, 12(9):853–862.

Saxonov, S., P. Berg, and D. L. Brutlag
2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences*, 103(5):1412–1417.

Schertel, C., M. Albarca, C. Rockel-Bauer, N. W. Kelley, J. Bischof, K. Hens, E. van Nimwegen, K. Basler, and B. Deplancke
2015. A large-scale, in vivo transcription factor screen defines bivalent chromatin as a key property of regulatory factors mediating Drosophila wing development. *Genome Research*, 25(4):514–523.

Schier, A. F.
2007. The Maternal-Zygotic Transition: Death and Birth of RNAs. *Science*, 316(5823):406–407.

Schneider, T. D., J. M. Arteaga-Salas, E. Mentele, R. David, D. Nicetto, A. Imhof, and R. A. W. Rupp
2011. Stage-specific histone modification profiles reveal global transitions in the Xenopus embryonic

epigenome. *PLoS ONE*, 6(7):e22548.

Schuettengruber, B., M. Ganapathi, B. Leblanc, M. Portoso, R. Jaschek, B. Tolhuis, M. van Lohuizen, A. Tanay, and G. Cavalli
2009. Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. *PLoS Biology*, 7(1):e13.

Secombe, J. and R. N. Eisenman
2007. The function and regulation of the JARID1 family of histone H3 lysine 4 demethylases: the Myc connection. *Cell cycle (Georgetown, Tex.)*, 6(11):1324–1328.

Senisterra, G., H. Wu, A. Allali Hassani, G. A. Wasney, D. Barsyte-Lovejoy, L. Dombrovski, A. Dong, K. T. Nguyen, D. Smil, Y. Bolshan, T. Hajian, H. He, A. Seitova, I. Chau, F. Li, G. Poda, J. F. Couture, P. J. Brown, R. Al-awar, M. Schapira, C. H. Arrowsmith, and M. Vedadi
2012. Small-molecule inhibition of MLL activity by disruption of its interaction with WDR5. *Biochemical Journal*, 449(1):151–159.

Seward, D. J., G. Cubberley, S. Kim, M. Schonewald, L. Zhang, B. Tripet, and D. L. Bentley
2007. Demethylation of trimethylated histone H3 Lys4 in vivo by JARID1 JmjC proteins. *Nature Structural & Molecular Biology*, 14(3):240–242.

Shah, M. A., E. L. Denton, C. H. Arrowsmith, M. Lupien, and M. Schapira
2014. A global assessment of cancer genomic alterations in epigenetic mechanisms. *Epigenetics & chromatin*, 7(1):29.

Shao, Z., Y. Zhang, G.-C. Yuan, S. H. Orkin, and D. J. Waxman
2012. MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. *Genome Biology*, 13(3):R16.

Shi, Y., F. Lan, C. Matson, P. Mulligan, J. R. Whetstine, P. A. Cole, R. A. Casero, and Y. Shi
2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell*, 119(7):941–953.

Shilatifard, A.
2012. The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. *Annual Review of Biochemistry*, 81(1):65–95.

Simmich, J., E. Staykov, and E. Scott
2012. Zebrafish as an appealing model for optogenetic studies. *Progress in brain research*, 196:145–162.

Southall, S. M., P.-S. Wong, Z. Odho, S. M. Roe, and J. R. Wilson
2009. Structural Basis for the Requirement of Additional Factors for MLL1 SET Domain Activity and Recognition of Epigenetic Marks. *Molecular Cell*, 33(2):181–191.

Sparmann, A. and M. van Lohuizen
2006. Polycomb silencers control cell fate, development and cancer. *Nature Reviews Cancer*, 6(11):846–856.

Steinbach, O. C., A. P. Wolffe, and R. A. Rupp
1997. Somatic linker histones cause loss of mesodermal competence in Xenopus. *Nature*, 389(6649):395–399.

Steinhauser, S., N. Kurzawa, R. Eils, and C. Herrmann
2016. A comprehensive comparison of tools for differential ChIP-seq analysis. *Briefings in bioinformatics*, P. bbv110.

Sterner, D. E. and S. L. Berger
2000. Acetylation of histones and transcription-related factors. *Microbiology and molecular biology reviews : MMBR*, 64(2):435–459.

Stoller, J. Z., L. Huang, C. C. Tan, F. Huang, D. D. Zhou, J. Yang, B. D. Gelb, and J. A. Epstein
2010. Ash2l interacts with Tbx1 and is required during early embryogenesis. *Experimental biology and medicine (Maywood, N.J.)*, 235(5):569–576.

Streisinger, G., C. Walker, N. Dower, D. Knauber, and F. Singer
1981. Production of clones of homozygous diploid zebra fish (Brachydanio rerio). *Nature*, 291(5813):293–296.

Subramanian, V., P. A. Fields, and L. A. Boyer
2015. H2A.Z: a molecular rheostat for transcriptional control. *F1000prime reports*, 7(1):01.

Subramanian, V., A. Mazumder, L. E. Surface, V. L. Butty, P. A. Fields, A. Alwan, L. Torrey, K. K. Thai, S. S. Levine, M. Bathe, and L. A. Boyer
2013. H2A.Z Acidic Patch Couples Chromatin Dynamics to Regulation of Gene Expression Programs during ESC Differentiation. *PLoS Genetics*, 9(8):e1003725.

Sulston, J. E., E. Schierenberg, J. G. White, and J. N. Thomson
1983. The embryonic cell lineage of the nematode Caenorhabditis elegans. *Developmental biology*, 100(1):64–119.

Sun, X.-J., P.-F. Xu, T. Zhou, M. Hu, C.-T. Fu, Y. Zhang, Y. Jin, Y. Chen, S.-J. Chen, Q.-H. Huang, T. X. Liu, and Z. Chen
2008. Genome-Wide Survey and Developmental Expression Mapping of Zebrafish SET Domain-Containing Genes. *PLoS ONE*, 3(1):e1499.

Surface, L. E., P. A. Fields, V. Subramanian, R. Behmer, N. Udeshi, S. E. Peach, S. A. Carr, J. D. Jaffe, and L. A. Boyer
2016. H2A.Z.1 Monoubiquitylation Antagonizes BRD2 to Maintain Poised Chromatin in ESCs. *Cell Reports*, Pp. 1–15.

Suzuki, Y., T. Tsunoda, J. Sese, H. Taira, J. Mizushima-Sugano, H. Hata, T. Ota, T. Isogai, T. Tanaka, Y. Nakamura, A. Suyama, Y. Sakaki, S. Morishita, K. Okubo, and S. Sugano
2001. Identification and characterization of the potential promoter regions of 1031 kinds of human genes. *Genome Research*, 11(5):677–684.

Takai, D. and P. A. Jones
2002. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proceedings of the National Academy of Sciences*, 99(6):3740–3745.

Terranova, R., H. Agherbi, A. Boned, S. Meresse, and M. Djabali
2006. Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. *Proceedings of the National Academy of Sciences of the United States of America*, 103(17):6629–6634.

Thakar, A., P. Gupta, W. T. McAllister, and J. Zlatanova
2010. Histone variant H2A.Z inhibits transcription in reconstituted nucleosomes. *Biochemistry*, 49(19):4018–4026.

Tolunay, H. E., L. Yang, W. F. Anderson, and B. Safer
1984. Isolation of an active transcription initiation complex from HeLa cell-free extract. *Proceedings of the National Academy of Sciences*, 81(19):5916–5920.

Tsai, F. T. and P. B. Sigler
2000. Structural basis of preinitiation complex assembly on human pol II promoters. *The EMBO Journal*, 19(1):25–36.

Vastenhouw, N. L., Y. Zhang, I. G. Woods, F. Imam, A. Regev, X. S. Liu, J. Rinn, and A. F. Schier
2010. Chromatin signature of embryonic pluripotency is established during genome activation. *Nature*, 464(7290):922–926.

Vermeulen, M., K. W. Mulder, S. Denissov, W. W. M. P. Pijnappel, F. M. A. van Schaik, R. A. Varier, M. P. A. Baltissen, H. G. Stunnenberg, M. Mann, and H. T. M. Timmers
2007. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell*, 131(1):58–69.

Volpe, T. A., C. Kidner, I. M. Hall, G. Teng, S. I. S. Grewal, and R. A. Martienssen
2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science*, 297(5588):1833–1837.

Wachter, E., T. Quante, C. Merusi, A. Arczewska, F. Stewart, S. Webb, and A. Bird
2014. Synthetic CpG islands reveal DNA sequence determinants of chromatin structure. *eLife*, 3:e03397.

Waddington, C. H.
2012. *The epigenotype. 1942.*, volume 41. Oxford University Press.

Wan, M., J. Liang, Y. Xiong, F. Shi, Y. Zhang, W. Lu, Q. He, D. Yang, R. Chen, D. Liu, M. Barton, and Z. Songyang
2013. The trithorax group protein Ash2l is essential for pluripotency and maintaining open chromatin in embryonic stem cells. *Journal of Biological Chemistry*, 288(7):5039–5048.

Wan, X., B. Hu, J.-x. Liu, X. Feng, and W. Xiao
2011. Zebrafish mll gene is essential for hematopoiesis. *Journal of Biological Chemistry*, 286(38):33345–33357.

Wang, P., C. Lin, E. R. Smith, H. Guo, B. W. Sanderson, M. Wu, M. Gogol, T. Alexander, C. Seidel, L. M. Wiedemann, K. Ge, R. Krumlauf, and A. Shilatifard
2009. Global Analysis of H3K4 Methylation Defines MLL Family Member Targets and Points to a Role for MLL1-Mediated H3K4 Methylation in the Regulation of Transcriptional Initiation by RNA Polymerase II. *Molecular and Cellular Biology*, 29(22):6074–6085.

Wang, X., K. Zhu, S. Li, Y. Liao, R. Du, X. Zhang, H. B. Shu, A. Y. Guo, L. Li, and M. Wu
2012. MLL1, a H3K4 methyltransferase, regulates the TNF~-stimulated activation of genes downstream of NF-~B. *Journal of Cell Science*, 125(17):4058–4066.

Wang, X.-X., L. Fu, X. Li, X. Wu, Z. Zhu, L. Fu, and J.-T. Dong
2011. Somatic mutations of the mixed-lineage leukemia 3 (MLL3) gene in primary breast cancers. *Pathology oncology research : POR*, 17(2):429–433.

Watanabe, Y., R. J. Castoro, H. S. Kim, B. North, R. Oikawa, T. Hiraishi, S. S. Ahmed, W. Chung, M.-Y. Cho, M. Toyota, F. Itoh, M. R. H. Estecio, L. Shen, J. Jelinek, and J.-P. J. Issa
2011. Frequent alteration of MLL3 frameshift mutations in microsatellite deficient colorectal cancer. *PLoS ONE*, 6(8):e23320.

Weber, C. M., S. Ramachandran, and S. Henikoff
2014. Nucleosomes Are Context-Specific, H2A.Z-Modulated Barriers to RNA Polymerase. *Molecular Cell*, 53(5):819–830.

Wen, B., H. Wu, Y. Shinkai, R. A. Irizarry, and A. P. Feinberg
2009. Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nature Publishing Group*, 41(2):246–250.

Wiekowski, M., M. Miranda, and M. L. DePamphilis
1993. Requirements for promoter activity in mouse oocytes and embryos distinguish paternal pronuclei from maternal and zygotic nuclei. *Developmental biology*, 159(1):366–378.

Woo, Y. H. and W.-H. Li
2012. Evolutionary conservation of histone modifications in mammals. *Molecular biology and evolution*, 29(7):1757–1767.

Wu, M. and J. C. Gerhart
1980. Partial purification and characterization of the maturation-promoting factor from eggs of Xenopus laevis. *Developmental biology*, 79(2):465–477.

Wu, R. S., S. Tsai, and W. M. Bonner
1982. Patterns of histone variant synthesis can distinguish G0 from G1 cells. *Cell*, 31(2 Pt 1):367–374.

Wysocka, J., T. Swigut, T. A. Milne, Y. Dou, X. Zhang, A. L. Burlingame, R. G. Roeder, A. H. Brivanlou, and C. D. Allis
2005. WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell*, 121(6):859–872.

Yagi, H., K. Deguchi, A. Aono, Y. Tani, T. Kishimoto, and T. Komori
1998. Growth disturbance in fetal liver hematopoiesis of Mll-mutant mice. *Blood*, 92(1):108–117.

Yang, C., E. Bolotin, T. Jiang, F. M. Sladek, and E. Martinez
2007. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene*, 389(1):52–65.

Yang, S.-M., B. J. Kim, L. Norwood Toro, and A. I. Skoultchi
2013. H1 linker histone promotes epigenetic silencing by regulating both DNA methylation and histone H3 methylation. *Proceedings of the National Academy of Sciences of the United States of America*, 110(5):1708–1713.

Yokoyama, A. and M. L. Cleary
2008. Menin critically links MLL proteins with LEDGF on cancer-associated target genes. *Cancer cell*, 14(1):36–46.

Yokoyama, A., Z. Wang, J. Wysocka, M. Sanyal, D. J. Aufiero, I. Kitabayashi, W. Herr, and M. L. Cleary
2004. Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Molecular and Cellular Biology*, 24(13):5639–5649.

Yu, B. D., R. D. Hanson, J. L. Hess, S. E. Horning, and S. J. Korsmeyer
1998. MLL, a mammalian trithorax-group gene, functions as a transcriptional maintenance factor in morphogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 95(18):10632–10636.

Yu, B. D., J. L. Hess, S. E. Horning, G. A. Brown, and S. J. Korsmeyer
1995. Altered Hox expression and segmental identity in Mll-mutant mice. 378(6556):505–508.

Zeleznik-Le, N. J., A. M. Harden, and J. D. Rowley
1994. 11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene. *Proceedings of the National Academy of Sciences*, 91(22):10610–10614.

Zhang, H., S. Gayen, J. Xiong, B. Zhou, A. K. Shanmugam, Y. Sun, H. Karatas, L. Liu, R. C. Rao, S. Wang, A. I. Nesvizhskii, S. Kalantry, and Y. Dou
2016. MLL1 Inhibition Reprograms Epiblast Stem Cells to Naive Pluripotency. *STEM*, Pp. 1–15.

Zhang, Y., N. L. Vastenhouw, J. Feng, K. Fu, C. Wang, Y. Ge, A. Pauli, P. van Hummelen, A. F. Schier, and X. S. Liu
2014. Canonical nucleosome organization at promoters forms during genome activation. *Genome Research*, 24(2):260–266.

Ziemin-van der Poel, S., N. R. McCabe, H. J. Gill, R. Espinosa, Y. Patel, A. Harden, P. Rubinelli, S. D. Smith, M. M. LeBeau, and J. D. Rowley
1991. Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias. *Proceedings of the National Academy of Sciences*, 88(23):10735–10739.

Zilberman, D., D. Coleman-Derr, T. Ballinger, and S. Henikoff
2008. Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature*,

456(7218):125–129.

Declaration of Authorship

I hereby declare that I have authored this doctoral thesis independently, without the prohibited assistance of third parties and without making use of aids other than those specified. Passages which are literally or in general matter taken over from other sources have been identified as such. The present document has not been previously presented in identical or similar form to any other german or foreign examination board.

I further declare that I have not undertaken any previous unsuccessful doctorate proceedings.

I declare that I recognize the doctorate regulations of the School of Science of the Technische Universität Dresden.

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig, ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Textpassagen oder Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder in Deutschland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Meine Person betreffend erkläre ich hiermit, dass keine früheren erfolglosen Promotionsverfahren stattgefunden haben.

Hiermit erkenne ich die Promotionsordnung der Fakultät für Mathematik und Naturwissenschaften der Technischen Universität Dresden an.

Place / Date Signature