Role of small RNAs and chromatin in transposable element silencing during global demethylation

Rebecca Vanessa Berrens TRINITY COLLEGE



This dissertation is submitted for the degree of Doctor of Philosophy

October 2016 Graduate School of Life Sciences



Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. This thesis does not exceed the word limit of 60,000 words required by the University of Cambridge School of Biological Sciences.

> Rebecca V. Berrens September 2016

ii

Table of Acknowledgments of Assistance

Data/materials provided by someone else:

- Unbiased analysis of TE classes in total RNA-seq and ChIP-seq data, antisense analysis in *Dnmt1* KO ESCs as well as the demethylation dynamics of TEs were done by Dr. Simon Andrews
- PGCLCs were made by Dr. Ferdinand von Meyenn
- iPS reprogramming was done by Dr. Inês Milagre
- in vivo PGCs were isolated by Dr. Wendy Dean

Data produced jointly:

- Microscopy pictures of IF staining were taken jointly with Dr. Fatima Santos
- CRISPR Cas9 conditional KO was designed with the help of Dr. Dominik Spensberger

Data obtained from technical service provider:

- Total RNA-seq libraries were done by Wellcome Trust Sanger Institute sample preparation pipeline
- Next generation sequencing were done by Wellcome Trust Sanger Institute and Babraham Institute NGS facility
- Flow cytometry for CRISPR clones was done by Flow cytometry facility at Babraham Institute

Abstract

DNA methylation entails the addition of a methyl group to the 5-carbon of the cytosine base of the DNA. This modification is important during many biological processes such as imprinting, X-chromosome inactivation, cell differentiation as well as silencing of transposable elements (TEs). DNA methylation is dynamic during early mammalian development, despite being a more static mark in somatic cells. Global hypomethylation is a hallmark of epigenetic reprogramming in mammalian primordial germ cells (PGCs), the early embryo and in naïve embryonic stem cells (ESCs). Genome integrity is crucial during early development, as the germline DNA needs to be protected for future generations. Therefore, epigenetic reprogramming presents a critical phase for TE defence since presumably alternative silencing pathways need to be employed to limit their activity. In this thesis I investigate the role of small RNAs to control TEs during global waves of DNA demethylation in cellular reprogramming, naïve pluripotency as well as early mammalian development.

Following an introduction into the research questions, in chapter 3 I investigate the mechanism of TE regulation in an *in vitro* model of *Dnmt1* deletion in mouse ES cells to recapitulate *in vivo* epigenetic reprogramming. I find that certain classes of TEs become transcriptionally upregulated and subsequently resilenced by a mechanism independent of DNA methylation. I identify ARGONAUTE 2 (AGO2) bound siRNAs as the prominent mechanism to control certain classes of TEs, while others appear to be regulated by redistribution of repressive histone modifications.

In chapter 4, I construct *Dicer* constitutive and conditional KO ESCs in the background of the $Dnmt1^{fl/fl}$ ESCs using CRISPR-Cas9. I dissect the role of DNA methylation and

of DICER dependent small RNAs on transcriptional changes of ESCs. Additionally, I find that DICER dependent small interfering RNAs (siRNAs) re-silence transcriptionally active TE classes.

Finally, in chapter 5, I examine the role of small RNAs in TE silencing in different models of global hypomethylation *in vivo* and *in vitro* PGCs, during iPSC reprogramming and in a transition from serum to 2i culturing of mouse ESCs.

Acknowledgements

A PhD is defined as independent research. Thus, I had to figure problems out and needed to make them work. However, I could have not managed my PhD, the last four years, without many people I feel I should thank.

First and foremost, I am deeply grateful to my supervisor Wolf Reik for guidance, support, patience, and trust. I was really lucky to be able to work on my own project driven by the excitement in small RNAs controlling transposable elements during epigenetic reprogramming.

I extend my gratitude to my collaborators. Especially, Dominik Spensberger, for his invaluable help with the CRISPR Cas9 knockouts; I would have given up on this without his optimism and experimental insights. Furthermore, I want to thank Simon Andrews for helping me to analyse transposable elements in various datasets.

I thank my fellow labmates in the Reik lab, for their stimulating discussions during hours of lab meetings. I would like to thank Tamir Chandra for his encouragement and way of thinking out of the box which widened my research from various perspectives. I thank Wendy Dean for her advice and discussions throughout my time at the Babraham Institute. I have to mention Ferdinand von Meyenn for all the discussions about small RNAs and his advice on various experimental questions. I would like to thank Inês Milagre and Solenn Patalano to remind me that you have to follow your heart when doing science. Furthermore, my time in the lab would have been not the same without the support of the current PhD students in the Reik lab, thank you for all the fun we have had in the last years.

This thesis would not read well without the editing assistance of Wendy Dean, Heather Lee, Ferdinand von Meyenn, Julia Spindel, Poppy Gould, Ola Kolodziejczyk, Annette LaRocco and Joaquina Delas Vives.

I would like to thank the Gates Cambridge Fellowship for funding to support my research.

I would like to thank my friends for sharing the time with me, for making me laugh and for making this life as joyful as it can be.

My Cambridge friends, since I arrived in Cambridge in January 2013 you have always been there for me. Without you the last four years would have not been the same. The Heidelberg crew, for all the fun we have on our weekends throughout the year. I would like to thank my friends from Kempen, who always make me feel home again when I visit.

A special thanks to my boyfriend Armin Lak, without your support, positive thinking and believe in me, I would have not been able to finish this work.

Ein ganz riesengrosses Danke gilt meinen Eltern und meiner "kleinen" Schwester. Eure Unterstützung und Liebe habe mich durch schwere Zeiten begleitet und mich immer wieder zum Lachen gebracht. Eure Care-Pakete waren immer vollgepackt mit tollen Sachen die das Leben schöner machen. Eure Besuche und Telefonate zeigen mir, dass ihr immer hinter mir steht und unsere gemeinsame Urlaube waren eine unbeschreibliche Zeit. Danke für Alles!

Contents

1	Intr	oducti	ion	1
	1.1	Chron	natin	2
	1.2	DNA	methylation	3
		1.2.1	Propagation of DNA methylation	4
		1.2.2	Loss of DNA methylation	6
		1.2.3	Role of DNA methylation	7
		1.2.4	Chromatin modifications	9
		1.2.5	Post-translational histone modifications	10
		1.2.6	Role of histone modifications	12
	1.3	Histor	ne modifications and DNA methylation	16
	1.4	Non-c	oding RNAs	18
		1.4.1	lncRNAs	19
		1.4.2	small RNAs	19
		1.4.3	miRNAs	20
		1.4.4	Endogenous siRNAs	23
		1.4.5	piRNAs	25
	1.5	Epiger	netic reprogramming	29
		1.5.1	Zygote reprogramming	29
		1.5.2	PGC reprogramming	30
	1.6	Trans	posable elements (TEs)	33
		1.6.1	TE families	34
		1.6.2	Lifecycle of TEs	41
		1.6.3	TEs and the genome: Friends or foe?	42

		1.6.4	Major Satellites	45
	1.7	Epigen	netic modifications of TEs	46
		1.7.1	Methylation of TEs	46
		1.7.2	Histone modifications of TEs	47
		1.7.3	Regulation of TEs by small RNAs	48
		1.7.4	TEs during epigenetic reprogramming	49
	1.8	Embry	ronic stem cells	50
		1.8.1	Transcription factor networks regulate ESC pluripotency	51
		1.8.2	H3K9 and H3K27 methylation in ESCs	52
2	Mat	terial a	nd Methods	53
	2.1	Materi	als	53
	2.2	Oligon	ucleotides	60
	2.3	Metho	ds	62
		2.3.1	Cell culture	62
		2.3.2	Reprogramming of MEF to iPSCs	64
		2.3.3	In vitro PGC like cells	64
		2.3.4	In vivo PGC collection	64
		2.3.5	DNA extraction	64
		2.3.6	RNA extraction	65
		2.3.7	Measurement of RNA and DNA concentration	65
		2.3.8	qRTPCR	65
		2.3.9	CRISPR KOs	65
		2.3.10	Small RNA qRTPCR	69
		2.3.11	AGO2 immunoprecipitation	69
		2.3.12	siRNA knock down	70
		2.3.13	Immunofluorescence	70
		2.3.14	Next generation sequencing library preparation $\ldots \ldots \ldots$	71
		2.3.15	Sequencing	74
		2.3.16	Bioinformatics	74

3 Acute demethylation upon *Dnmt1* conditional KO

79

	3.1	Introdu	uction	79
	3.2	Results	5	82
		3.2.1	Acute deletion of $Dnmt1$ as a model for global demethylation	
			dynamics	82
		3.2.2	Dnmt1 KO leads to transcriptional upregulation of imprinted loci	85
		3.2.3	Small RNAs from Dlk and X-chromosome locus $Xlr3$ become	
			upregulated upon <i>Dnmt1</i> KO	88
		3.2.4	No genome-wide chromatin changes upon global demethylation	
			in ESCs	91
		3.2.5	Mapping of TEs in next generation sequencing libraries	96
		3.2.6	TE classes become demethylated upon $\textit{Dnmt1}$ KO in ESCs $~$	97
		3.2.7	$Dnmt1~{\rm KO}$ leads to transcriptional activation of ERV elements	98
		3.2.8	Sense/antisense transcription of repeat families feed into the	
			RNAi pathway	101
		3.2.9	Small RNAs are produced from TEs upon loss of $Dnmt1$	104
		3.2.10	Small RNAs are actively loaded into the RNAi machinery	106
		3.2.11	EndosiRNAs and not miRNAs play a critical role in resilencing	
			of IAPs	107
		3.2.12	Histone modifications may account for different behaviour of TE	
			families upon acute <i>Dnmt1</i> deletion	110
	3.3	Discus	sion	113
4	TE	regulat	tion upon <i>Dicer</i> KO and <i>Dicer/Dnmt1</i> KO	119
	4.1	Introdu	uction	119
	4.2	Results	S	121
		4.2.1	Constitutive KO of <i>Dicer</i> by CRISPR-Cas9	121
		4.2.2	Transcriptional changes in $Dicer\ {\rm KO}\ {\rm and}\ Dicer/Dnmt1\ {\rm double}$	
			KO ESCs	123
		4.2.3	Genome-wide histone marks upon <i>Dicer</i> KO and	
			Dnmt1/Dicer DKO	127

		4.2.4	LINE-1 and major satellite expression in $Dicer$ constitutive KO	
			ESCs	129
		4.2.5	Histone modifications may explain different behaviours of TEs	
			to acute Dnmt1 KO	132
		4.2.6	TE response after double conditional KO of $Dicer$ and $Dnmt1$.	135
		4.2.7	miRNA production against IAPs suggest multiple levels of gene	
			regulation	138
		4.2.8	IAP protein production upon <i>Dicer</i> KO	140
	4.3	Discus	sion	141
5	Sma	all RN.	As as suppressors of TEs upon demethylation	147
5	Sma 5.1	all RN . Introd	As as suppressors of TEs upon demethylation	147 147
5	Sma 5.1 5.2	all RN . Introd Result	As as suppressors of TEs upon demethylation uction	147 147 150
5	Sma 5.1 5.2	all RN . Introd Result 5.2.1	As as suppressors of TEs upon demethylation uction	 147 147 150 150
5	Sma 5.1 5.2	all RN . Introd Result 5.2.1 5.2.2	As as suppressors of TEs upon demethylation uction	 147 147 150 150 154
5	Sma 5.1 5.2	Introd Result 5.2.1 5.2.2 5.2.3	As as suppressors of TEs upon demethylation uction	 147 147 150 150 154 157
5	Sma 5.1 5.2 5.3	all RN . Introd Result 5.2.1 5.2.2 5.2.3 Discus	As as suppressors of TEs upon demethylation uction	 147 147 150 150 154 157 164

List of Figures

1.1	DNA methylation	4
1.2	Maintenance DNA methylation and H3K9me3	17
1.3	De novo DNA methylation and H3K4 methylation	18
1.4	Small RNA pathway in mice	20
1.5	miRNA biogenesis	22
1.6	endosiRNA biogenesis	25
1.7	piRNA biogenesis	27
1.8	Epigenetic reprogramming in mammalian development	29
3.1	Schematic of <i>in vitro Dnmt1</i> deletion time-course in mouse ESCs	80
3.2	Genome demethylation upon <i>Dnmt1</i> deletion	83
3.3	Genome demethylation in LMRs and enhancer regions	83
3.4	WGBS on chromosome 2 upon <i>Dnmt1</i> KO	84
3.5	Regions resisting global demethylation upon $Dnmt1$ KO \ldots	84
3.6	Genes respondent to <i>Dnmt1</i> driven demethylation	86
3.7	Highest respondent genes upon <i>Dnmt1</i> deletion	86
3.8	Normal expression of pluripotency network upon $Dnmt1$ KO	87
3.9	Transcriptional expression did not recapitulate methylation status	87
3.10	Small RNA-seq libraries are 90% made up of miRNAs	88
3.11	Not significantly altered expression of endogenous miRNAs in ESCs $$.	89
3.12	miRNAs from the Dlk and $Xlr3$ locus are dependent on $Dnmt1$	90
3.13	Upregulation of mmu-miR-367 and mmu-miR-543 upon $Dnmt1\ KO$	90
3.14	Expression of Dlk locus upon $Dnmt1$ KO	91

3.15	Enrichment of repressive histone marks in the genome	92
3.16	Histone marks over gene body and TSS	93
3.17	Histone mark enrichment in the genome	93
3.18	Minor redistribution of global histone marks upon $Dnmt1~KO.$	94
3.19	H3K27me3 strong enrichment at TSSs of lowly expressed genes $\ . \ . \ .$	95
3.20	No change of histone marks in expressed genes across $Dnmt1~{\rm KO}~$	95
3.21	Histone marks are not enriched on regions dependent on DNA mainte-	
	nance methylation	96
3.22	Mapping of TEs in WGBS-seq, RNA-seq and ChIP-seq libraries $\ . \ . \ .$	97
3.23	TE family demethylation upon $Dnmt1$ deletion $\ldots \ldots \ldots \ldots$	98
3.24	Resistance to demethylation of some TEs	99
3.25	TE classes upregulated upon $Dnmt1$ KO \ldots	100
3.26	Unique TE elements upregulated upon <i>Dnmt1</i> KO	100
3.27	Increased genome-wide pervasive transcription upon $Dnmt1~{\rm KO}~\ldots~.$	102
3.28	Pervasive transcription overlaps with TEs	102
3.29	TE orientation antisense to genes	102
3.30	Sense/antisense transcription in TE families	103
3.31	Model of TE regulation by RNAi	104
3.32	Small RNAs map to TEs	105
3.33	Small RNAs map to the IAP consensus sequence	105
3.34	Quality of AGO2 small RNA-seq libraries	106
3.35	AGO2 siRNAs map to TE classes	107
3.36	Analysis of AGO2 bound small RNAs that mapped to repeats $\ . \ . \ .$	108
3.37	Sense/antisense small RNAs map to TE families	108
3.38	IAP is upregulated upon Dicer and Ago2 KD	109
3.39	siRNA knockdown of RNAi pathway	109
3.40	High H3K9me3 enrichment in IAPEZ elements	110
3.41	H3K27me3 and H3K9me3 enrichment over MERVL elements	111
3.42	Repressive histone marks control TEs upon Dnmt1 KO	112
3.43	Schematic of endosiRNAs becoming expressed to control TEs upon	
	Dnmt1 deletion in ESCs	118

4.1	CRISPR-Cas9 KO of <i>Dicer</i> in $Dnmt1^{fl/fl}$ ESCs \ldots \ldots \ldots \ldots	122
4.2	Characterisation of <i>Dicer</i> KO in $Dnmt1^{fl/fl}$ ESCs	122
4.3	Differentially expressed genes between <i>Dicer</i> KO and WT ESCs	123
4.4	GO term analysis of differentially expressed genes between <i>Dicer</i> KO	
	and WT ESCs	124
4.5	Differentially expressed genes between <i>Dicer</i> KO, <i>Dicer/Dnmt1</i> DKO	
	and WT ESCs	124
4.6	Expression of <i>Lin28</i> , <i>Dnmt3l</i> , <i>Fbln2</i> and <i>Oct4</i> in <i>Dicer</i> KO, <i>Dicer/Dnmt1</i>	
	DKO ESCs	125
4.7	Expression of the pluripotency genes in $Dicer$ KO, $Dicer/Dnmt1$ DKO	
	ESCs	126
4.8	Expression of the Dnmts in <i>Dicer</i> KO, <i>Dicer/Dnmt1</i> DKO ESCs	126
4.9	Enrichment of repressive histone marks at genomic features in $Dicer$ KO	
	and <i>Dicer/Dnmt1</i> DKO	127
4.10	Histone mark enrichment in the genome upon $Dicer{\rm KO}$ and $Dicer/Dnmt1$	
	DKO	128
4.11	Transcriptional profile in <i>Dicer</i> KO is not explained by histone mark	
	enrichment	129
4.12	LINE-1 and major satellites are upregulated in a $Dicer$ constitutive KO	130
4.13	Dicer dependent TE upregulation.	131
4.14	H3K9me3 enrichment in TEs dependent on $Dicer$ and $Dnmt1$	132
4.15	H3K9me2 enrichment in TEs dependent on $Dicer$ and $Dnmt1$	133
4.16	H3K27me3 enrichment in TEs dependent on $Dicer$ and $Dnmt1$	134
4.17	Construction of $Dicer^{fl/fl}$ in $Dnmt1^{fl/fl}$ cDKO ESCs	136
4.18	Dicer mRNA and miRNAs expression in $Dicer/Dnmt1)$ cDKO \ldots .	136
4.19	TE expression in $Dicer^{fl/fl}$ in $Dnmt1^{fl/fl}$ cDKO ESCs	137
4.20	Long and short small RNAs map to major satellites	138
4.21	miRNAs map to IAPEZ consensus sequence independent of DNA demethy-	-
	lation	139
4.22	Mmu-miR-7081 complementary to IAP consensus sequence	139
4.23	Mmu-miR-7081 is conserved in the murine lineage	140

4.24	Expression of endogenously expressed mmu-miR-7081 upon $Dnmt1~{\rm KO}$	140
4.25	IAP protein production upon <i>Dicer</i> KO in <i>Dnmt1</i> KO ESCs	141
4.26	Schematic of DICER generated endosi RNAs control TEs upon $Dnmt1$	
	deletion in ESCs	145
5.1	Classification of small RNAs mapped to 2i grown ESCs	151
5.2	21nt small RNAs in 2i grown ESCs	151
5.3	Small RNAs map to SINE elements in 2i medium	152
5.4	miRNAs are the most abundant class of small RNAs mapped to the	
	genome in 2i grown ESCs	153
5.5	Heatmap of miRNAs expression in 2i grown ESCs	154
5.6	Classification of small RNAs in iPSCs reprogramming	155
5.7	No specific fraction of siRNAs in early stages of iPSCs reprogramming.	155
5.8	Classification of small RNAs of iPSCs mapped to the genome	156
5.9	Heatmap of miRNAs expression in iPSCs during reprogramming	157
5.10	Classification of small RNAs mapping to <i>in vitro</i> PGCs	158
5.11	piRNAs from PGCLCs mapped to TEs	159
5.12	Classification of small RNAs mapping to <i>in vivo</i> PGCs	159
5.13	TE bound small RNAs in vivo PGCs and mouse ESCs	160
5.14	IAP bound small RNAs in <i>in vitro</i> and <i>in vivo</i> PGCs	161
5.15	Contribution of 20-24nt small RNAs mapped to repeats	161
5.16	In vivo PGC small RNAs mapped to TEs	162
5.17	Heatmap of miRNAs expression <i>in vitro</i> and <i>in vivo</i> PGCs	163

List of Tables

2.1	Instruments I	53
2.2	Instruments II	54
2.3	Antibodies	54
2.4	Laboratory materials	55
2.5	Kits	56
2.6	Chemicals and Reagents I	57
2.7	Chemicals and Reagents II	58
2.8	Chemicals and Reagents III	59
2.9	CRISPR gRNAs and sequences	60
2.10	qRTPCR Primer and sequences	61
2.11	PCR Primer and sequences	62

Chapter 1

Introduction

The relationship between genotype and phenotype during development has been a question in biology since over 60 years (Waddington, 1942). The whole organism arises from one cell and shares the same genotype, however a plethora of cell types, with specialised cellular functions, are needed to generate a functional multicellular organism (reviewed in Allis and Jenuwein, 2016). Cellular identities are generated through tight regulation of transcriptional programmes. Transcription factor (TF) networks direct cell differentiation and lineage commitment, however "epigenetic modifications" are necessary to reinforce cell-fate decision to prevent reversion into preceding cellular states (reviewed in Allis and Jenuwein, 2016). Epigenetic modifications can be inherited from one cell to the next without any variation in the DNA sequence (reviewed in Allis and Jenuwein, 2016). This propagation of "epigenetic landscapes" during cell divisions rather than genetic inheritance creates cellular identities and drives development (Waddington, 1957). Epigenetic modifications such as DNA methylation, histone modifications, higher order chromatin structure, chromatin interacting factors and non-coding RNAs build the epigenome of a cell. Those epigenetic modifications play together as well as autonomously from each other to form the epigenome. Disruption of proper modifications can lead to a disease phenotype, loss of cellular identity and incomplete early development (reviewed in Allis and Jenuwein, 2016).

1.1 Chromatin

The higher order chromatin structure is governed by an interplay of all epigenetic modifications. DNA is tightly packed in macromolecules, termed chromatin (Flemming, 1882). This organisation allows the compaction of the human two-metre long DNA molecule into the nucleus. The identification of a phosphorus-rich acid, termed nucleic acid as well as the proteins which bind to the DNA are the main building blocks of DNA (Miescher, 1871, Kossel, 1884).

In 1944 it was found that not the proteins but the nucleic acid contains the genetic information (Avery et al., 1944). The molecular basis of genetic inheritance was resolved when the structure of the DNA as a double-helix was identified (Watson and Crick, 1953).

The genome is largely organised in two distinct chromatin states: (1) heterochromatin is densely packed in mitosis (Heitz, 1928) is a silenced chromatin environment, while (2) euchromatin is the diffusely organised chromatin during interphase. Heterochromatin has been referred to as the inactive state of the DNA (Heitz, 1929, Heitz, 1932). This hypothesis was confirmed by the observation that a gene close to heterochromatin was silenced and could be activated by decondensation of the DNA region in *Drosophila* melanogaster (Schultz, 1936). Heterochromatin was further subdivided into (1) constitutive heterochromatin, of constantly condensed regions in all cell types, typically associated with centromeres and telomeres, and (2) facultative heterochromatin that contains regions that switch between active and inactive states during development, like the X chromosome in females as well as gene imprinting (Brown, 1966). To the present day this genome organisation is valid as heterochromatin is found to be typically a gene-poor and transcriptionally silenced part of the genome, while euchromatin is gene-rich and allows transcriptional activity. The DNA is organised in these two distinct chromatin states by post-translational modification of the histone proteins and by modification of the DNA itself (reviewed in Allis and Jenuwein, 2016). Additionally, non-coding RNAs are involved in organisation of the chromatin; the long non-coding RNA Xist has been found to be invaluable for the silencing of the X chromosome (Galupa and Heard, 2015). Furthermore, spatial organisation of the chromatin in the

nucleus is not random. Local and long-range interaction of sequences in the DNA is highly controlled and the DNA can be divided into discrete topological domains of 100kb to 1Mb in length (Dixon et al., 2012, Nora et al., 2012) and genes that sit in domains which are bound by the nuclear lamina have been found to be transcriptionally silent (Guelen et al., 2008, Pickersgill et al., 2006).

Even though hundreds of chromatin modifications have been identified to date, the modifications associate in preferential combinations with each other and hence give rise to a small number of chromatin states (Rando, 2012, van Steensel, 2011). Indeed, four chromatin states were identified in *Arabidopsis thaliana*, five states were found in *Drosophila melanogaster* and the human chromatin organises in six chromatin compartments (Rao et al., 2014).

1.2 DNA methylation

DNA methylation entails the only covalent modification of the DNA itself and presents the most studied epigenetic mark. DNA methylation is the addition of a methyl group to the 5-carbon of the cytosine base in the DNA (5mC) (Holliday and Pugh, 1975, Riggs, 1975). In mammals 4-6% of cytosines are methylated in somatic cells. DNA methylation (5mC) plays a key role in developmental processes in mammals. This epigenetic modification controls genome regulation, genome stability, X-inactivation, genomic imprinting as well as transposon silencing (reviewed in Reik and Walter, 2001). In the mammalian genome, 5mC is mostly found in cytosine residues of CpG dinucleotides, while methylation of other DNA bases has been found at low levels (Guo et al., 2014). CpG dinucleotides are mostly found in CpG islands (CGIs), while the majority of the genome is depleted of CpG sites (Bird et al., 1985, reviewed in Goll and Bestor, 2005). CGIs are mostly associated with promoters and remain unmethylated while CpGs randomly distributed in the genome are methylated (Meissner et al., 2008, reviewed in Suzuki and Bird, 2008).

1.2.1 Propagation of DNA methylation

DNA methylation is established and maintained by DNA methyltransferases (DNMTs). DNMTs are the enzymes that catalyse the addition of a methyl group to the DNA using S-adenosylmethionine as a donor for an activated methyl group (reviewed in Bestor, 2000).

Three major DNMTs exist to maintain DNA methylation in the mammalian genome. The establishment of *de novo* DNA methylation is accomplished by the two active DNA methyltransferases DNMT3A and DNMT3B as well as their catalytically inactive partner DNMT3L. However, during every cell division the newly synthesised DNA strand has to be remethylated to maintain symmetrical DNA methylation patterns both strands of a CpG dinucleotide are methylated - by the DNA maintenance methyltransferase DNMT1 (Figure 1.1). The DNA methylation can be lost passively, through dilution during replication in the absence of the maintenance machinery, or actively, through enzymatic demethylation.

Another DNMT enzyme, DNMT2 has been found to methylate RNAs, specifically tR-NAs, instead of DNA (Goll and Bestor, 2005).

Furthermore, DNMT3C has been discovered to be involved in transposable element silencing of specifically young TE classes in the mouse genome (Barau et al., 2016).



Figure 1.1: DNA methylation. DNMT3A and DNMT3B work together with DNMT3L to establish *de novo* DNA methylation. DNMT1 together with E3 ubiquitin-protein ligase (UHRF1) is responsible for the maintenance of DNA methylation after DNA replication (blue). If the maintenance methylation is impaired DNA methylation gets lost through passive dilution (yellow). Figure from (Wu and Zhang, 2014).

De novo methylation

De novo methylation is known as the establishment of methylation patterns in mammals. DNA *de novo* methylation is established by the methyltransferases DNMT3A and DNMT3B (Okano et al., 1999, Li et al., 1992). Both enzymes can bind hemimethylated as well as non-methylated DNA (Okano et al., 1999), but unlike DNMT1, they do not preferentially bind hemimethylated DNA. DNMT3A and B both contain a PWWP motif, a PHD-related domain called ADD (Atrx-Dnmt3-Dnmt3l) and a catalytic domain (Okano et al., 1998).

Deletion of Dnmt3a and Dnmt3b during early mammalian development showed that these two enzymes overlap in their functions. The knock out of Dnmt3b leads to demethylation of pericentric satellite DNA and severe developmental phenotypes which result in embryonic lethality from mid-gestation onwards (Okano et al., 1999). Knock out of Dnmt3a leads to offspring, with prominent developmental defects which result in their death a few weeks after birth (Okano et al., 1999). However, the double knockout of Dnmt3a and Dnmt3b results in severe developmental effects and lethality from embryonic day (E)11.5 onwards (Okano et al., 1999).

Loss of Dnmt3a and Dnmt3b in mouse embryonic stem cells (ESCs) leads to gradual loss of DNA methylation (Chen et al., 2003). This hypomethylation happens over progressive cell divisions, therefore the remaining methylation is attributed to the activity of the DNA maintenance machinery (von Meyenn et al., 2016).

DNMT3L (DNMT3-like) is a truncated form of DNMT3A and DNMT3B without any enzymatic activity (Aapola et al., 2001). The enzymatic activity of those two methyltransferases is enhanced through binding of their catalytically inactive partner DNMT3L (Chen et al., 2005, Suetake et al., 2004, Gowher et al., 2005). DNMT3L plays an important role in *de novo* methylation of transposable elements during spermatogensis. Knock out of *Dnmt3L* results in viable offspring with a severe germline defect, as well as a lack of imprinting (Bourc'his et al., 2001).

Maintenance methylation

DNMT1 is the enzyme that maintains DNA methylation after DNA replication during S phase of the cell cycle. DNMT1 was the first DNA methyltransferase discovered in mammals (Gruenbaum et al., 1982, Bestor and Ingram, 1983). DNMT1 contains a replication foci-targeting domain (RFD) a DNA-binding CxxC domain, two Bromoadjacent homology domains (BAH) as well as the catalytic domain, which is composed of a target recognition domain and the core catalytic motifs (Song et al., 2011, Takeshita et al., 2011). The maintenance methylase DNMT1 is targeted to the genome by the E3 ubiquitin-protein ligase (UHRF1) (previously known as NP95) which recognises hemimethylated DNA at replication foci via its SRA-domain (Sharif et al., 2007, Bostick et al., 2007). The methyltransferase copies the 5mC mark onto the replicated daughter strand (Bestor et al., 1988, reviewed in Bestor, 2000). The high affinity of DNMT1 to hemimethylated DNA, by binding with the RFD domain to PCNA proteins and UHRF1, has been shown *in vivo* (Fatemi et al., 2001). Furthermore, the CXXC domain of DNMT1 binds to unmethylated CpGs and thereby inhibits the access of the catalytic domain to those regions (Song et al., 2011).

Dnmt1 as well as Uhrf1 -null mice have reduced DNA methylation levels and die in utero (Li et al., 1992, Sharif et al., 2007). The knockout of both Dnmt1 as well as Uhrf1 in mouse embryonic stem cells (ESCs) leads to a rapid DNA demethylation of the genome. Even though self-renewal of those cells is maintained, they fail to differentiate (Lei et al., 1996, Li et al., 1992). The remaining DNA methylation level of 20%, after knockout of the maintenance methylation machinery (Meissner et al., 2005) can be attributed to the activity of the *de novo* methyltransferase DNMT3A and DNMT3B (Lei et al., 1996).

1.2.2 Loss of DNA methylation

The enzymes of DNA methylation are DNMT1, DNMT3a and DNMT3b, however the erasure of DNA methylation has been under debate for many years. DNA methylation

can be lost by passive dilution through the impairment of the methylation maintenance machinery - DNMT1 or UHRF1. However, rapid demethylation dynamics that cannot be explained by passive dilution have also been detected (Guo et al., 2011, Mayer et al., 2000). The discovery of Ten-Eleven-Translocation 1 (TET1) was a ground breaking finding because this enzyme enables active demethylation in mammals (Tahiliani et al., 2009). TET1, TET2 and TET3 comprise the TET dioxygenase family. These enzymes oxidise methylcytosine to 5-hydroxymethylcytosine (5hmC) and subsequently 5-formylcytosine (5-fC) and 5-carboxycytosine (5caC) (Ito et al., 2011, He et al., 2011) Due to their role in DNA demethylation in plants, DNA glycosylases have been proposed to play a similar role in DNA demethylation in mammals (reviewed in Zhu, 2009). The thymine DNA glycosylase (TDG) has been found to remove 5fC and 5caC from the genome (He et al., 2011, Maiti and Drohat, 2011). Active demethylation by TET and TDG have been found during *in vivo* in early embryonic development (Gu et al., 2011, Seisenberger et al., 2012, reviewed in Wossidlo et al., 2011). TDG can also work together with active deamination enzymes such as AID and APOBEC to lead to DNA demethylation (Kohli and Zhang, 2013). Deamination has been found to play a role during global demethylation processes in primordial germ cell (PGC) development (Popp et al., 2010), in the zygote (Santos et al., 2013) as well as during reprogramming of induced pluripotent stem cells (iPSCs) (Kumar et al., 2013).

1.2.3 Role of DNA methylation

The functional role of DNA methylation has long been affiliated to transcriptional repression (Razin and Riggs, 1980), however whole genome bisulfite sequencing has shown that the relationship is not unambiguously defined.

I will highlight four different ways in which DNA methylation could interfere with transcription, which are active in the genome, among others: (1) promoter methylation, (2) gene body methylation, (3) preventing the binding of transcription factors to promoters, and (4) indirect control of promoter enhancer interactions (reviewed in Klose and Bird, 2006).

(1) DNA methylation at the Transcription start side (TSS) is implicated with gene repression. promoter CGI methylation plays a profound role in imprinted genes and X-chromosome inactivation during development. Imprinted genes are monoallelically expressed depending on their parent of origin (reviewed in John, 2017). Imprinting is controlled by differential methylation at imprinted control regions (ICR). The methylation level of the ICR can influence single or multiple genes simultaneously (Bartolomei, 2009, Ferguson-Smith, 2011).

Nevertheless, most of the CGI promoters in the genome are unmethylated through the binding of chromatin modifiers, transcription factors as well as the activity of TET enzymes (Ooi et al., 2007, Stadler et al., 2011, Williams et al., 2011). However, the plethora of unmethylated TSSs in the genome does not necessarily lead to transcriptional activation of the associated genes. Rather, they may be silenced through post transcriptional histone modifications.

(2) Although DNA methylation at the TSS is a repressive signal, gene body methylation correlates with activation of transcription (Lister et al., 2009). During transcriptional elongation histone 3 lysine 36 trimethylation (H3K36me3) recruits DNMT3B which in turn deposits DNA methylation over gene bodies (Baubec et al., 2015). Gene bodies are mostly depleted of CpGs, though intragenic CGIs exist and show tissue specific methylation patterns (Maunakea et al., 2010). Altogether, genome-wide studies showed that methylation of the promoter regions leads to transcriptional regulation and that the local CpG density plays an important role (Meissner et al., 2008, Weber et al., 2007). While methylation of high CpG-density promoters (HCP) is strongly associated with transcriptional silencing, low CpG-density promoters (LCP) do not have this property as they are mostly methylated without any effect on the transcriptional regulation (Borgel et al., 2010). Correlation between transcriptional inactivation and methylation is best described at intermediate CpG-density promoters (ICP), like germline specific promoters (Meissner et al., 2008).

(3) The binding of transcription factors (TFs) can depend on the methylation level and therefore affect transcription. The methylation of a single CpG in the binding site of the TF YY1 abolishes its binding (Kim et al., 2003). Nevertheless, the zinc finger proteins (ZFPs) with a Krueppel-Associated box (KRAB) domain solely bind to methylated DNA motifs (Quenneville et al., 2012). The binding of the majority of TFs seems to be unaffected by the methylation state of the binding site (Thurman et al., 2012).

(4) Another way to control gene expression is through distal regulatory regions called enhancers. Enhancers are CpG-poor, and low methylated regions (LMRs) (Stadler et al., 2011). ChIP-seq experiments of TET enzymes showed their enrichment over LMRs (Chen et al., 2013b; Williams et al., 2011). Additionally, TET enzymatic products - 5hmC and 5fC - have been found in high abundance at enhancers in ESCs (Ficz et al., 2011, Shen et al., 2013, Williams et al., 2011). Their methylation status could influence gene expression and studies of global chromatin structures will help to shed light on this part of gene regulation controlled by DNA methylation.

However, no clear picture of a role of DNA methylation in regulating gene expression can yet be drawn (Deaton et al., 2011) and further investigation has to follow to understand the regulatory function of this epigenetic mark in the genome.

1.2.4 Chromatin modifications

In addition to higher order chromatin structure as well as DNA methylation, chromatin modifications are contributing to the epigenome of a cell.

The basic unit of chromatin is the nucleosome composed of 147 bp DNA that is wound around an octamer of two histone dimers H2A, H2B and H3, H4, respectively (Kornberg, 1974, Luger et al., 1997). The nucleosomes are connected by a linker DNA of about 10-60bp which is often bound by the histone 1 (H1) linker histone (Thoma et al., 1979). A fiber of DNA under the electron microscope appears like "beads-ona-string" first observed by Olins and Olins, 1974. The chromatin is further packaged and condensed. Chromatin binding proteins, like heterochromatin protein 1 (HP1), methyl-CpG-binding protein 2 (MeCP2) and polycomp group proteins (PcG) play a role in additional chromatin compaction (reviewed in McBryant et al., 2006).

Gene expression can be regulated through the orchestrated modification of histone tails. The posttranslational modification of histones is often found in the N-terminal histone tails, which represent 25-30% of the histone protein (reviewed in Zhang et al., 2015). The ability of chromatin to regulate gene expression was originally hypothesised over 60 years ago, when the first histone acetylase (HAT) was identified (Allfrey et al., 1964). Since then, a plethora of histone modifications have been found including acetylation, phosphorylation, methylation, and ubiquitination (reviewed in Arnaudo and Garcia, 2013). Histone modifications can also recruit non-histone proteins to the chromatin to control gene expression. Furthermore, histone protein variants can replace the canonical histone proteins and influence chromatin environments. H3 is replaced by CENP-A in centromeres as well as by H3.3 in actively transcribed genes (reviewed in Biterge and Schneider, 2014).

Altogether, post-translational histone modifications, and histone core organisation can change chromatin states and impact compaction/decompaction and thereby alter promoter accessibility and control gene expression (reviewed in Bannister and Kouzarides, 2011).

1.2.5 Post-translational histone modifications

Histone acetylation

Since the discovery of the first HAT, histone acetylation has been associated with transcriptional activation. The lack of histone acetylation on the inactive X chromosome was a first evidence suggesting that this post-translational modification (PTM) is marking transcriptional activity (Jeppesen and Turner, 1993). The identification of the histone acetyltransferase (HAT) as well as histone deacetylases (HDACs) confirmed the link between histone acetylation and transcription (Brownell, 1996, Kuo, 1996, Taunton et al., 1996). Histone acetylation neutralises the positive charge of the lysine residue and inhibits the binding of histone tails to the negatively charged DNA (Hong et al., 1993). This results in a chromatin decondensation and facilitates the transcription machinery to access the DNA (Norton et al., 1989, Lee et al., 1993a). The activity of HAT and HDAC presented the first confirmation that histone modifications play a role in gene expression with HAT being the "writer" of the histone mark

and HDAC the "eraser" (Marmorstein, 2001, reviewed in Thiagalingam et al., 2003). Histone acetylation recruits bromodomain proteins to the chromatin among these are transcriptional activators (TAFs), methyltransferases (MLLs), SWI/SNF chromatin remodellers as well as helicases (Smarca), which are commonly referred to as chromatin "readers" (reviewed in Muller et al., 2011).

Histone methylation

While histone acetylation is regarded as an activating mark, histone methylation is a more complex PTM, as it can lead to transcriptional repression as well as activation. Histone methylation is a highly specific histone modification. By mono-, dior trimethylation of a single lysine on one of the histone is sufficient for transcriptional repression or activation (reviewed in Bannister and Kouzarides, 2011). The SET (Su(var), Enhancer of zeste, and Tritorax) domain of histone lysine methyltransferases (KMTs) contains the enzymatic activity which catalyses the transfer of the methyl group donor SAM to the amino group of a lysine residue (reviewed in Bannister and Kouzarides, 2011). KTMs are the "writers" of histone methylation (reviewed in Bannister and Kouzarides, 2011).

Histone methylation is erased by histone demethylases (KDMs) (reviewed in Bannister and Kouzarides, 2011). KDMs either have an LSD1 or an JumonjiC (JmjC)-domain to remove methyl groups from modified histones (reviewed in Bannister and Kouzarides, 2011). LSD1 demethylates the active histone mark, lysine 4 on histone 3 (H3K4), which leads to a transcriptional inhibition (Shi, 2004). LSD1 as well as several JmjC-domain containing demethylases can antagonise the repressive H3K9 methylation and thereby activate transcription (Metzger et al., 2005, Fodor, 2006, Klose, 2006).

Histone methylation can attract proteins with chromo, tudor or MBT domains, which present the "readers" of this modification.

Additional histone modifications

Next to histone methylation and acetylation, further PTMs have been studied. Histone ubiquitination is a histone modification which is, unlike the other histone modifications, mostly found on the C-terminal tail of histones. Histone ubiquitination has been found at lower levels than other modifications. Histone 2A lysine 119 ubinquitination (H2AK119ub1) is involved in gene silencing through the interaction with polycomb group proteins (Wang et al., 2004). Furthermore, histone phosphorylation is takes place during DNA damage response and phosphoruated H2A is largely enriched at sites of DNA breakage (reviewed in Allis and Jenuwein, 2016).

1.2.6 Role of histone modifications

Histone modifications are important regulators of gene transcription. In mammals, inactive genes are marked with methylation of H3K9, histone 3 lysine 27 (H3K27) and histone 4 lysine 20 (H4K20), albeit active genes are marked with high levels of acetylation and trimethylaltion of histone 3 lysine 4 (H3K4me3), histone 3 lysine 36 (H3K36me3) as well as histone 3 lysine 79 (H3K79me3) (Noma et al., 2001, reviewed in Kouzarides, 2007, Barski, 2007). H3K4me3 can be found in promoter regions and H3K36me3 is found in transcribed regions (Liang et al., 2004, Pokholok et al., 2005). Genome-wide analysis of distribution of histone marks allowed for prediction of transcriptional states of promoters and enhancers (Heintzman, 2007). H3K4me1 and H3K27ac mark cell type specific enhancers while another class of enhancers has been identified by H3K122ac but lack off H3K27ac (Rada-Iglesias, 2011, Pradeepa et al., 2016). In ESCs the Nucleosome Remodelling and Deacetylation (NuRD) complex has been shown to directly regulate the expression of pluripotency genes by removal of H3K4me1 and H3K27ac from active enhancers (Reynolds et al., 2012).

In ESCs, genes important for developmental are termed "poised" and carry simultaneously activating H3K4me3 and repressing H3K27me3/H2AK119ub1 "bivalent" histone marks (Bernstein et al., 2006a, Azuara, 2006). RNA polII binding to those regions showed that they express low levels of RNA (Stock et al., 2007). This poised state allows rapid lineage specification upon differentiation of ESCs (Bernstein et al., 2006a, Azuara, 2006). Bivalent chromatin domains are not restricted to ESCs but also exist in other cell types (Mikkelsen et al., 2007, Barski, 2007).

In the present study, I concentrated on the effect of global DNA demethylation on H3K9me2/3 and H3K27me3 genome wide, therefore I will expand on those specific modifications.

H3K9 methylation

H3K9 methylation is generally known as a repressive histone mark and several H3K9 methyltransferases are known to repress specific regions of the genome. H3K9 methylation silences pericentric repeats, centromeres as well as telomeres (Bannister, 2001). Pericentric repeats are a combination of large arrays of AT-rich repeats interspersed with transposable elements (TEs), also called major satellites. Centromeres represent shorter repeated DNA fragments termed minor satellites (Guenatri et al., 2004). H3K9 can be mono- di- or trimethylated and KMTs are responsible for the specific methylation level. Additionally, H3K9 modifications have been found to interact with HP1 as well as KRAB-associated protein 1 (KAP1) proteins. HP1 is a highly evolutionary conserved 25 kDa protein with an chromo- and a chromoshadow-domains. HP1 binds the chromatin with its chromodomain (Bannister, 2001, Lachner et al., 2001, Nielsen et al., 2002). With the chromoshadow-domain HP1 recruits additional chromatin binding proteins and can also attract additional HP1 molecules, resulting in a block of HP1 that inhibits transcription (Smallwood et al., 2007). HP1 is only able to bind H3K9 methylation in vivo if it is associated with KMTs (Chin et al., 2007, Sripathy et al., 2006, Stewart et al., 2005). The interaction of HP1 with H3K9me3 is also important for the recruitment of HDACs (Schultz et al., 2001, Stewart et al., 2005).

Even though eight KMTs for H3K9me3 have been identified in mammals only three have been attributed to specific functions in the genome.

The KMT Suv39h1/2 is important for H3K9me3 at pericentric repeats. *Suv391/2* dKO leads to a loss of HP1 binding to pericentric repeats as well as chromosomal instability (Peters et al., 2001, Peters et al., 2003, Rea, 2000, Lachner et al., 2001). Nevertheless,

Suv39h1/2 KO mice are viable and heterochromatic foci are still forming without these KTMs (Pinheiro et al., 2012, Peters et al., 2001).

G9a/GLP are the main methyltransferases responsible for non-pericentric H3K9 monoand dimethylation. G9a/GLP form heterodimers and single KO of either G9a or GLP lead to reduction of H3K9me2 and H3K9me1 (Tachibana et al., 2001, Tachibana et al., 2005). Additionally, G9a and GLP KO mice are embryonic lethal with an inability to complete meiosis in both male and female (Tachibana et al., 2002, Tachibana et al., 2005, Tachibana et al., 2007). Both G9a and GLP can read and write H3K9me1/2, which leads to a self-enforcing H3K9me2 spreading in the genome of up to 5Mb (Wen et al., 2009). H3K9me2 is highly enriched at genes bound by the nuclear lamina, and G9a KO in ESCs leads to an upregulation of those genes (Guelen et al., 2008, Yokochi et al., 2009).

ESET/SETDB1 has been studied extensively, as this KMT is responsible for most of the H3K9me3 in non-pericentric regions (Schultz et al., 2001, Wang et al., 2003). ESET is targeted to H3K9me3 by KAP1. KAP1 contains a RBCC domain, a PHD-bromo domain and an HP1 binding domain. The RBCC domain can bind to KRAB-ZFP (Friedman et al., 1996). KRAB-ZFP recruits KAP1 to specific genomic loci, KAP1 recruits ESET and NuRD and stabilises the interaction of HP1 with H3K9me3 and KAP1 (Ryan et al., 1999, Schultz et al., 2002). *Kap1* and *Eset* KO mice die shortly after gestation (Cammas et al., 2000, Dodge et al., 2004). Mouse ESCs with *Kap1* or *Eset* KO are not viable (Dodge et al., 2004, Rowe et al., 2010). It as been shown that KRAB-ZFPs bind to TEs specifically; and therefore, the KAP1 and ESET mediated H3K9me3 deposition seems to play a major role in TE silencing (Matsui et al., 2010, Najafabadi et al., 2015, Rowe et al., 2010).

H3K27 methylation

H3K27 methylation works in close relationship with protein complexes called polycomb repressive complexes (PRC) to induce gene silencing.

PRC1 and PRC2 are the two polycomb complexes that exist in mammals. PRC1 is

mostly involved in chromatin compaction and catalyses H2AK119ub, while PRC2 is important for methylation of H3K27 (Shao et al., 1999, Cao, 2002).

The PRC2 complex is composed of the four core proteins EZH1 or 2, SUZ12, EED and RbAp46 or 48 (Cao, 2002), with the first three presenting the minimal requirement for PRC2 enzymatic activity. PRC2 is further stabilised by the interacting factor AEBP2 and JARID2 or PCL are proteins involved in recruiting PRC2 to the chromatin (reviewed in Margueron and Reinberg, 2010). EZH1 and EZH2 contain a SET domain for H3K27 di- or trimethylation. However, EZH1 depletion does not affect global H3K27me2/3 in the genome (Margueron et al., 2008), but it is impossible to generate ESCs depleted of Ezh2 (O'Carroll et al., 2001). The interaction of EED with H3K27me3 is essential for methylation of H3K27 (Margueron, 2009). PRC2 binds the chromatin through the SUZ12 and RbAp24/48 proteins (Nekrasov et al., 2005).

H3K27me3 can cover large domains of about 100kb to maintain X-chromosome inactivation, as well as the highly coordinated expression of the *Hox* gene cluster during embryo development (Silva et al., 2003, Boyer, 2006, Bracken et al., 2006). In those cases, PRC2 is recruited to the chromatin by long non-coding RNAs (lncRNAs). *Xist* lncRNA recruits PRC2 *in cis* to the X chromosome (Plath, 2003), while lncRNA *HO-TAIR*, transcribed in the *HoxC* locus, tethers PRC2 *in trans* to the *HoxD* locus (Rinn et al., 2007).

Small domains of H3K27me3 are enriched in promoter regions of ESCs. H3K27me3 promoter methylation is found in 10% of genes in ESCs. Developmental regulatory genes are mostly enriched for H3K27me3 (Boyer, 2006, Marks et al., 2012, Zhao et al., 2007). PRC2 and H3K27me3 generate the chromatin environment to allow gene expression by recruiting PRC1 to genes that need to be silenced.

PRC1 is a multiprotein complex and a key component of the complex are the E3 ubiquitin ligases RING1a and RING1b. The E3 ubiquitin ligases catalyse the H2AK119 ubiquitination. While *Ring1b* KO is embryonic lethal in mice, *Ring1a* KO mice survive (del Mar Lorente et al., 2000, Voncken et al., 2003). Another member of the PRC1 complex is one of the six PCGF protein, giving rise to six PRC1 complexes. (Gao et al., 2012). Additionally, either CBX or RYBP proteins are members of the PRC1 complex (Wang et al., 2010). CBX proteins contain a chromodomain and can therefore bind H3K27me3, which has been confirmed *in vitro* (Bernstein et al., 2006b, Fischle et al., 2003). The interaction of CBX with H3K27me3 recruits PRC1 to PRC2 targets to drive chromatin compaction and thereby gene repression. In ESCs depleted of Cbx proteins PRC2 targets are largely expressed (Leeb et al., 2010). The exact mechanism by which PRC1 leads to transcriptional repression is still under investigation. Ubiquitination is one form of silencing developmental genes and while Hox genes are repressed independently of this histone modification, Cbx KO leads to an activation of the Hox genes (Endoh et al., 2012, Morey et al., 2012).

RYBP-PRC1 complexes can bind to chromatin independent of PRC2 and H3K27me3, however in this thesis I am concentrating on the H3K27me3 modification, therefore I will not describe this role of PRC1.

1.3 Histone modifications and DNA methylation

Post-translational histone modifications and DNA methylation can be attracted to the same chromatin loci and work together to control gene expression.

Negative feedback loops enforce repressive chromatin, for example ESET catalyses H3K9me3 at pericentric repeats, which in turn recruits HP1 to the chromatin. HP1 recruits the *de novo* methyltransferases to the chromatin. The DNA methylation is then recognised by the DNA binding protein MECP2. MECP2 has a methyl binding domain (MBD), which is also able to recruit SUV39 histone methylase (Meehan et al., 1992). This closes the circle through further H3K9me3 deposition and heterochromatic persistence at pericentric repeats (reviewed in Zhang et al., 2015). While the knock out of the H3K9me3 methyltransferase Suv39h1/2 does not lead to a complete global loss of DNA methylation, hypomethylation was found in heterochromatic repeats (Lehnertz et al., 2003). Cooperative effects of DNA methylation and histone modifications can be induced by proteins which multiple binding domains.

UHRF1 can bind to hemimethylated DNA, via a SET and RING-associated (SRA) domain but this protein can also bind to methylated lysine 9 of histone H3 (H3K9me2/3),
via a tudor domain (Arita et al., 2008, Avvakumov et al., 2008). Conditional double knock out studies of *Dnmt1* as well as *Uhrf1* suggested an interaction between UHRF1 and the histone 3 lysine 9 (H3K9) methyltransferase ESET. In a hemimethylated genome, due to deletion of *Dnmt1* the binding affinity of UHRF1 to hemimethylated DNA confounds DNA binding of ESET and thereby H3K9me3 deposition (Sharif et al., 2016) (Figure 1.2).



Figure 1.2: Maintenance DNA methylation and H3K9me3. The DNMT1 protein has an RFD domain (brown), an CXXC domain (yellow), a BAH1 domain (pink), a BAH2 domain (blue) and the enzymatic domain (green). UHRF1 has a UBL domain (brown), a tandem tudor domain (light blue) a PHD domain (pink) an SRA domain (dark blue) and a RING domain (orange). UHRF1 binds with its SRA domain to hemimethylated CG sites (hmCG) and directs DNMT1 to the sites by binding it with the RFD domain. UHRF1 can also bind H3K9me3 with its RING domain. Figure from Du et al., 2015.

Additionally, KMTs can directly interact with DNMTs to repress transcription. G9a/GLP interacts directly with the maintenance methylase DNMT1 (Estève et al., 2006) and is needed for *de novo* DNA methylation of retrotransposons in ESCs (Dong et al., 2008).

Histone modifications can recruit DNMTs to repress transcription. Both DNMT3A and DNMT3B can bind to H3K36me3 *via* their PWWP domain (Baubec et al., 2015, Dhayalan et al., 2010). This interaction has been shown in yeast, where DNMT3B is recruited to sites of transcriptional elongation by H3K36me3 (Morselli et al., 2015).

However, histone modifications can also antagonise DNA methylation and the same region. The enzymatically inactive isoform DNMT3L can bind to the tails of histone H3 through the ADD domain. This interaction is disrupted if the lysine 4 (K4) at the N terminus of histone H3 is methylated (Ooi et al., 2007). Therefore, DNA methylation is found at regions depleted of H3K4 methylation. This occupation of different regions in the genome is visible at CGIs, where low levels of DNA methylation but high levels of H3K4 methylation is found (Weber et al., 2007) (Figure 1.3).



Figure 1.3: *De novo* DNA methylation and H3K4 methylation. The DNMT3A protein has a PWWP domain (blue), an ADD domain (brown) and the enzymatic domain (green). DNMT3L has a ADD domain (brown), a an enzymatically inactive DNA methyl-transferase domain (pink). DNMT3A can bind to unmethylated DNA and lead to *de novo* methylation which is enhanced by binding to DNMT3L. DNMT3L can also bind H3K4 in its unmethylated state. Figure from Du et al., 2015.

1.4 Non-coding RNAs

Non-coding RNAs are also involved in shaping the epigenome of a cell.

The genome comprises only 1.5% protein coding genes, about three-quaters of the genome produces non-coding RNAs (ncRNAs) (Djebali et al., 2012). The size of ncR-

NAs ranges from 21-24nt of small RNAs to 2.3 to 17.2 kilobases of long non-coding RNAs (lncRNAs) (reviewed in Costa, 2007, Fu, 2014).

1.4.1 lncRNAs

lncRNAs can have *cis* or *trans* regulatory functions on gene activity (Wang et al., 2014, Lu et al., 2014, Fort et al., 2014, reviewed in Wang and Chang, 2011, Melé and Rinn, 2016). The first lncRNA of the H19 locus was discovered in 1991 (Brannan et al., 1990). lncRNAs can work as scaffolds for higher-order chromatin organisation (reviewed in Nagano and Fraser, 2011). The lncRNA X-inactive specific transcript (Xist) plays an important role in X-chromosome inactivation. The inactivation of one X chromosome in female embryos is done to obtain dosage compensation. The silencing is accomplished by the binding of the noncoding RNA Xist with a preceding histone remodelling by loss of the active histone marks H3K4 di- and trimethylation concurrent with the gain of the repressive histone marks H3K9me2 and H3K27me3, and ubiquitination of H2A (Brown et al., 1991), reviewed in Galupa and Heard, 2015).

1.4.2 small RNAs

Since the discovery in 1998, RNA interference (RNAi) has been found to be important in gene regulation, heterochromatin formation as well as regulation of transposable elements (TEs) (reviewed in Malone and Hannon, 2009). RNAi mediated mechanisms are defined as mechanisms in which small RNAs (20-30 nucleotides long) serve for recognition and regulation of its target RNA. Small RNAs are furthermore defined by association with ARGONAUTE (AGO) family proteins. At least three different classes of small RNAs exist in mammals: endogenous short-interfering RNAs (endosiRNAs), microRNAs (miRNAs) and PIWI-interacting RNAs (piRNAs). Small RNAs play a role in post-transcriptional gene silencing by direct base-pairing with its targets. However, small RNAs can also work as transcriptional silencers (TGS) by directing chromatin modifications to their targets (Figure 1.4) (reviewed in Slotkin and Martienssen, 2007).

1.4.3 miRNAs

miRNAs have been found to play a role in almost all cellular processes ranging from development to oncogenesis through post-transcriptional gene regulation (Wightman et al., 1993, reviewed in Ameres and Zamore, 2013). miRNAs are the most abundant class of small RNA in somatic tissues in mammals (miRNAs) (reviewed in Ha and Kim, 2014). The first miRNA *lin-4* was discovered in *Caenorhabditis elegans*, (Lee et al., 1993b). Over 1500 miRNAs are expressed in the mouse genome (Griffiths-Jones, 2004) and it has been shown that most mRNAs contain evolutionary conserved miRNA target sites (Friedman et al., 2009).



Figure 1.4: Small RNA pathway in mice. Different small RNA pathways are active in the mouse genome. The miRNAs (pink) and endosiRNAs (esiRNAs) (pink) are dependent on DICER, while piRNAs (blue) are dependent on MIWI, MILI and MIWI2. Figure from Iwasaki et al., 2015.

Post-transcriptional regulation by miRNAs

miRNAs control gene expression by post-transcriptional gene silencing (PTGS). PTGS of mRNAs is accomplished by partial or full complementary binding of the miRNA and AGOs are acting as the effector proteins. By recruitment of GW182 proteins AGOs induce translational repression, mRNA deadenylation, mRNA decapping and mRNA decay by 5'-to-3' degradation (reviewed in Ameres and Zamore, 2013).

The 5' pairing of nucleotide 2-7 of the miRNA to the target mRNA is conserved and called the miRNA "seed" (Bartel, 2009, Grimson et al., 2007). Additionally to the seed pairing a 3' pairing at nucleotide 13-16 enhances target repression further repression is accomplished by a 5' adenosine (Grimson et al., 2007, Wee et al., 2012, Lewis et al., 2005, Nielsen et al., 2007). miRNAs mostly bind to the target mRNAs in their 3

1

UTR (Grimson et al., 2007, Forman and Coller, 2010). Additionally, the general accessibility of the miRNA binding site in the mRNA sequence is of importance. However, Watson-Crick complementarity of the seed region to the target mRNA suffices for post-transcriptional gene repression, therefore half of all protein-coding genes in mammals can be regulated by this class of small RNAs (reviewed in Ameres and Zamore, 2013). In plants miRNAs which bind with perfect complementarity to their targets trigger mRNA cleavage (Tang et al., 2003, Llave et al., 2002), albeit mammalian miRNAs mostly block translation by lack of complementarity to the target mRNA (Shin et al., 2010, Yekta et al., 2004, Davis et al., 2005). Furthermore, out of the four mammalian AGO proteins, three are catalytically inactive and only AGO2 has endonuclease functions. Ribosome profiling showed, that mRNAs are mostly translationally repressed through mRNA decay by miRNAs in mammals (Baek et al., 2008). Nevertheless about 15% of miRNAs block translational initiation (Guo et al., 2010).

Although miRNAs are described as translational repressors, binding of certain miRNAs to the target mRNA leads to activation of transcription. miR128 in the mammalian brain controls the nonsense-mediated decay machinery and hence increases mRNA abundance (Bruno et al., 2011).

In general, miRNAs seem to not completely disrupt gene expression but rather "tunes" their targets (Bartel, 2009). Furthermore, miRNAs can serve as diagnostic tools and potentially interfere with disease outcomes as was shown for hepatitis C infection (reviewed in Shrivastava et al., 2015).

miRNA biogenesis

Several key proteins are involved in the biogenesis of miRNAs. The miRNAs are encoded as miRNA genes in the genome. RNA polymerase II (Pol II) is responsible for the transcription of miRNA genes into primary miRNA (pri-miRNA). The primiRNA is capped with a 7-methylguanosine cap (m7Gppp) and a 3' poly(A) tail. The miRNA transcripts folds into an integral hairpin structure which incorporates the miRNA sequence. The RNase III endonuclease DROSHA cleaves the pri-miRNA into the 60nt long stemloop precursor miRNA (pre-miRNA) (Lee et al., 2003, Denli et al., 2004, Gregory et al., 2004). DROSHA can cleave more than one pre-miRNA out of the pri-miRNA stem loop.



Figure 1.5: miRNA biogenesis. miRNAs are transcribed and the pri-miRNA is bound by the complex composed of DGCR8 (black) and DROSHA (dark grey). The pri-miRNA is cleaved into the pre-miRNA stemloop, which is exported from the nucleus to the cytoplasm. The pre-miRNAs is then further cleaved by DICER (white) to give rise to the miRNA/miRNA* duplex, and one of the strands is then loaded into the miRISC complex with the main protein being AGO (red).

However, DROSHA is part of the multisubunit complex called the microprocessor complex. The microprocessor complex can be made of a large complex or the smaller heterodimer complex comprised of DROSHA bound to the dsRNA-binding protein DiGeorge syndrome critical region gene 8 (DGCR8). DGCR8 can bind dsRNA and therefore directs DROSHA to the pri-miRNA (Landthaler et al., 2004, Han et al., 2004, Gregory et al., 2004, Denli et al., 2004). Animal pri-miRNAs are processed in the nucleus (Lee et al., 2003, Billy et al., 2001, Provost et al., 2002, Lee et al., 2002). The pre-miRNA is then exported from the nucleus by the RAN-GTP protein EXPORTIN-5 (EXP-5) (Bohnsack et al., 2004, Lund et al., 2004, Okada et al., 2009). Dicer is a cytoplasmic ribonuclease which generates 22 nucleotide miRNA/miRNA* duplex from the pre-miRNA (Bernstein et al., 2001, Hutvágner et al., 2001, Grishok et al., 2001). Even though both miRNAs are produced as duplexes only one of the strands is finally loaded into the RNA-induced silencing complex (RISC) (Hammond et al., 2000, Tuschl et al., 1999). ARGONAUTE (AGO) proteins are making up the core of the RISC. AGOS have two domains, the PIWI and the PAZ domain. The PIWI domain encodes an RNase H protein, and harbours the nuclease activity, while the PAZ domain binds to the 3' overhang of two nucleotides of an siRNA (Lingel et al., 2003, Song et al., 2004) (Figure 1.5).

1.4.4 Endogenous siRNAs

Endogenous siRNAs (endosiRNAs) present another class of small RNAs involved in PTGS. EndosiRNAs were first discovered in *Drosophila melanogaster* (Czech et al., 2008, Kawamura et al., 2008, Okamura et al., 2008). EndosiRNAs are around 21nt long and originate from convergent sense/antisense transcription of genomic loci and can be formed through inter- or intramolecular interactions. TEs can give rise to the double-stranded RNA (dsRNA) that feeds into the endosiRNA pathway.

Post-transcriptional gene silencing by endosiRNAs

EndosiRNAs can control transcription of both protein-coding genes and TEs by PTGS. Mutations in proteins important for the endosiRNA pawthay in flies does not impair fertility and leads to viable offspring (Förstemann et al., 2005, Okamura et al., 2008, Lee et al., 2004, Liu et al., 2003), even though TEs are silenced by endosiRNAs *in vitro* (Rehwinkel et al., 2006).

The idea arose that piRNAs and endosiRNAs might cooperate to compensate for each other's loss of function (reviewed in Malone and Hannon, 2009). ' In mammals, the endosiRNA pathway is able to take over the role of the piRNA pathway in the female germline (Tam et al., 2008, Watanabe et al., 2008). The MT transposon family is highly upregulated in *Dicer* KO in oocytes (Murchison et al., 2007), although *Piwi* KO does not lead to transcriptional upregulation of this TE family. dsRNA feeding into the siRNA pathway can be generated through sense transcription of protein-coding transcripts while the antisense species can arise from pseudogene transcripts in the mouse oocyte (Tam et al., 2008, Watanabe et al., 2006).

endosiRNA biogenesis

The generation of endosiRNAs is similar to miRNA biogenesis. However the two small RNAs originate from different sources. While the precursor of miRNAs is encoded in the DNA, endosiRNAs arise from double stranded transcription. DICER is an essential player in the biogenesis of siRNAs (Grishok et al., 2001, Hutvágner et al., 2001). Dicer processes double stranded RNA (dsRNA) into 21-23nt long duplexes of siRNAs (Bernstein et al., 2001). siRNAs are produced as duplexes and only one strand is loaded into the RNA-induced silencing complex (RISC). The so-called guide strand recognises its target mRNAs by Watson-Crick base pairing. Perfect pairing leads to endonucleolytic cleavage of the target mRNA which is also referred to as slicing (Figure 1.6).

AGO proteins are making up the core of the RISC. Four AGO proteins (AGO 1-4) exist in mammals and all of them can bind small RNAs to trigger translational repression. Yet, in mammals AGO2 is the only endonuclease that possesses the enzymatic activity to cleave its target RNA (Meister et al., 2004, Liu et al., 2004, Song et al., 2004, Okamura et al., 2004, Rand et al., 2005, Matranga et al., 2005).



Figure 1.6: endosiRNA biogenesis. endosiRNAs are generated from dsRNA, which is generated either *in cis* by sense (black) /antisense (grey) transcription in the same locus, by generation of a hairpin structure of the transcript or *in trans* through sense and antisense strand derived from different genomic locations. dsRNA is then loaded into DICER in the cytoplasm (white) and processed on the siRNA guide and passenger stand. The guide strand is normally loaded into the siRISC complex with the key component AGO2 (red).

Noncanonical biogenesis of endosiRNAs and miRNAs

Next to the canonical pathways, miRNAs and endosiRNAs can also be generated through non-canoncial pathways.

One of those pathways is DGCR8 and DROSHA-independent, while another one is TUTase-dependent and a third has been described to be DICER-independent (reviewed in Ha and Kim, 2014). A non-canonical biosynthetic pathway has been shown to take place for mirtrons, miRNA precursors consisting of spliced introns. As mirtrons resemble premiRNAs, they do not undergo Drosha dependent cleavage. Instead, they are directly exported to the cytoplasm where they are further processed by Dicer (reviewed in Babiarz and Blelloch, 2009). Furthermore, DICER and DGCR8- independent miR-NAs and siRNAs have been found in mouse ESCs (Babiarz et al., 2008).

1.4.5 piRNAs

miRNAs and endosiRNAs can be found in differentiated as well as undifferentiated cell types, however piRNAs are confined to the germline as well as during cancer progression. piRNAs are transcribed from a long intergenic region called the piRNA cluster. They are longer than the other small RNAs (24-30nt) and have a 2'-O-methyl-group at their 3' end. Furthermore, other than miRNAs and siRNAs, piRNAs are Dicer-independent and mainly control transposable elements (TEs). P-element-induced-

wimpy-testes (PIWI) proteins were originally identified in *Drosophila melanogaster* and are essential for germline development (reviewed in Thomson and Lin, 2009, Cox et al., 1998). In *Mus musculus* three PIWI proteins - MIWI, MIWI2 and MILI - exist. However, the three proteins are expressed at different stages during early development (Figure 1.7). Interestingly, mutations in PIWI proteins specifically affects the male germline, while the female germline remains unaffected (Carmell et al., 2007, Aravin et al., 2007, Kuramochi-Miyagawa et al., 2008, Aravin et al., 2008). MILI or MIWI2 knock down lead to activation of TEs of long interspersed elements (LINE) as well as the long terminal repeat class (LTR) retrotransposons (Carmell et al., 2007, Kuramochi-Miyagawa et al., 2006, Girard et al., 2006, Grivna et al., 2006).

Transcriptional and post-transcriptional gene silencing of piRNAs

piRNAs can regulate transcription on the level of transcriptional gene silencing (TGS) or PTGS. While prepachytene piRNAs mostly function through TGS by *de novo* methylation of their targets, pachytene piRNAs cleave their target and therefore work through PTGS. The main difference between the two classes is that prepachytene piRNAs originate from TEs and are processed by MILI and MIWI2 in the so called gonocyte stages (just after cell cycle arrest). Pachytene piRNAs, on the contrary, mostly originate from piRNA clusters in unannotated regions of the genome and are bound by MILI and MIWI in round spermatids at post-natal day 14 (P14) (Aravin et al., 2007, Kuramochi-Miyagawa et al., 2008, Aravin et al., 2008, Girard et al., 2006, Lau et al., 2006). MILI and MIWI are endonucleases and therefore necessary for cleavage of the target RNA. Only in the presence of MIWI and MILI, LINE-1 (L1) TEs are cleaved in mouse testes (Reuter et al., 2011, De Fazio et al., 2011). Silencing of intracisternal A particles (IAPs) and L1 elements by *de novo* DNA methylation is abolished in *Mili* and *Miwi2* KO in the male germline (Aravin et al., 2007, Kuramochi-Miyagawa et al., 2008, Bourc'his and Bestor, 2004).

The deletion *Maelstrom* - important for piRNA biogenesis - has been shown to induce spermiogenic arrest (Castañeda et al., 2014). Accordingly, PIWI protein and piRNAs

are thought to be invaluable for genome integrity and fertility of the organism. Nevertheless, L1 transcriptional activity during spermatogensis by deletion of a pachytene piRNA cluster does not effect germline development (Xu et al., 2008)

Thus, the regulation of TEs by PIWI-piRNAs and the effect on gametogenesis are still not unambiguously defined (reviewed in Iwasaki et al., 2015).



piRNA biogenesis

Figure 1.7: piRNA biogenesis. Mouse PIWI proteins are expressed at different stages during spermatogensis. MILI is expressed throughout spermatogenesis (orange), MIWI2 is expressed just until P6.5 (violet) and MIWI is expressed from P14 onwards (green). piRNAs are classified into prepachytene piRNAs (red) and pachytene piRNAs (blue). MILI and MIWI2 bind to prepachytene piRNA clusters. MILI performs to ping-pong cycle to silence targets in PTGS and produce piRNAs that associates with MIWI2. MIWI2 then localises to the nucleus and leads to *de novo* methylation of its targets. Pachytene piRNAs regulate their targets through PTGS. Figure from Iwasaki et al., 2015.

piRNAs are generated by two distinct pathways. The primary piRNA pathway and the secondary piRNA ping-pong cycle for amplification.

Primary piRNAs are known to have a bias for a 5' uridine (U) overhang, and secondary piRNAs bind with 10 nt complementarity to primary piRNAs. Additionally, secondary piRNAs are biased towards the sense strand and have an adenosine base at the tenth nucleotide.

Transcription of primary piRNA clusters gives rise to long single-stranded RNA transcripts, which harbour several truncated TE sequences as well as transposons in antisense direction to their coding strands (Brennecke et al., 2007, Prud'homme et al., 1995, Zanni et al., 2013). Those long transcripts are exported from the nucleus and cleaved by a cytoplasmic endonuclease MitoPLD in the mouse (Huang et al., 2011, Watanabe et al., 2011). The resulting precursor RNAs are loaded into *nuage*, a structure found between aggregated mitochondria. A key component of a *nuage* is the RNA helicase MOV10L1 (Frost et al., 2010, Zheng et al., 2010). A 3'-5' endonuclease needed for the generation of the mature piRNAs is still not identified. Deletion of *Armitage* (Drosophila homologue of mouse MOV10L1) leads to reduction of germline piRNAs (Malone et al., 2009).

A methylatransferase, not identified in mouse, then tranfers the 2'-O-methyl group to the 3' ends of the piRNAs and the piRNAs are loaded into the PIWI-piRISC complex (Saito et al., 2007; Horwich et al., 2007). The Piwi-piRISC complexes are transported from the cytoplasm to the nucleus (Sienski et al., 2012).

MILI-piRNA complexes are associated with primary piRNAs and only secondary piR-NAs are associated with MIWI2. Those secondary piRNAs bound by MIWI2 are imported to the nucleus and direct TGS by targeted DNA methylation (Aravin et al., 2008, Kuramochi-Miyagawa et al., 2008, Reuter et al., 2011, De Fazio et al., 2011). The details regarding ping-pong amplification of piRNAs in the mouse are still being debated. MILI is localised in the cytoplasm and can only induce a single ping-pong cycle, but it is thought that MILI-piRNAs can amplify piRNAs by a pathway yet to be identified (De Fazio et al., 2011) (Figure 1.7).

1.5 Epigenetic reprogramming

Although DNA methylation is a covalent modification of the DNA and regarded as a stable epigenetic modification, during epigenetic reprogramming a global methylation erasure followed by *de novo* methylation occurs in two waves in every generation during mammalian development. The first wave happens shortly after fertilisation in the early embryo and a second wave takes place in the primordial germ cells (PGCs) which are the embryonic progenitors for sperm or oocytes (reviewed in Hackett and Surani, 2013, Seisenberger et al., 2013b, Seisenberger et al., 2013a) (Figure 1.8).



Figure 1.8: Epigenetic reprogramming in mammalian development. Schematic representation of the genome-wide waves of DNA methylation and demethylation in the mouse genome in early mammalian development. Image from (Lee et al., 2014b).

1.5.1 Zygote reprogramming

In order to form the totipotent zygote, epigenetic modifications are being reprogrammed. The zygote has the potential to generate an entire organism with multiple cell types. Those cell types all have a different developmental potency and the canalisation of cells into different cell fates is defined by epigenetic features (Surani et al., 2007). Global demethylation of both female (oocyte) and male (sperm) pronuclei occurs just after fertilisation. Nuclear exclusion of DNMT1 leads to a passive loss of DNA methylation in the female pronucleus (Howell et al., 2001). However, global demethylation in the highly methylated male pronucleus happens by active demethylation through the oxidation of 5mC to 5hmC and 5fC, shown by immunofluourescence (IF) (Gu et al., 2011, Inoue and Zhang, 2011, Iqbal et al., 2011, Santos et al., 2013, Wossidlo et al., 2011). This oxidation is most likely accomplished by TET3, which is the only TET enzyme expressed in the zygote and knockdown as well as knockout of *Tet3* inhibits global demethylation of the male pronucleus (Gu et al., 2011, Peat et al., 2014).

Active base excision repair (BER) pathways acting solely on 5mC are also involved in demethylation of the paternal pronucleus, as small molecule inhibition of BER results in global hypermethylation (Hajkova et al., 2010, Santos et al., 2013).

The difference in global hypomethylation in the two pronuclei is of interest. The maternal factor *Stella* presents a prominent explanation. *Stella* protects the maternal genome and paternal imprints with the deposition of H3K9me2 (Hajkova et al., 2010, Nakamura et al., 2012). Deletion of *Stella* from the zygote leads to incomplete preimplantation development and demethylation of the maternal genome and paternal imprinted regions (Nakamura et al., 2006).

Another way to protect imprint demethylation is through the action of the zinc-finger protein 57 (ZFP57) and KAP1 by recruitment of the maintenance methylase DNMT (Quenneville et al., 2011, Li et al., 2008). The hypomethylated zygote has to acquire DNA methylation from blastocyst stage onwards, in order to separate trophectoderm from the inner cell mass (ICM) (Santos et al., 2002). To reinforce lineage commitment gene promoter methylation is highly apparent at developmental genes (Borgel et al., 2010). In the transition from blastocyst to egg cylinder stage further remethylation is occurring, finally giving rise to a hypermethylated genome in the epiblast stage (Smith et al., 2012).

1.5.2 PGC reprogramming

In mice, PGCs are the precursors of the future gametes and in order to gain developmental potency, PGCs also need to go through a wave of global demethylation. Furthermore, the imprinted regions have to be reset (Hajkova et al., 2002).

PGCs can be found in the epiblast of the developing embryo around embryonic day (E)7.25 (Seisenberger et al., 2012, Ginsburg et al., 1990) and share their epigenetic modifications with the epiblast. At E8 PGCs are first localised at the base of the allantois before they reach the developing gonads (E11.5) by migrating through the hindgut (Anderson et al., 2000, Molyneaux et al., 2001). Methylation reprogramming in PGCs occurs in two phases during their development, an early reprogramming event during their migration phase (E7.5 - E8.5) and a late reprogramming phase in their gonadal stage (E8.5 - E10.5) (Guibert et al., 2012, Seisenberger et al., 2012, Kagiwada et al., 2013, Hackett et al., 2013, Seki et al., 2005).

The early phase of reprogramming affects all genomic features and is accomplished by passive demethylation due to impairment of the maintenance machinery by transcriptional regulation of *Uhrf1* (Seisenberger et al., 2012, Hackett et al., 2013). An additional exclusion of UHRF1 from the nucleus results in a great reduction of DNMT1 activity and results in global hypomethylation. Recent *in vitro* studies accompanied by modelling have shown that global demethylation pathways are largely reliant on the maintenance methylation machinery, DNMT1 and UHRF1 (von Meyenn et al., 2016). However, ICR as well as CGI promoters of germ cell and meiosis specific genes further to CGIs associated with the inactive X chromosome stay methylated until the gonadal stage (Seisenberger et al., 2012, Hackett et al., 2013). The mechanism of their resistance to demethylation is not unambiguously defined.

The second phase of demethylation is accomplished through active demethylation, as the dynamics of demethylation do not allow for a passive pathway (Bagci and Fisher, 2013). During epigenetic reprogramming 5hmC levels are increasing from E10.5 onwards (Hackett et al., 2013, Yamaguchi et al., 2013), suggesting the involvement of TET enzymes. Yet, the knockout of *Tet1* in PGCs resulted in minor changes of methylation levels in E13.5 PGCs (Yamaguchi et al., 2012). Furthermore, *Tet1/Tet2* double KO did not abolish PGC differentiation (Vincent et al., 2013). Thus, active demethylation by TET enzymes might be important at very targeted loci, such as ICRs as well as CGIs of gametic genes (Hackett et al., 2013, Vincent et al., 2013). Another mechanism that might play a role in active demethylation are activation-induced cytidine deaminase (AID) and TDG. *Aid* KO results in hypermethylated PGCs at stage E13.5 (Popp et al., 2010) but does not lead to any fertility problems (Kagiwada et al., 2013).

Epigenetic reprogramming does also encompass chromatin remodelling. While H3K9me3 levels stay unchanged over the course of global demethylation in PGCs, H3K9me2 levels follow global demethylation dynamics and get erased from E8 to E13.5 and global levels of H3K27me3 increase (Hajkova et al., 2008, Seki et al., 2007).

In mice, following the wave of demethylation of PGCs to about 5-8% of 5mC at E13.5, de novo methylation begins in male germ cells during embryonic days E14.5 and E16.5 and goes on until prospermatogonia phase (Davis et al., 1999, Davis et al., 2000, Kato et al., 2007). The de novo methylation rises from 7% at E13.5 to 55% at E16.5 before the beginning of meiosis and reaches a final methylation level of 90% in sperm (Kobayashi et al., 2013, Pastor et al., 2014). The remethylation is accomplished by the de novo methylatransferases DNMT3A, DNMT3B and DNMT3L (Kato et al., 2007 Bourc'his et al., 2001, Bourc'his and Bestor, 2004, Kaneda et al., 2004, Ciccone et al., 2009). TE remethylation is accomplished with the help of piRNAs (Bourc'his and Bestor, 2004). Most piRNAs in the foetus are complementary to TEs and in Dnmt3L, Mili, Miwi2 KO LINES, ERVs are transcriptionally active resulting in meiotic catastrophe (Aravin et al., 2007, Bourc'his and Bestor, 2004). Also conditional KO of DNMT3a or DNMT3L leads to TE reactivation, as well as meiotic arrest and finally germ cell apoptosis (Bourc'his and Bestor, 2004, Kaneda et al., 2004). Male and female remethylation happens with different kinetics, as the final methylation level of the two sexes is different (Smith et al., 2012, Popp et al., 2010, Smallwood et al., 2011, Kobayashi et al., 2013) (reviewed in Seisenberger et al., 2013b, Seisenberger et al., 2013a). - The different kinetics of remethylation of the two germlines can be attributed to the activity of Dnmt3L in the male germline (Bourc'his et al., 2001, Bourc'his and Bestor, 2004, Aravin et al., 2007, Aravin et al., 2008). Additionally, Dnmt3C has been found to play a role in methylating especially young elements in the male germline in mice (Barau et al., 2016).

Although the whole genome becomes hypomethylated during epigenetic reprogram-

ming, several regions resist global demethylation. Interestingly, regions relatively resistant to DNA reprogramming harbour two recently acquired transposon families in the rodent lineage - IAPs and ETns. Furthermore, genes within the proximity of IAPs partially escape methylation erasure in PGCs (Guibert et al., 2012, Seisenberger et al., 2012, Hackett et al., 2013). However, there is also a limited number (a couple of hundred) of CGIs not linked to IAPs, which show variable resistance to reprogramming and may thus contribute to transgenerational epigenetic inheritance (Seisenberger et al., 2012, Hackett et al., 2013).

1.6 Transposable elements (TEs)

Transposable elements (TEs) are mobile genomic regions that are able to change their location within the genome by transposition. TEs were discovered in plants in the 1950s (McCLINTOCK, 1951). The presence of mobile elements was also found in bacteria and in *Drosophila melanogaster*, where the cross of a wild strain of fruit flies to a lab strain led to an increase of chromosomal rearrangenemtns, recombination effects and sterility (Shapiro, 1969, Kidwell et al., 1977). The incompatibility of the two fruit fly genomes, termed "hybrid dysgenesis", could be explained by the activity of the DNA transposon called P element TEs (Rubin et al., 1982, reviewed in Majumdar and Rio, 2015). The P element had expanded in genomes of wild fruit flies, while the genome of the lab strain had been protected (Anxolabéhère et al., 1988).

The non-coding genome was termed "junk" DNA in the 1970s (Ohno, 1972). Due to the accumulation of large amounts of non-coding DNA the term "selfish DNA" arose and TEs were called "parasites of the genome" (Doolittle and Sapienza, 1980, Orgel and Crick, 1980).

Whole genome sequencing allowed a closer examination of the TE contribution to the genome, and it was found that TEs comprise $\sim 70\%$ of the mammalian genome (Lander et al., 2001, Chinwalla et al., 2002, de Koning et al., 2011).

33

While TEs are still regarded as the main danger for genome integrity, their regulatory function and role in evolution are being progressively discovered (Fedoroff, 2012b, Chuong et al., 2016, Elbarbary et al., 2016, Ecco et al., 2016).

The mouse genome is comprised of ancestral TEs as well as newly acquired TEs. The mouse genome (2.7 Gb) is about 0.5 Gb smaller than the human genome (3.2 Gb). The lineage separation of rodent and primates happened 65 to 100 Myrs ago (Chinwalla et al., 2002). As the mutation rate in the mouse is very high only 5 % of ancestral TEs persisted in the mouse genome. 44 % of the mouse genome are TEs and 85 % of those are mouse specific. In general, mouse repeats are younger than human TEs, most of them have integrated in the last 25Myrs and still contain around 1000 active LINEs and ERVs, while most of the human TEs are inert (Chinwalla et al., 2002).

1.6.1 TE families

TEs can be divided in to two different classes, according to their mechanism of transposition: Class I transposons, which are also called retrotransposons, work in a "copy and paste" mechanism as an RNA intermediate is first produced and the cDNA is then reinserted in another place of the genome (reviewed in Smit, 1996) and class II transposons are called DNA transposons and can directly move by a "cut and paste" mechanism. While the major class are class I transposons with about 47% of the genome, only about 3% of the genome is comprised of DNA transposons (Lander et al., 2001). Class I transposons, or retrotransposons are further classified into endogenous retroviruses (ERVs) as well as long interspersed elements (LINEs) and short interspersed elements (SINEs) (Okamura and Nakai, 2008).

DNA transposons

The different coding sequences of DNA transposons define their mechanism of transposition. DNA transposons are mostly 1.5-5kb in long. They encode a transposase gene which is flanked by tandem inverted repeats (TIRs). The transposase allows DNA transposons to move by excising and reintegrating in a new genomic location (Muñoz-López and García-Pérez, 2010). The integration leads to a target site duplication (TSDs) which is the addition of a short sequence of DNA, about 4-8bp, at either site of the DNA transposon. All DNA transposons transpose by this mechanism with the exception of *Helitrons* and *Mavericks* that replicate through a single stranded DNA intermediate.

Activity of DNA transposons is linked to transcription, and retrotransposition leads to a gain of a transposon copy while excision leads to DNA double-stranded breaks that need repair. The break can be repaired by homologous recombination and results in the regeneration of the transposon copy, or non homologous end joining (NHEJ), with only the TSDs remaining. These TSD sequences in the genome are called transposon "footprints" (reviewed in Feschotte and Pritham, 2007). DNA transposons can be classified by their different transposases into eleven superfamilies (Wicker et al., 2007) - with the main ones being Tc1/mariner, hAT, piggyback, Maverick and Helitrons. Tc1/mariner existing in most eukaryotic species (Capy et al., 1996, Plasterk et al., 1999). Miniature inverted repeat transposable elements (MITEs) have lost their transposase but have accumulated in many genomes as they use transposases encoded elsewhere in the genome for mobilisation (reviewed in Feschotte et al., 2002).

Endogenous retroviruses

LTR-retrotransposons make up approximately 8-10% of mouse genomes (Chinwalla et al., 2002). While they were thought to be relics of primary infectious retroviruses which lost their *Env* gene and became *bona fide* endogenous genomic sequence (reviewed in Smit, 1996), LTR transposons have been found to acquire *Env* genes and become infection competent during evolutionary time (Malik et al., 2000). Due to this, another name for LTR retrotransposons is endogenous retroviruses or short ERVs. Interestingly, not one single colonisation event, but rather several separate events, can be attributed to the accumulation of the highly polymorphic class of 400 ERV families in the genome (Maksakova et al., 2013, Zhang et al., 2008, reviewed in Goodier and Kazazian, 2008).

The ERV life-cycle is characterised by an RNA intermediate. In mammals, ERV transposons are a few hundred bp to 7kb long. All full-length ERVs encode at least two genes, *Gag* and *Pol*. While the *Gag* gene encodes a polymorphic capsid protein, the *Pol* gene encodes three proteins: a protease (PR), a reverse transcriptase (RT) with an upstream RNAseH domain and an integrase (IN) (Wicker et al., 2007). Transcription of the ERV mRNA is accomplished by RNA pol II which binds to the promoter located in the 5'LTR. Both *Gag* and *Pol* genes are transcribed into one mRNA. Depending on the host the mRNA is either (1) translated by fusion into one open reading frame (ORF) that is subsequently cleaved to generate two open reading frames or (2) ribosomal frame shifting enables the translation of both proteins (Craig, 2015).

Following the translation of the ERV proteins, the reverse-transcription initiates downstream of the 5'LTR at the primer binding site (PBS). Host tRNAs are used as primers resulting in double stranded cDNA (Lund et al., 2000). The integration of ERVs also creates TSDs (Mager and Stoye, 2015). A lot of ERVs lack ORFs and therefore are unable to replicate. However, ERV retrotransposition is solely dependent on the LTR sequences as well as the PBS sequence for reverse transcription, therefore ERVs can mobilise *in trans*. In mammals, the VL30 and MaLR families are so called non-autonomous ERV elements (reviewed in Stocking and Kozak, 2008). Furthermore, ERV copies can be fixed through homologous recombination of the two LTRs as a solitary LTRs in the genome (Mager and Goodchild, 1989, reviewed in Sverdlov, 1998). Those LTRs can be identified in the genome by their TSDs, and a polyadenylation signal (reviewed in Smit, 1996).

Several families of ERVs have been classified. Phylogenetic analysis of the RT, as well as the organisation of RT and IN domain in the *Pol* gene, allowed the classification of ERVs into *gypsy*-like or *copia*-like transposons (Xiong and Eickbush, 1990, reviewed in Havecker et al., 2004). It has long been thought that in vertebrates, ERVs have likely originated from the endogenisation of a retrovirus (reviewed in Boeke and Stoye, 1997), however phylogenetic analyses of LTR transposons suggests that these elements can acquire *Env* genes which allow them to gain infectious potential (Malik et al., 2000). Depending on their original retrotransposon endogenisation, ERV transposons in mammals can be further subdivided into three subclasses. The Class I ERVs are most closely related to gamma- and epsilonretroviruses. Class II ERVs are related to lentiviruses as well as alpha- beta- and deltaretroviruses, while spumaviruses share highest sequence homology with class III ERVs (Gifford et al., 2005, Jern et al., 2005, Bénit et al., 1997, Cordonnier et al., 1995, Ono et al., 1986, Peters and Glover, 1980). ERVs make up around 12% of the mouse genome.

Class I ERVs comprise only 1.2% of the mouse genome and only about 10 to 300 full-length copies are still active (reviewed in Mager and Stoye, 2015). The mouse leukemia virus (MLV) family is the best described ERV1 element. MLVs integrated into the mouse genome about 1.5Myrs ago and are still able to infect other cells. However, MLVs are undergoing endogenisation and therefore differing amounts of MLVs (10 to 100 copies) can be found in different mouse strains (Stocking and Kozak, 2008). RLTR6 and VL30 are also ERV1 elements and share the same LTR elements. While RLTR6 encodes *gag*, *pol* and *env* genes, VL30 does not encode any protein coding genes. Therefore, RLTR6 is autonomous and VL30 coopts the RLTR6 coding genes to transpose *in trans* (Bromham et al., 2001, French and Norton, 1997).

Class II ERVs make up 4.9% of the mouse genome and comprise 8 members which account for around 2000 active ERV2 elements (reviewed in Mager and Stoye, 2015). ERV2 elements are sometimes also referred to as ERVK elements, because the first identified ERV2 class uses lysine-tRNAs as primers. MMTV was initially identified as a *bona fide* retrovirus, however only 2-3 full length MMTV exist in the mouse genome (Nusse and Varmus, 1982).

Early transposons (ETn) and MusD transposons are highly related and make up around 300 copies of the mouse genome. ETns are highly active, however they lack coding sequences. Therefore, their transposition depends on the expression of the MusD proteins (Baust et al., 2003, Mager and Freeman, 2000, reviewed in Mager and Stoye, 2015).

Intracisternal A-type particles (IAPs) are the most studied mouse ERVs. IAPs were first discovered as viral particles in the endoplasmatic reticulum of several cancer cells (Dalton et al., 1961, De Harven and Friend, 1958). MMERVK10C as well as RLTR10 are also members of the ERV2 class, with MMERVK10C representing an internally deleted RLTR10 element. IAP and ETn families are two high copy number ERVs responsible for most of the insertional germ-line mutations described in mice (Maksakova et al., 2006, Zhang et al., 2008).

Class III ERVs are the most abundant as well as most ancient ERV elements in the mouse genome (5.9%). ERV3 elements are sometimes referred to as ERVL elements as human retrotransposition starts with a leucine-tRNAs. As they have resided in the mouse genome for a long time mostly solo-LTRs of ERV3 elements play an important role in embryonic development as well as gametogenesis as early activators of developmental genes important for two-cell stage embryos. Mouse ERV with a leucine tRNA primer-binding site elements (MERVL) and mammalian apparent LTRretrotransposons (MaLRs) are the two members of the ERV3 class. Both elements share a similar LTR sequence, hence it is thought that they arose from a common ancestor (McCarthy and McDonald, 2004). MERVLs have the ability to retrotranspose while MaLR elements are non-autonomous (Bénit et al., 1997, Smit, 1993). The two main MaLR elements - ORR1 and MT - are about 2-2.4kb in length and are ten times more abundant in the genome than MERVL elements.

LINE elements

LINE elements are autonomous TEs that lack LTRs. Based on the RT domain eleven superfamilies of LINE elements have been identified in eukaryotes: CRE, R2, R4, L1, RTE, Tad, R1, LOA, I, Jockey, CR1 (Malik et al., 1999, Kapitonov et al., 2009, Wicker et al., 2007). LINE elements are dominant in mammals and comprise 20% of the genome. L1 elements in human and mouse are the most studied LINE elements. The full-length human L1 is about 6kb long and composed of an 900nt long 5' untranslated region (UTR), two ORFs (ORF1 and ORF2) as well as a 3'UTR with a polyadenylation signal and a polyA tail (Babushok and Kazazian, 2007). The ORF2 encodes 150kDa protein with an reverse transcriptase (RT) domain and an endonuclease domain (EN), while ORF1 encodes for a 40kDa protein with unknown function (Feng et al., 1996, Mathias et al., 1991, Martin et al., 2003, Holmes et al., 1992). R2 and RTE LINE elements only encode ORF2, while most other LINE elements encode both ORFs. To allow retrotransposition both ORF1 and ORF2 are required (Moran et al., 1996). In mice, the promoter of L1 sits in the 5'UTR and the amount of GC-rich tandem repeat monomers in the UTR is directly linked to the promoter activity (De-Berardinis and Kazazian, 1999). Furthermore, the 5'UTR also harbours an antisense promoter with yet unknown function (Li et al., 2014, Speek, 2001). In humans deletion of this promoter and knock-down of *Dicer* led to increased L1 retrotransposition (Yang and Kazazian, 2006) and in mouse 20-24nt siRNAs complementary to L1 sequences were found in the mouse germline (Watanabe et al., 2006). The two ORFs encode a bicistronic mRNA, even though internal ribosome entry sites (IRES) in front of each ORF in mice have been found (Li et al., 2006). After translation ORF1 and ORF2 form ribonucleoprotein (RNP) complexes and associate in cis with their encoding mRNA (Babushok and Kazazian, 2007, Kulpa and Moran, 2006). The RNP complex is transported into the nucleus and the L1 endonuclease nicks the host genome to generate a primer for the L1 reverse-transcriptase (Jurka, 1997). The reaction is known as target primed reverse transcription (TRPT) and the cDNA is generated at the site of integration (Cost et al., 2002, Luan et al., 1993). Integration creates TSDs and leaves a often non-functional L1 without a 5'UTR as 5' truncation or inversion of the 5' sequence happen during integration (Szak et al., 2002). Pseudogenes are sometimes created through an improper association of the LINE proteins with their mRNA but instead transcribe genic mRNAs in trans (Esnault et al., 2000, Garcia-Perez et al., 2007).

L1 insertions account for 20% of the mouse genome, with about 600,000 copies. Of those, about 20,000 L1s are full-length and around 3,000 are competent to retrotranspose (Goodier et al., 2001). The active L1s are divided into L1A and L1F, while the L1V class is extinct. This classification is based on the L1 promoter. About 6.4 Myrs ago the L1F class arose by a replacement of the V promoter with an F promoter, this got further replaced by an A-promoter and then replaced back to an F-promoter, giving rise to the active classes of L1A, L1Gf and L1Tf-types (Sookdeo et al., 2013). In the mouse genome 900 L1A, 400 L1Gf and 1800 L1Tf elements are presumably still active (Goodier et al., 2001).

SINE elements

SINE elements are non-autonomous TEs that rely on the enzymatic machinery of LINE elements. Most of the SINE elements' 5'ends are derived from small cellular RNAs which are transcribed by polymerase III. Mouse SINE elements can be divided into four classes: B1, B2, ID and B4, depending on their RNA pol III promoter. The pol III promoter originates either from tRNAs, ribosomal 5S RNA or signal recognition particle 7SL RNA (Nishihara et al., 2006, reviewed in Kramerov and Vassetzky, 2005). The 7SL RNA gave rise to most abundant SINEs: B1s (650000 copies) and Alu elements (1.1 million copies) in mouse and human, respectively. The ID and B2 SINEs in mouse (350000 copies) harbour a tRNA-related RNA pol III element (Vassetzky and Kramerov, 2013). The B4 SINE elements arose through a fusion of the B1 and ID elements (Kramerov and Vassetzky, 2001).

SINEs are about 500-800bp long and are non-autonomous TEs. While LINE and ERV elements have the ability to retrotranspose, the retrotransposition of SINE elements depends on the activity of LINE elements. This hypothesis was first formed because of the 3'end sequence homology between LINE and SINE elements in the polyA tail. Cell culture studies confirmed a dependency of SINE retrotransposition on LINE activity (Kajikawa and Okada, 2002) and studies have since then found L1 activity responsible for the activity of over a million SINE elements (Zhang et al., 2003, Vinckenbosch et al., 2006). However, not all SINE 3' UTRs are similar to LINE polyA tails, and the short tandem repeat or polyT tail of Alu elements in humans are essential for their mobility (Dewannieux et al., 2003). SINE RNAs are transcribed by their own polymerase, the RNAs are transported to the cytoplasm where they are incorporated into LINE RNPs, and integrated by coopting the LINE EN and RT machinery. Specifically the 3? poly A tail of the Alu elements bind to the L1ORF2p (Doucet et al., 2015, Ahl et al., 2015) Only B1 and B2 SINE elements are still active in the mouse genome and about 1.4 million SINEs are integrated in the mouse genome and occupy 8% of the mouse genome (Gibbs et al., 2004, Gilbert et al., 2004).

1.6.2 Lifecycle of TEs

The accumulation of TEs in the genome is divided into a burst of newly acquired TE activity, followed by a period of decay. Vertical or horizontal transfer leads to acquisition of new TEs in the genome. Vertical transfer is the deriviation of a new TE through modification of an existing one while horizontal transfer happens through endogenisation of TEs from other species or through viral integration. TEs are masters of vertical transfer. It has be proposed that the 5'UTR of LINE elements is not conserved in LINE elements and LINEs acquire new regulatory units by incorrect template switching in the retrotransposition event (Hayward et al., 1997). In the mouse genome L1 familes have been found to exchange coding sequences to give rise to mosaic elements which in turn can regain activity (Saxton and Martin, 1998, Sookdeo et al., 2013). While in most species, TEs horizontal transfer is leading to an increasing variety of TE species, mammals are subjected to relatively fewer instances of horizontal transfer. This might explain why mammals have fewer TE families which in itself are made up of many different members. 50% of LINE elements in both human and mouse originated from different variants of the LINE family trough vertical gene transfer (Khan et al., 2006). An example of horizontal gene transfer is the endogenisation of retrotransposons to give rise to ERVs in vertebrates, like the Bov-B LINE element which was transferred from snakes to cows (Kordis and Gubensek, 1999).

The accumulation of TEs in the host genome is controlled by the insertion rate versus the loss of their functional activity (Charlesworth and Charlesworth, 1983). TEs transpose through insertion of an identical genetic copy into a new genomic region while the original sequence becomes inactivated through genetic drift. This efficient process leads to the distribution of identical copies of TEs, which decay over time by accumulating mutations and blending into the background (reviewed in Smit, 1996). As TEs are not positively selected for in the genome, they acquire mutations over evolutionary time. Therefore, only small amounts of TEs are still being active and account for most of the retrotransposition events with only 80 - 100 L1s being active and 4 - 5 very active (HOT) L1s are still active in the human genome (Brouha et al., 2003). TEs are responsible for 10-15% mutations in mice (Kazazian and Moran, 1998, Maksakova et al., 2006), accordingly the host genome acquired a series of defence mechanisms to control TE activity (reviewed in Craig, 2015). Newly acquired TEs can initially proliferate in the genome, as the genome does not recognise them. However, the genome defence pathways are subsequently set in place and finally TEs acquire mutations which inactivate them. The activity of TEs can lead to mutations in both host genome and TE, therefore TE accumulation is controlled by natural selection (Charlesworth and Charlesworth, 1983). A lot of TEs in the genome are inactive. The genome of mouse and human share about 165Mb of repeat sequences from 100Myrs ag (Chinwalla et al., 2002).

The accumulation of TEs is not random in the genome. Gene-rich regions are mostly spared from TE integration. While L1s were thought to be at higher frequency at pericentric repeats in humans, this could not be confirmed by whole genome sequencing (Lander et al., 2001, Laurent et al., 1997, Schueler et al., 2001). L1s and ERVs are highly enriched at AT-rich sites, while SINEs are enriched at GC-rich sites in the human and mouse genome (Smit, 1999, Soriano et al., 1983). L1 transgene insertion in mouse and human genomes showed that they inserted randomly in the genome (Babushok et al., 2006, Gilbert et al., 2002, Symer et al., 2002). Furthermore, new insertions of LINEs are mostly found throughout the genome, with no preference (Ovchinnikov et al., 2001), and while SINEs are integrate preferentially in AT-rich regions, they are selected for in GC-rich regions (Lander et al., 2001, Chinwalla et al., 2002). ERVs transgenes as well as *bona fide* retroviruses insert preferentially in genic regions, al-though they are selected for preferentially intergenically (Brady et al., 2009, Mitchell et al., 2004, Medstrand et al., 2002, reviewed in Lander et al., 2001).

1.6.3 TEs and the genome: Friends or foe?

The genome is comprised of genes as well as regulatory sequences. Studying the mosaic patterning of maize Barbara McClintock already defined TEs as mutable loci which have the ability to regulate the expression of nearby genes (McCLINTOCK, 1951).

Genome evolution by TEs

TEs are creating genetic variation due to their transposition and have therefore been described as drivers of genomic evolution (Fedoroff, 2012a). In the mouse genome, L1 insertions lead to 2-3% of mutations while ERV insertions make up 10-12% of mutations (Druker and Whitelaw, 2004, Maksakova et al., 2006). Human TEs are largely silent with only 0.3% of mutations in human occuring through transposition of LINEs and SINEs (Callinan and Batzer, 2006).

TEs can directly insert into protein-coding regions or regulatory regions that disrupt transcription (Kazazian et al., 1988). TEs can interfere with splicing, lead to early termination of a gene transcript, work as an alternative enhancers or work as an alternative promoters (Nekrutenko and Li, 2001, Kreahling and Graveley, 2004, Chen et al., 2009, Cohen et al., 2009, Gifford et al., 2013). An example of this presents the insertion of an LTR element into the Dicer mRNA which leads to an alternative Dicer isoform in the mouse oocyte (Flemr et al., 2013). LINE elements can mobilise SINE as well as genic mRNAs in trans and lead to the distribution of cellular RNA in the host genome (Esnault et al., 2000, Wei et al., 2001). TSDs are also described as the footprints of TEs, however L1s and Alus in humans can also transpose independent of the L1 endonuclease and therefore lack this TSD. The retrotransposition is rather linked to DNA double-stranded break repair, as the TEs insert into regions and repair the break (Morrish et al., 2002). Furthermore, insertions of L1s can lead to additional deletions of 1bp to 20kb in the host genome (Symer et al., 2002), while TE flanking genomic sequences can also transpose together with the TE into new locations (Moran et al., 1999, Han et al., 2005). It is referred to as 5' or 3' transduction and happens when TEs make use of alternative 5' promoters or 3' polyadenylation signals. 3' transduction happens with high frequency and 15-20% carry an additional 200bp of genomic DNA sequence into new locations (Goodier et al., 2000, Pickeral et al., 2000, Tubio et al., 2014).

TEs can have a profound effect on the genome when homologous recombination of

non-allelic TEs leads to chromosomal rearrangements. If the two TEs are located on different chromosomes it leads to chromosomal translocations and if it takes place on the same chromosome it results in deletions, duplication or inversions (Sasaki et al., 2010). Throughout human evolution 27% of duplication, 44% of inversions and and over 500 deletion events were linked to improper homologous recombination of L1 and Alus (Bailey et al., 2003, Lee et al., 2008, Han et al., 2008, Sen et al., 2006).

TEs as disease-forming mutations

Insertions of TEs are mostly found in intergenic regions and therefore do not affect the host genome. In humans, the activity of LINEs and SINEs was directly linked to 50 genetic diseases ranging from muscle and blood disorders to immunodeficiencies, visual impairment and cancer (Kaer and Speek, 2013, Lee et al., 2012). While Alu insertions into the Breast cancer genes 1 and 2 (BRCA1-2) can directly cause breast cancer, TE mobilisations can also increase by disruption of cellular defence mechanisms that control TEs activity (Ross et al., 2010).

TEs as regulatory elements

TE insertions can lead to the deriviation of protein-coding genes resulting in beneficial functions in evolution. This happens mostly through gene duplications (Zhang et al., 2003).

Furthermore, L1 *in trans* mobilisation of genic mRNA can lead to intron-less genes without a promoter. In the mouse genome about 3000 so called "retrogenes" have been found; 1000 of those have been found to be transcribed and 120 are now *bona fide* protein coding genes (Vinckenbosch et al., 2006, Zhang et al., 2003).

Furthermore, TEs have been endogenised by the host genome and evolved new functions. The recombination activating Rag1/Rag2 protein was created by the domestication of a DNA transposon 500Myrs ago (Kapitonov and Jurka, 2005, Thompson, 1995). VDJ recombination driven by Rag1 and Rag2 is a key component of the adaptive immune system in vertebrates. Furthermore, the imprinted genes *Peg10* and *Rtl1* arose through endogenisation of ERV transposons during placenta formation (Youngson et al., 2005). Additionally the sycytins are envelope genes of retroviral orignis and are expressed specifically in the mammalian placenta (Cornelis et al., 2014).

1.6.4 Major Satellites

Transposable elements are able to retrotranspose and can present a danger to genome integrity, however the activity of additional repetitive regions in the genome can result into genome instability. Pericentric repeats present AT-rich long repetitive domains also called major satellites (Wong and Rattner, 1988, Joseph et al., 1989). Major satellites are 6Mb long and composed of 234bp repetitive units (Choo, 1997). These regions play a very important role throughout organisation of chromosomes during interphase of mitosis (reviewed in Comings, 1980, Manuelidis, 1990). In ESCs, major satellites are highly enriched for H3K9me3 which works together with HP1 to maintain heterochromatin of the centromeric region (Peters et al., 2001, Taddei et al., 2001). While H3K9me3 and DNA methylation are lost on pericentric repeats upon Suv39h1/2 KO, H3K27me3 gets enriched at those repeats, which suggests a compensatory mechanism to silence these repetitive elements (Lehnertz et al., 2003, Peters et al., 2003) Additionally, in hypometylated ESCs, H3K9me3 is unaffected, yet H3K27me3 becomes enriched at pericentric heterochromatin (Marks et al., 2012, Saksouk et al., 2014, Cooper et al., 2014). In Dnmt triple KO ESCs, major satellites are not transcriptionally active but are enriched for PRC2. It was shown that the DNA binding protein BEND3 recruits PRC2 to major satellites upon DNA demethylation (Dai et al., 2013, Saksouk et al., 2014). BEND3 binding to major satellites is impaired upon DNA methylation (Dai et al., 2013).

1.7 Epigenetic modifications of TEs

Transposable elements (TEs) comprise around 70% of the mammalian genome (de Koning et al., 2011). TEs can randomly transpose in the genome, therefore they are sometimes described as the parasites of the host genome. Mobilised TEs are able to lead to chromosomal breakage, large-scale genomic rearrangements as well as interfere with gene expression by disruption protein coding genes, as well as by altering transcriptional regulatory networks (McCLINTOCK, 1951, Chuong et al., 2016, Elbarbary et al., 2016).

In order to preserve genome integrity, the cell must guard the host genome by discriminating TEs from protein-coding genes. However, the plethora of TEs classes and very little sequence similarity between them makes this a difficult task for the genome (reviewed in Malone and Hannon, 2009).

1.7.1 Methylation of TEs

DNA methylation is a stable epigenetic mark that ensures the silencing of transposable elements. Studies suggest that DNA methylation evolved to protect the genome from TEs (Yoder et al., 1997). DNA methylation is thought to play a main role in TE silencing in differentiated tissues. However, differentiated cells are not viable without DNA methylation. *Dnmt1* KO mice show 1000 fold upregulation of IAPs (Walsh et al., 1998, Sharif et al., 2016). Furthermore, KO of *Dnmt3L* lead to activation of LINEs and ERVs in the male germline (Bourc'his and Bestor, 2004). *Dnmt3C* has been found to be involved in methylating especially young TEs in the male germline (Barau et al., 2016). Additionally, *Tet1* KO has been found to lead to transcriptional activation of LINE-1 elements in mouse ESCs (de la Rica et al., 2016). As reported before, TEs are active in severely hypomethylated cancer types (Schulz, 2006). Although *Dnmt1* KO leads to demethylation of IAP elements (Arand et al., 2012), *Dnmt1* constitutive KO in ESCs does not result in IAP activity (Hutnick et al., 2010). The long term deletion of DNA methylation at TEs in ESCs seem to be compensated for by deposition of repressive chromatin modifications (Karimi et al., 2011, Matsui et al., 2010). However, in acute

global hypomethylation of ESCs IAPEZ, LINEs, MERVL as well as MMERGLN are upregulated (Walter et al., 2016).

1.7.2 Histone modifications of TEs

In addition to DNA methylation, histone modifications have been implicated in TE regulation. The KO of *Eset* in mouse ESCs led to upregulation of a large class or ERV1 and ERV2, as well as an upregulation of L1s (Karimi et al., 2011, Matsui et al., 2010). Also Kap1 KO transcriptionally activated all ERV classes, as well as L1 elements (reviewed in Rowe and Trono, 2011). Furthermore, chimeric transcripts derived by TEs used as alternative promoters for genes arise upon *Eset* KO in mouse ESCs (Karimi et al., 2011). Therefore, H3K9me3 plays a major role in TE silencing through ESET and KAP1 pathways, as KRAB-ZFP targets the histone methylase to TE regions. However, HP1 plays only a minor role in keeping TE elements transcriptionally inactive (Maksakova et al., 2013). KRAB-ZFP bind in the 5'region of TEs; and therefore, H3K9me3 is highly enriched at the promoter region of TEs. MLV TEs are silenced by binding of ZFP809 to the proline tRNA binding site (PBS) and recruitment of KAP1 and ESET (Wolf and Goff, 2007, Wolf and Goff, 2009). KRAB-ZFPs evolve very quickly in the genome and are thought of as a genome defense against TEs. They are regarded as an example of the arms race of the genome against TEs; as new ERVs evolved, KRAB-ZFPs appeared (Thomas and Schneider, 2011), therefore young L1s might have still escaped ZFP targeting now.

Two L1 classes, L1PA3 and L1PA4, in the human genome escaped KRAB-ZFP control as they deleted the binding site of the zinc finger protein (Jacobs et al., 2014). On the other hand, ancient L1s, that were highly active before KRAB-ZFPs evolved, are not recognised by ZFP elements, but guarded by KAP1 and H3K9me3 (Bulut-Karslioglu et al., 2014).

SUV39H plays a major role in controlling LINE elements. *Suv39h* dKO in ESCs leads to depletion of H3K9me3 at ERVs as well as L1 elements, but while ERVs do not get transcriptionally upregulated, L1 type A are strongly expressed (Bulut-Karslioglu et al., 2014). TE silencing by H3K9me3 has not been found in differentiated cells yet, as KO of Suv39h, Eset and Kap1 in mouse embryonic fibroblasts or upon differentitation of ESCs did not lead to transcriptional upregulation of TEs (Bulut-Karslioglu et al., 2014, Matsui et al., 2010, Rowe et al., 2013). ERV3 seem to be silenced by H3K9me2, as MERVL elements are enriched for H3K9me2 in ESCs and G9a KO leads to transcriptional upregulation of MERVL elements (Maksakova et al., 2013). H3K9me3 is not enriched at ERV3 transposons, neither in differentiated or undifferentiated cells, nevertheless KO of Kap1, Suv39 as well as *Eset* led to minor MERVL transcription. Therefore, it is thought that these elements are silenced by all three H3K9 methylation pathways but through an indirect yet to be identified mechanism

(Bulut-Karslioglu et al., 2014, Maksakova et al., 2013). Additionally, demethylation of histone methylation has been found to activate ERV elements; the deletion of the $KDM \ Lsd1/Kdm1a$ of H3K4me1/2 led to upregulation of MERVL elements (Macfarlan et al., 2012). Finally, upon global hypomethylation in ESCs MERVL elements have been found to be silenced by redistribution of H3K27me3 (Walter et al., 2016).

1.7.3 Regulation of TEs by small RNAs

Further to DNA methylation and histone modifications, small RNAs have been described as playing a role in TE control. Two recent studies in the mouse oocyte have indicated a role of miRNAs as well as endogenous siRNAs (endosiRNAs) to control TE expression (Flemr et al., 2013, and Stein et al., 2015). In the mouse oocyte and in preimplantation embryos, endosiRNAs have been found to control IAPs and MERVL elements (Tam et al., 2008, Watanabe et al., 2006, Svoboda et al., 2004). In ESCs SINE derived endosiRNAs were discovered; two tandem inverted SINE elements give rise to a hairpin siRNA that feeds into the endosiRNA pathway (Babiarz et al., 2008). Additionally, studies have shown that upon *Dicer* KO in ESCs, L1 and IAP mRNA were upregulated (Kanellopoulou et al., 2005) and in human cancer cell lines LINEs have been found to be regulated by endosiRNAs (Yang and Kazazian, 2006). Artificially generated L1 siRNAs have been found to reduce retrotransposition of L1 in cell culture (Soifer et al., 2005). Furthermore, miRNAs have been found to potentially play a role in TE silencing as 12% of the miRNAs so far reported share homology with TEs (Smalheiser and Torvik, 2005). In the male germline IAPs are repressed by so-called piwi-interacting small RNAs (piRNAs) which lead to transcriptional gene silencing (TGS) through *de novo* DNA methylation (Aravin et al., 2008, Kuramochi-Miyagawa et al., 2008). The PIWI proteins MILI, MIWI and MIWI2 are important during spermatogenesis and *Dnmt3L*, *Mili*, or *Miwi2* KO leads to LINEs, ERVs and meiotic catastrophe, with impairment of the *de novo* methylation of the promoters of IAPs as well as L1 (Aravin et al., 2007, Bourc'his and Bestor, 2004). Also KO of *Maelstrom*, which is important for perinuclear chromatoid body formation important for piRNA biogenesis, leads to impaired spermatogenesis. Furthermore, the KO of the MAELSTROM leads to retrotransposition of IAP and L1 elements in cell culture (Soper et al., 2008).

1.7.4 TEs during epigenetic reprogramming

During early mammalian development, TEs become transcriptionally activated and contribute to genome regulation. 15-20% of total capped RNA in the ooctye and at blastocyst stage stems from TEs, with LINEs as well as ERVs being expressed (Fadloun et al., 2013). IAP, MusD and MERVL virus particles have been detected in the mouse embryo (Kuff and Lueders, 1988, Ribet et al., 2008).

ERV3 elements are highly active during the first wave of global demethylation in early mammalian development. 13% of RNA in the unfertilised oocyte is comprised of MaLR RNAs (Peaston et al., 2004). MERVL elements present the best studied ERV during embryonic development. Already eight hours after fertilisation, at the onset of zygotic genome activation, MERVL are the earliest transcripts in the oocyte. MERVL are upregulated 300-fold in the oocyte and two-cell embryo, while transcription decreases rapidly in the following cell divisions (Kigami et al., 2003, Macfarlan et al., 2012). Interestingly, this ERV3 upregulation also controls about 300 endogenous genes, which get expressed as chimeric transcripts between MERVL LTR and their exons. MERVL elements are potentially important early activators in development as most of the genes they regulate are confined to the two-cell stage. Furthermore, in serum-grown ESCs a small proportion of cells express high levels of MERVL as well as genes controlled by

MERVL (Macfarlan et al., 2012).

Furthermore, IAPs have been found to be transcriptionally active at blastocyst stage and during PGC development at E13.5 (Peaston et al., 2004, Molaro et al., 2014). Nevertheless, *in vivo* ERV2 have not been found to regulate gene expression of developmental genes.

During epigenetic reprogramming, H3K27me3 and H3K9me3 are both enriched at L1s and ERV1 and ERV2 elements in the absence of DNA methylation (Liu et al., 2014), while in pluripotent and differentiated cells those two histone modifications are mostly mutually exclusive (Hawkins et al., 2010, Mikkelsen et al., 2007).

1.8 Embryonic stem cells

Mouse embryonic stem cells (ESCs) can be isolated from the inner cell mass of the early blastocyst and cultured *in vitro* and serve as an invaluable tool for *in vitro* studies in developmental biology (Evans and Kaufman, 1981, Martin, 1981).

Even though ESCs can self-renew infinitely without immortalisation, which is normally attributed to a tumorigenic potential (Suda et al., 1987), ESCs are still able to contribute to an embryo after prolonged *in vitro* culturing.

It is important to culture ESCs in optimal conditions to preserve the self renewal and pluripotency to prevent differentiation. Mouse ESCs are cultured on a layer of inactivated mouse embryonic fibroblasts (MEFs) that provide nutrients. The addition of serum and Leukaemia Inhibitory Factor (LIF) preserves pluripotency of ESCs (Smith et al., 1988, Williams et al., 1988) and allows culturing ESCs without MEFs. Nevertheless, ESCs need to be grown on plates coated with gelatine to provide elasticity of the surface, which is invaluable to maintain pluripotency (Skardal et al., 2013). Although ESCs are derived from the inner cell mass (ICM) of the blastocyst, which is depleted of DNA methylation, cultured ESCs have a methylation landscape largely similar to somatic cells, with about 75% of CpGs methylated (Stadler et al., 2011, Popp et al., 2010).

The addition of two small molecule inhibitors of FGF and GSK signalling (2i inhibitors,

PD0325901 and CHIR99021) allows ESC culturing in serum-free conditions (Ying et al., 2008, Kunath et al., 2007, Wray et al., 2010). Not only do 2i grown ESCs show a more homogenous morphology than serum grown ESC (Marks et al., 2012), but also the transcriptional profiling led to the conclusion that those 2i grown ESCs resembled the "ground state" of pluripotency (Ying et al., 2008). Additionally, 2i grown ESCs have a much lower methylation level than serum grown ESCs, more closely resembling the methylation of the ICM (Ficz et al., 2013, Angermueller et al., 2016, von Meyenn et al., 2016).

1.8.1 Transcription factor networks regulate ESC pluripotency

Self-renewal is a key property of ESCs which is tightly regulated by a network of transcription factors. The pluripotency network in ESCs is governed by the master regulator OCT4, encoded by the *Pou5f1* gene (reviewed in Pan et al., 2002). Knockout of *Pou5f1* leads to embryonic lethality at the blastocyst stage (Nichols et al., 1998), while deletion in ESCs leads to differentiation. NANOG is the second member of the pluripotency network (reviewed in Saunders et al., 2013). ESC depleted of NANOG start to differentiate but can be maintained (Chambers et al., 2007). However, NANOG knock out mouse embryos die at blastocyst stage (Mitsui et al., 2003, Silva et al., 2009). This is not surprising as OCT4 and NANOG work jointly and regulate the same genes (Loh et al., 2006). The transcription factor SOX2 can regulate *Nanog* expression (Rodda et al., 2005). And OCT4, NANOG as well as SOX2 have been found to control the same genes by co-occupying the same promoters (Boyer et al., 2005, Chambers et al., 2009). Therefore, we can think of the regulation of pluripotency genes as a highly organised network (Boyer et al., 2005).

In mouse ESCs 5 to 20% of the binding sites of NANOG and OCT4 have been found in ERV2 TEs (Kunarso et al., 2010). Therefore, ERV2 might also be important drivers of the pluripotent state. In human, naïve ESCs primate-specific ERVs provide functional binding sites for pluripotency transcription factors and can thereby drive expression of alternative long non-coding (lncRNAs) transcripts (Wang et al., 2014, Lu et al., 2014, Fort et al., 2014).

1.8.2 H3K9 and H3K27 methylation in ESCs

In mouse ESCs, DNA methylation and the repressive histone marks H3K9me2, me3 and H3K27me3 mark distinct genomic loci.

Even though, H3K9 methylation largely overlaps with DNA methylation, H3K9me2/3 is highly enriched in TEs, telomeres, centromeres and pericentric regions and depleted in CGIs.

In ESCs lacking DNA methylation, H3K9me3 is unaffected, however H3K27me3 becomes enriched at pericentric heterochromatin (Marks et al., 2012, Saksouk et al., 2014, Karimi et al., 2011, Tsumura et al., 2006). Additionally, an even broader enrichment of H3K27me3 can be found at chromocentres upon depletion of SUV39h in *Dnmt* triple KO ESCs (Schmitges et al., 2011). This is very intriguing, as the repressive marks H3K27me3 and H3K9me3 are normally mutually exclusive in pluripotent and differentiated cells (Hawkins et al., 2010, Mikkelsen et al., 2007). The mutually exclusive enrichment of those two repressive histone marks might be explained by the ability of HP1 proteins to prevent CBX-PRC1 and PRC2 binding in the zygote (Tardat et al., 2015).

Additionally, DNA methylation and H3K27me3 are mutually exclusive at CGIs in ESCs (Brinkman et al., 2012, Statham et al., 2012). PRC2 is attracted to GC-rich unmethylated regions and therefore H3K27me3 is enriched at GC-sites (Jermann et al., 2014, Lynch et al., 2012, Mendenhall et al., 2010). One way to prevent DNA methylation at promoters in ESCs is the ability of PRC2 to recruit TET enzymes to these genomic regions in ESCs (Neri et al., 2013). Although H3K27me3 is confined to CGIs in ESCs, in mouse differentiated cells, H3K27me3 expands away from CGIs into methylated regions and can encompass 10% of the genome with 15-40kb domains (Tanay et al., 2007, Hawkins et al., 2010, Pauler et al., 2009). Furthermore, upon ESC differentiation DNA methylation gets enriched at promoter regions (Mohn et al., 2008, Ohm et al., 2007). PRC could be important to silence the genes during development with DNA methylation being a long-term silencer in differentiated cells.
Chapter 2

Material and Methods

2.1 Materials

Product	Commercial supplier
Pipettes	Gilson Pipetman P
NanoPhotometer	NanoDrop [®] Technologies
Thermocycler	Biorad
qRTPCR Thermocycler	Biorad
LSR Fortessa Cell Analyser	BD Biosciences
Tablecentrifuge	Eppendorf
Centrifuge	Eppendorf
Vortex	Genius 3
Bioanalyzer	Agilent
PHERAstar FS Plate reader	BMG Labtech
Invitrogen Chamber	Invitrogen

Table 2.1: Instruments I

Product	Commercial supplier
Agarose gel chambers	BioRad
Agarose gel imager	BioRad
Covaris E220 sonicator	Covaris
780 AxioObserver point scanning confocal microscope	Zeiss
Sonicator 3000	Misonix
MACSquant Magnet	Milteney Biotech

Table 2.2: Instruments II

Table 2.3: Ant	ibodies
----------------	---------

Antibody	Commercial supplier	Method
anti-CD4 microbead	Miltenyl Biotec	CD4 pull down
anti-IAP	Dr. Cullens lab	IF
anti-rabbit Alexa-Fluor 568	Life-Molecular Probes	IF
anti-AGO2	Dr. O'Carrolls lab	AGO2 IP
anti-H3K9me3	Active Motif, Clone	ChIP-seq
	MABI0319	
anti-H3K27me3	Active Motif, 39155	ChIP-seq
anti-H3K9me2	Abcam ab1220	ChIP-seq

Materials	Commercial supplier
1.5 ml reaction tubes	Axygen
0.5 ml reaction tubes	Axygen
Falcon tubes	BD Biosciences
TC dishes	Fisher Nunc
Petridish	Scientific Laboratory Supplies Ltd
Gloves (Latex or Nitrile)	Microflex
General glass ware	Fisherbrand
Parafilm	Pechiney Plastic Packaging
Filter Tips	Starlab
Tissues	Kimwipes
Phase Lock tubes	5-Prime
1.8 ml Cryotube	Fisher Scientific UK Ltd
8 ml polyprylene round bottom tubes	BD Biosciences
14 ml round bottom tubes	BD Biosciences
Cell strainer 40 $\mu {\rm m}$	Corning
96 well PCR plates	VH Bio Ltd
384 well qRTPCR plates	Biorad
96 well qRTPCR plates	Biorad
PCR strips and lids	Axygen
qRTPCR lids	Biorad
Cell Scraper	Appleton Woods Ltd
Scalpel	Scientific Laboratory Supplies Ltd
MACS columns	Milteney Biotech

 Table 2.4:
 Laboratory materials

Product	Commercial supplier
Miniprep kit	Qiagen
Gel extraction kit	Qiagen
PCR Purification kit	Qiagen
Qiamp kit	Qiagen
TURBO DNA-free kit	Qiagen
PicoGreen® dsRNA kit	Thermo Fisher Scientific
Platinum SYBR Green qRTPCR	Life Technologies Ltd
small RNA library kit	Illumina
NEBNext	New England Biolabs
Imprint kit	Sigma
QuickExtract	Epicentre
Kapa Library Quantification kit	Kapa Biosystems
MicroPlex Library Preparation kit	Diagenode
SmallRNA qRTPCR miRNA kit	Taqman
SMARTpool siRNA KD	Dharmacon
TruSeq reagents	Illumina
TruSeq RNA library preparation kit v2	Illumina
High Sensitivity DNA kit	Agilent
High Sensitivity total RNA kit	Agilent
Tri-Reagent	Sigma

Table 2.5: Kits

Chemical	Commercial supplier
Phenol/chloroform/isoamylalcohol (25:24:1)	AppliChem
Isopropanol	VWR Chemicals
Ethanol	VWR Chemicals
Chloroform	VWR Chemicals
Glycerol	VWR Chemicals
Agarose	Melford
Natrium Chloride	VWR Chemicals
TritonX 100	Sigma
DMSO	Sigma
Tween [®] 20	Sigma
Ampicillin	Life Technologies
Tamoxifen	Sigma
Penicillin/Streptavidin	Gibco
L-glutamine	Gibco
Sodium pyruvate	Gibco
Beta-mercaptoethanol	Sigma
mouse LIF	Stem Cell Institute, Cambridge
NP-40	Sigma
Tris/HCl	Sigma
EDTA	Sigma
Glycine	Sigma
EGTA	Sigma
Na/Deoxycholate	Sigma
N-lauroylsarcosine	Sigma
LiCl	Sigma
SDS	Sigma
NaHCO ₃	Sigma
RNase A	Thermo Fisher Scientific

Table 2.6: Chemicals and Reagents I

Chemical	Commercial supplier
Proteinase K	Sigma
PFA	Sigma
Gelatine	Sigma
$MgCl_2$	Sigma
DTT	Sigma
Fetal Bovine Serum (FBS)	Gibco
DMEM	Life technologies
NEBuffer	New England Biolabs
Bioline Ladder I, IV and V	Bioline
SYBR Green	Invitrogen
SYBR Gold	Invitrogen
T4 RNA Ligase 2, truncated	New England Biolabs
PvuI	New England Biolabs
EcoRI HF	New England Biolabs
T4 Polynucleotide Kinase	New England Biolabs
T4 DNA ligase	New England Biolabs
Ampure XP beads	New England Biolabs
T5 Exonuclease	New England Biolabs
Shrimp Alkaline Phosphatase	Fermentas
Exonuclease I	New England Biolabs
Klenow exo-	New England Biolabs
Glycoblue	Ambion
DMEM	Gibco
DMEM/F12	Gibco
Neurobasal	Gibco
N2	Stem Cell Sciences
PBS (Phosphate Buffered Saline)	Thermo Fisher Scientific
Lipofectamine 2000	Invitrogen
Protease Inhibitors	Roche

 Table 2.7:
 Chemicals and Reagents II

Chemical	Commercial supplier
Optimem	Gibco
DAPI	Sigma
SlowFade Gold	Thermo Fisher Scientific
BSA	New England Biolabs
HiFi Uracil+ ReadyMix	KAPABiosystems
Vanadyl ribonucleoside complex	New England Biolabs
Trypsin	Gibco
PD0325901	Stem Cell Institute Cambridge
CHIR99021	Stem Cell Institute Cambridge
formaldehyde	Sigma
KCl	Sigma
Na_2HPO_4	Sigma
$\mathrm{KH}_{2}\mathrm{PO}_{4}$	Sigma
T4 Ligase	New England Biolabs
T4 Polynucleotide Kinase	New England Biolabs
MyTaq Redmix	Bioline
Sucrose	Sigma
Orange G dye	Sigma
Protein G-coupled Dynabeads	Thermo Fisher Scientific
HEPES Buffer	Gibco

 Table 2.8: Chemicals and Reagents III

2.2 Oligonucleotides

 Table 2.9:
 CRISPR gRNAs and sequences

Primer	Sequence $(5' \text{ to } 3')$
Dicer KO	
Dicer1_X23_gRNA1_FW	CACCGAGTAATCAAAAGGACCAGCC
Dicer1_X23_gRNA1_RV	AAACGGCTGGTCCTTTTGATTACTC
$Dicer1_X24_gRNA1_FW$	CACCGTTACCAGCGCTTAGAATTCC
Dicer1_X24_gRNA1_RV	AAACGGAATTCTAAGCGCTGGTAAC
Dicer fl/fl	
Dicer_20_21_gRNA1_FW	CACCGAGCAATGATCCGGTCTCAGG
Dicer_20_21_gRNA1_RV	AAACCCTGAGACCGGATCATTGCTC
Dicer_14_15_gRNA2_FW	CACCGCACTCAGCATCGAGTCTCGT
$Dicer_{14}_{15}gRNA2_RV$	AAACACGAGACTCGATGCTGAGTGC
Dicer_20_21_donor_loxP	gacaaggaccactgtactgtttatccctgaagtagcagacta
	gaccattgagatcttgtcaagttagagagcagcaagaattct
	ATAACTTCGTATAGCATACATTATACGAAG
	TTATgagaccggatcattgctcctgtagcagtgatgc
	tggaataggggtgagaatggatatagttcttctcaaaactaa
$Dicer_{14}_{15}donor_{lox}P$	ggcaagaaaagacatttatttctggttgtggggttaaacaaag
	cag cag cag cag cag cag agg cact cag cat cg ag t ct
	ATAACTTCGTATAATGTATGCTATACGA
	AGTTATcgatcgaagccagagctgcacactgcccaattttacc
	tatgetgettattacagttttatggaatatcaaaagtatttaaaatag

	0 (52 + 92)
Primer	Sequence (5' to 3')
$Hspcb_FW$	GCTGGCTGAGGACAAGGAGA
$Hspcb_RV$	CGTCGGTTAGTGGAATCTTCATG
$Atp5b_FW$	GGCCAAGATGTCCTGCTGTT
$Atp5b_RV$	GCTGGTAGCCTACAGCAGAAGG
Dicer_FW	GCATTCCTAGCACCAAGTATTCA
Dicer_RV	GGAAGGAAATTTACTGAGTGGGG
$Dnmt1_FW$	TGAGGAAGGCTACCTGGCTA
$Dnmt1_RV$	ACAACCGTTGGCTTTTTGAG
IAPEZ_F	AAGCAGCAATCACCCACTTTGG
IAPEZ_RV	CAATCATTAGATGTGGCTGCCAAG
MERVL_FW	TTCTTCTAGACCTGTAACCAGACTCA
MERVL_RV	TCCTTAGTAGTGTAGCGAATTTCCTC
ETn_FW	GTGGTATCTCAGGAGGAGTGCC
ETn_RV	GGGCAGCTCCTCTATCTGAGTG
U1_FW	CTTACCTGGCAGGGGGGAGATA
U1_RV	CAGTCCCCCACTACCACAAA
$MajSat_FW$	GACGACTTGAAAAATGACGAAATC
$MajSat_RV$	CATATTCCAGGTCCTTCAGTGTGC
${\rm MmERVK10C\text{-}gag_FW}$	ATGTGAGCTAGCTGTTAAAGAAGGAC
$\rm MmERVK10C\text{-}gag_RV$	CTCTCTGTTTCTGACATACTTTCCTGT
LINE-ORF2_FW	GACATAGACTAACAAACTGGCTACACAAAC
LINE-ORF2_RV	GGTAGTGTCTATCTTTTTTTTTCTCTGAGATGAG

 Table 2.10:
 qRTPCR Primer and sequences

Primer	Sequence (5' to 3')
U6-FW	GAGGGCCTATTTCCCATGATTCC
Dicer_23_24_screen_FW	AGCAGTGCATTGCTGACAAGAG
Dicer_23_24_screen_RV	CTTGTGGTAGTCATACTTCACAGCC
Dicer_14_15_screen_FW1	CCTTTCCCTCTTGCACATTTACCT
Dicer_14_15_screen_RV1	TGAAACCAGACTTCTTCAGCTCG
Dicer_20_21_screen_FW1	GGTGTCAGATCACTTCCCGT
Dicer_20_21_screen_RV1	TGACCAGAATAAGAAGGAGCGGA

 Table 2.11: PCR Primer and sequences

2.3 Methods

2.3.1 Cell culture

ESCs were cultured in standard serum-containing media (DMEM (Gibco, 11995-040), 4 mM L-glutamine, 110 μ g/ml sodium pyruvate, 15% fetal bovine serum, 1 μ g/ml Penicillin- Streptomycin (Gibco, 15140-122), 0.1 mM nonessential amino acids (Gibco, 11140-050), 50 μ M beta-mercaptoethanol (Sigma, M6250) and 103 U/ml mouse LIF (Stem Cell Institute, Cambridge).

For serum-free culturing of ESCs under 2i culturing conditions (Ying et al., 2008), the medium was replaced by serum-free N2B27, composed of 1:1 DMEM/F12 (Gibco, cat. 21331-020) and Neurobasal (Gibco, 21103-049), 1X N2 (Stem Cell Sciences, cat. SF-NS-01-005), 1X B27 (Gibco cat. 17504-044), 1% L-glutamine (Gibco, 25030-081), 1 μ g/ml Penicillin- Streptomycin (Gibco, cat 15140-122), and 50 μ M beta-mercaptoethanol (Sigma, M6250), 103 U/ml LIF and the two inhibitors (MEK inhibitor PD0325901 (1 μ M, Stem Cell Institute Cambridge) and GSK3 inhibitor CHIR99021 (3 μ M, Stem Cell Institute Cambridge).

ESCs were maintained in an incubator at $37 \,^{\circ}$ C with 5 % CO2 humidified atmosphere. ESCs were grown on immortalised mouse embryonic fibroblasts (MEFs) or on 0.1 % gelatine. A tamoxifen-inducible *Cre* recombinase was used to induce recombination of LoxP sites in the floxed cell lines using tamoxifen (800 nM) (Sigma).

Passaging cells

The medium of ESCs was changed every day and ESCs were split (1:4) every 2 days. For splitting ESCs, the cells were washed with 1 x PBS (Thermo Fisher Scientific). TrypLE Express (Thermo Fisher Scientific, 12604013; for serum-free ESCs) or Trypsin (Gibco; for serum-grown ESCs) was added for 3 min at 37 °C and was inactivated by adding 10 x ESC medium. The cells were spun down at 300 x g at room temperature for 3 min and resuspended in ESC medium and passaged to a new gelatine-coated or MEF coated plate.

Thawing cells

To thaw ESCs from liquid nitrogen, 15 ml falcons with 10 ml ESC medium were prepared. Cells were thawed in a 37 °C water bath, resuspended in ESC medium and spun down at 300 x g at room temperature for 3 min. The ESCs were resuspended in ESC medium and plated into 10 cm dishes.

Freezing cells

Cells from a 10 cm dish were split (1:4) and resuspended in 0.5 ml ESC medium. Two times ESC freezing medium (20 % DMSO, 60 % FBS and 20 % ESC medium) was added and the mixture cells were transferred into cryotubes. Cryotubes were transferred into an isopropanol freezing container and frozen at -80 °C overnight (ON). The following day the frozen vials were transferred to the liquid nitrogen storage tank.

2.3.2 Reprogramming of MEF to iPSCs

Wildtype or AidKO MEF cultures were established from female 13.5 to 14.5 dpc embryos. For each transfection, 0.8×10^6 MEF were nucleofected using Amaxa Nucleofection Technology (Lonza AG; program A-023), according to the manufacturer's instructions, with 1 μ g of each plasmid. Plasmids for reprogramming pB-TRE-OCKS, pBASE, pB-CAG-rtTA were obtained from Sanger Institute's plasmid repository. Reprogramming was performed in ES cell medium in the presence or absence of Doxycycline, in a 5 % O2 incubator. The medium was refreshed every other day. Colonies were picked on day 6 of reprogramming and expanded for at least 29 passages.

2.3.3 In vitro PGC like cells

PGC like cells (PGCLC) were generated following the protocol in (Meyenn et al., 2016).

2.3.4 In vivo PGC collection

All embryonic samples for library preparation were collected from timed matings of C57Bl/6J female mice PGCs carrying the Oct4-GFP transgene expressed in the developing gonad (Yoshimizu et al., 1999) described in Seisenberger et al., 2012. Animal work carried out as part of this study is covered by a project license (to W.R.) under the 1986 animal (scientific procedures) act, and is further regulated by the Babraham Institute Animal Welfare, Experimentation, and Ethics Committee.

2.3.5 DNA extraction

The Qiamp kit from Qiagen was used according to supplier's protocol. Elution was done with 100 μ l EB buffer. If DNA was used for CRISPR-Cas9 screening or polymerase chain reaction (PCR) purposes, the QuickExtract (Epicentre, QE09050) solution was used following the supplier's protocol.

2.3.6 RNA extraction

RNA was extracted using Trizol isolation reagent (Tri-Reagent, Sigma) and Phase Lock tubes following manufacturer's instructions (5-Prime), and was subjected to DNAse treatment using the Ambion DNA-free kit according to the manufacturer's instructions (Invitrogen).

2.3.7 Measurement of RNA and DNA concentration

The concentration of DNA and RNA was estimated using the NanoDrop 2000 UV-Vis Spectrophotometer (ThermoScientific). The genomic DNA concentration was measured using the Picogreen kit according to manufacturer's protocol. For library preparation, DNA and RNA were analysed on a Bioanalyzer (Agilent) following the manufacturer's protocols.

2.3.8 qRTPCR

100 ng -1 μ g of DNAse treated RNA was reverse transcribed (Thermo RevertAid K1622) using random hexamer primers. Platinum SYBR Green qRTPCR (Life Technologies Ltd) was used for qRTPCR reaction. The cDNA was diluted 200 times and 2 μ l was used for each qRTPCR reaction. The qRTPCR was run in the Biorad 384 well machine following the manufacturer's condition and measuring SYBR signal. Endogenous controls (Atp5b, Hspcb, U1) were used to normalise expression. Primers are listed in Materials (Table 2.10).

2.3.9 CRISPR KOs

For CRISPR KO the protocol published by Ran et al., 2013 was followed with the subsequent changes. Guide RNAs were designed using the online tools http://crispr.mit.edu/ and http://crispr.dbcls.jp/ as well as

https://chopchop.rc.fas.harvard.edu/. The gRNAs which were predicted best by all three prediction tools, taking into consideration off target effects, were used.

Constitutive Dicer KOs

The gRNAs for the DICER KO (Table 2.9) were then cloned into the pSpCas9(BB)-2A-GFP plasmid (Addgene plasmid ID: 48138). ESCs were cultured on gelatinised plates and transfected with 1 μ g of gRNA construct. Transfection was done using Lipofectamine 2000 transfection reagent (Invitrogen).

Two days after transfection, ESCs were split and resuspended in 1 ml of 1 x PBS with 1% FBS. The cells were then pipetted through a cell strainer (40 μ m, Corning, Product #352340) into a round bottom polypropylene tube of flow cytometry. The cells were sorted on the LSR Fortessa Cell Analyser (BD Biosciences). In order to distinguish live from dead cells, DAPI incorporation was measured. Single cells were sorted for GFP expression into 96-well plates which had been gelatine-coated and filled with 100 μ l ESC medium. Medium was changed every day and colonies were grown until confluent. The cells were then split 1:2, with half of the colonies being expanded and DNA extractions carried out from the other half using QuickExtract. Clones were screened by PCR using screening primer pairs for Dicer_23_24_screen (Table 2.11) and MyTaq Redmix following manufactor's protocols. Positive colonies identified by PCR were send for Sanger sequencing.

Conditional *Dicer* KOs

The conditional KOs were generated the same way as the constitutive KOs with the following changes. The gRNAs for Intron 14_15 and Intron 20_21 of Dicer (Table 2.9) were cloned into the pSpCas9(BB)-2A-CD4 plasmid. Cells were cultured on feeder plates and transfected with 1 μ g of gRNA construct as well as 100 ng donor single stranded DNA for either Intron 14_15 (Dicer_14_15_donor_loxP) or Intron 20_21 (Dicer_20_21_donor_loxP). Cells were trypsinised and resuspended in 70 μ l PBS and stained with human CD4 microbead antibody (Miltenyl Biotec, Cat. No. 130-045-101)

according to manufacturer's instructions. The CD4 positive cells were enriched using MACS columns. Negative cells were collected from flow through. The cells were eluted in 500 μ l PBS, spun down and resuspended in ESC medium. 4000 to 5000 cells were plated in 10 cm dishes to have single cells. The medium was changed every day and after 1 week the colonies were picked. The colonies were picked into Trypsin in 96 well round bottom plates. The colonies were then divided into two. Half of the colonies were plated in 96 well plates coated with gelatine and containing 200 μ l ESC medium, and the other half were transferred into a 96 well PCR plate for DNA extraction and screening. For screening, 96 well plates were centrifuged for 10 min at 3000 x g at room temperature. The ESC medium was discarded and the cells were resuspended in 100 μ l H₂O. 10 μ l of Proteinase K was added and cells were incubated for 30 min at 55°C. Afterwards 5 μ l of the DNA was used for PCR using MyTaq Redmix and screening primers for intron 14_15 and intron 20_21 of Dicer (Table 2.11). As a diagnostic test, 10 μ l of PCR product were digested with *PvuI* in the case of intron 14 15 or *EcoRI* for intron 20 21. 20 μ l of the product was run on a 2 % agarose gel. Positive colonies identified by PCR were send for Sanger sequencing.

Cloning of gRNA constructs

gRNAs primers were ordered with sticky end overhangs and ligated according to (Ran et al., 2013) into the Cas9 plasmid using T4 Ligase (New England Biolabs) following to manufacturer's instructions. The ligation mix was incubated for 10 min at room temperature. The ligation batch was mixed with 100 μ l of Top10-bacteria (Invitrogen) and incubated on ice for 10 min. The heat shock was performed for 45 sec at 42°C and the samples were subsequently put on ice for 2 min. After having added 200 μ l of LB-medium (1 % (w/v) Bacto tryptone, 0.5 % (w/v) Bacto yeast extract, 170 mM NaCl, adjusted to pH 7.6 with NaOH) without ampicillin the bacteria were shaken for 30 min at 300 rpm and 37°C for recovery on the thermocycler. They were plated on LB-agar plates with ampicillin (100 μ g/mL) which had been prewarmed at room temperature (1.5 % (w/v) Bacto agar in LB medium, PBS (130 mM NaCl, 100 mM Na₂HPO₄, pH 7.0). The plates were incubated over night at 37 °C. The next day, 5 ml

LB-medium with 100 μ g/mL ampicillin was inoculated in a round bottom tube (BD Biosciences) with single bacterial colonies, which grew ON on a shaker at 37 °C. The plasmid DNA was isolated using QiAprep Spin Miniprep kit from Qiagen and following the manufacturer's protocol. The U1 reverse primer (Table 2.11) was used to test for positive cloning of gRNAs into the plasmid. The PCR product was send for Sanger sequencing at Sigma.

Testing gRNA efficiency

Firstly, the T7 Endonuclease assay was performed to test for the cutting efficiency of the gRNAs. Thereafter, gRNAs cloned into the Cas9 plasmid were transfected into ESCs. After two days the cells were either sorted for GFP or CD4 expression and the DNA was isolated out of the bulk population of the positive culture using 300 μ l of QuickExtract (Epicentre, QE09050) following supplier's instructions. Afterwards the DNA was incubated with the T7 endonuclease for 5 min and 10 μ l of the sample were run on a 10% agarose gel. The agarose gel was stained with SYBR Gold and analysed using a gel imager to calculate the gRNA efficiency.

Screening for CRISPR clones

For each PCR reaction of 30 μ l sample volume: forward Primer (100 μ M) 0.5 μ l, reverse Primer (100 μ M) 0.5 μ l, template 50 ng, 2 x MyTaq Polymerase mix 15 μ l, H₂O (to 15 μ l). The PCR was run on the Biorad thermocycler with the following conditions: 2 min denaturation at 95 °C, 35 cycles with 30 sec denaturation at 95 °C, 1 min annealing at 55 °C and 30 sec elongation at 72 °C were completed. Before cooling to 4 °C, an additional elongation step for 5 min at 72 °C was accomplished.

Agarose gel electrophoresis

Agarose flat-bed gels in various concentrations (0.6 - 2 % agarose in 1 x TE buffer)and sizes were run to separate DNA fragments in an electrical field (10-20 V/cm) for analytical or preparative use. The desired amount of agarose was boiled in 1X TE buffer until it was completely dissolved. After it cooled to approximately 55 °C, SYBR green (0.5 g/ml final concentrations) was added to the liquid agar, which was then poured in a flat-bed tray with combs. As soon as the agarose solidified, the running buffer (1 x TE buffer) was added. DNA in the 6 x loading buffer (1.5g Ficoll 400, Orange G dye, H₂O to 10mL) was loaded into the wells and separated electrophoretically. 10 x TE: 400 mM Tris, pH 8.0, 10 mM EDTA, acetic acid for pH titration.

2.3.10 Small RNA qRTPCR

For small RNA qRTPCR Taqman miRNA kits were used according to the manufacturer's instructions for mmu_miR93 (Taqman, Cat. No. TM001090), mmu_miR7081_mat (Taqman, Cat. No. TM467052_mat) and snoRNA202 (Taqman, Cat. No. 001232) was used as a positive control.

2.3.11 AGO2 immunoprecipitation

Cells were cultured on 15 cm dishes and harvested in 1x PBS. Pellets were frozen at -80°C until further processing. ESCs were resuspended in 300 μ l lysis buffer (50 mM Tris, pH 8, 150 mM NaCl, 5 mM MgCl₂, 15 % Glycerol, 1 mM DTT , 0.5 % Na deoxycholate, 0.5 % Triton X-100, Protease inhibitor cocktail (Roche), 50 μ g/ml yeast tRNA , 2 mM Vanadyl ribonucleoside complex) and cells were pelleted at 10,000 rpm, 10 min, 4 °C. The supernatant was used as whole ESC extract. 25 μ l Protein G-coupled Dynabeads (Thermo Fisher Scientific, Cat. No. 10003D) were washed 3 x with 1 ml of Wash Buffer (10 mM Tris pH 8, 150 mM NaCl, 1 mM MgCl₂, 0.01% NP-40). 50 μ l of purified AGO2 antibody (kind gift of the Dr. O'Carroll lab) was added, filled up to 1 ml with Wash Buffer and incubated ON at 4 °C in a rotating wheel. On the next day the beads were washed three times with wash buffer and the negative control (beads with extract but without serum) was prepared. The ESC extract was prespun to remove precipitated proteins and 200 μ L extract was added to the beads and filled up to 600 μ L with Lysis buffer. The mix was incubated for 2-4 h at 4 °C in a rotating wheel and subsequently washed five times with wash buffer. The IP was

eluted with 300 μ L Proteinase K buffer (10 mM Tris pH 7.5, 0.5% SDS, 5 mM EDTA, 1 μ l Proteinase K/reaction) after 30 min of 50 °C incubation on the thermomixer, at 850 rpm. RNA was isolated by phenol extraction and eluted in 8 μ l H₂O.

2.3.12 siRNA knock down

For the siRNA knockdown, the Dharmacon SMARTpool siRNAs were used according to the manufacturer's description. For the first round of transfection, cells were split and plated at a density of 1 x 10⁵ onto gelatine coated 12 well plates. 24 h after plating, cells were transfected with siRNAs using Lipofectamine 2000 (Invitrogen, Cat. No. 11668019) and following manufacturer's instructions: 2.5 μ l of 10 μ M siRNAs in 100 μ l Optimem were transfected using 2 μ l of Lipofectamine 2000 per well. Dharmacon siGENOME SMARTpool siRNA against mouse Dicer (Dharmacon, Cat. No. MU-040892-01-0005), Dgcr8 (Dharmacon, Cat. No. M-051365-00) or Ago2 (Dharmacon, Cat. No. MU-058989-01-0005) and siGENOME non-targeting siRNA2 (Dharmacon, Cat. No. D-001210-02-05) were used for transfection in triplicates. The second transfection was done 48h after the first transfection. Cells were split onto 6 well plates. The second transfection was done with the following conditions: 5 μ l of 10 μ M siRNAs in 200 μ l Optimem using 5 μ l Lipofectamine 2000 per well. Cells were harvested 48 h after the second transfection and RNA was extracted using Tri-Reagent.

2.3.13 Immunofluorescence

Cells were grown directly on sterile cover slips coated with 1 % gelatine and were fixed in 2 % PFA (Polysciences, Inc. Cat 18814) for 30 minutes at room temperature, washed in PBS and permeabilised in PBS 0.5 % Triton X-100 for 30 minutes at room temperature. After washing, they were blocked in PBS 0.05 % Tween-20, 1 % BSA (blocking solution, BS) ON at 4 °C and subsequently incubated 60 minutes at room temperature with a rabbit polyclonal anti-IAP antibody (kind gift of the Dr. Cullen lab) diluted 1:200 with BS. After 60 minutes of washing in BS, the cells were incubated with anti-rabbit Alexa-Fluor 568 (Life-Molecular Probes) diluted 1:500 with BS for 60 minutes at room temperature in the dark. After washing ON at 4 °C in PBS 0.05 % Tween® 20, the cells were stained with DAPI (5 μ g/mL) and mounted with SlowFade Gold (Life-Molecular Probes). Slides were imaged on a Zeiss 780 AxioObserver point scanning confocal microscope with a Plan Apochromat 63x/1.40 oil objective. Single optical slices were captured. Images were pseudocoloured using Adobe Photoshop CS4 and levels were adjusted according to internationally accepted guidelines for image manipulation.

2.3.14 Next generation sequencing library preparation

Whole Genome Bisulfite sequencing libraries

For whole genome bisulfite sequencing (WGBS-seq) library preparation, 100 ng of DNA (measured by PicoGreen) were fragmented by sonication (Covaris E220 sonicator), with the following conditions: duty factor, 10; peak incident power, 140; cycles/burst, 200; time, 55 sec; temperature, at 12 °C. The fragmented DNA was end-repaired, A-tailed and the adapters were ligated following the NEBNext Ultra kit protocol (NEB, cat: E7370L). The adapter-ligated DNA was treated with bisulfite reagent following the two step protocol from the Sigma Imprint kit (Sigma, cat: MOD50). The libraries were amplified using the HiFi Uracil+ ReadyMix (KapaBiosystems, cat. KK2801) for 16 cycles and iPCRtagged indexing primers. Subsequently the libraries were purified using AmPure beads. The library concentration was quantified using the Kapa Library quantification kit (KapaBiosystems, KK4844) and equimolar amounts of each indexed library were pooled to run two WGBS-seq libraries on one lane of HiSeq 2000 Illumina system sequencing 100 bp paired-end.

Stranded Total RNA sequencing libraries

Total RNA was extracted using Tri-Reagent (Sigma). 100 ng - 1 μ g of RNA with a RNA integrity number (RIN) of 8-10 were used for total RNA-seq libraries. After Ribozero depletion, stranded total RNAseq libraries were prepared using the TruSeq RNA library preparation kit v2 (Illumina) following manufacturer's instructions. The RNA-seq libraries were indexed and 50 bp single end sequencing was performed using TruSeq reagents (Illumina) according to manufacturer's instructions. Libraries were sequenced on Illumina HiSeq 2000.

Small RNA-seq libraries

Small RNA-seq libraries were produced according to the Illumina protocol (RS-200-0012), with the following changes: 10 ng or 1 μ g RNA with a RIN of 8-10 were used as input material. The instructions were followed until the cDNA purification. In order to purify the cDNA, the samples were run on 10 % Novex PAGE gel. DNA sample buffer (5x Qiagen sample buffer, miniprep kit) was used, and the Illumina custom and high resolution markers were prepared following supplier's instructions. A 10 % agarose gel was run for 1h at 145 volts. The gel was stained for 5 min in SYBR gold and visualised under UV light. Four holes were punched into 200 μ l PCR tubes using the tip of a 21 G needle. The entire area between the 145 and 160 bp markers was excised and placed in the 200 μ l PCR tubes. The 200 μ l PCR tube was placed into a 1.5 ml eppendorf tube and spun for 2 min at 10,000 x g. 400 μ l freshly prepared 0.3 M NaCl was added to the gel samples and the DNA was eluted from the gel by rotation ON at 4° C. The samples were subsequently spun for 5 min at max speed and the supernatant was transferred to a clean tube. In order to eradicate any remaining gel, the supernatant was spun down again for 5 min and a 10 μ l pipette tip was inserted in 200 μ l pipette tip to transfer the supernatant to a new tube. The DNA was precipitated ON by adding 1153 μ l EtOH and 1 μ l glycogen and the next day the DNA was eluted in 10 μ l EB and the library was quantified using the HighSensitivity Bioanalyzer kit. The small RNA-seq libraries were additionally quantified by Kapa Library Quantification. The libraries were pooled according to their molecular weight and high-throughput sequencing of all libraries was carried out with single-end protocols on a HiSeq 2000 instrument (Illumina).

ChIP-seq libraries

ESCs were grown on 15 cm dishes coated with 0.1 % gelatine until they were 80 % confluent. Subsequently cells were were cross-linked with 1 % methanol-free formaldehyde in fresh medium for 10 minutes. To quench the cross-linking, 0.2 M final concentration of glycine was added. ESCs were washed twice with ice cold 1 x PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 2 mM KH₂PO₄ dissolved in 800 ml distilled H₂O, pH was adjusted to 7.4 with HCl) and harvested using a cell scraper. Cells were then pelleted by centrifugation at 8,000 x g at 4 °C for 3 min. Pellets were resuspended in LB1 buffer (50 mM HEPES' KOH, pH 7.5; 140 mM NaCl; 1 mM EDTA; 10 % glycerol; 0.5 % NP-40; 0.25 % Triton X-100, protease inhibitors) for 10 minutes at 4°C, pelleted and resuspended in LB2 buffer (10 mM Tris/HCl, pH 8.0; 200 mM NaCl; 1 mM EDTA; 0.5 mM EGTA, protease inhibitors) for 10 minutes at 4 °C. Cells were pelleted and resuspended in LB3 buffer (10 mM Tris-HCl, pH 8; 100 mM NaCl; 1 mM EDTA; 0.5 mM EGTA; 0.1% Na/Deoxycholate; 0.5% N-Lauroylsarcosine, protease inhibitors). Next the cells were sonicated using Misonix Sonicator 3000. Triton X-100 was added to a final concentration of 1% and the lysate was centrifuged at 20,000 x g for 10 min to pellet the debris. The bead-antibody complexes were prepared before adding the sonicated DNA. Protein G-coupled Dynabeads (Thermo Fisher Scientific, Cat. No. 10003D) and the primary antibodies in PBS with 5 mg/ml BSA were incubated ON. Subsequently, the bead-antibody complexes were added to the sonicated chromatin and both were incubated at 4 °C ON. On the following day, beads were washed extensively with RIPA buffer (50 mM HEPES pH 7.6, 1 mM EDTA, 0.7 % Na deoxycholate, 1 % NP-40, 0.5M LiCl), once with 1x TE buffer (1 M Tris-HCl (pH approximately 8.0), 0.1 M EDTA) and eluted in 200 μ l of buffer containing 1 % SDS and 0.1 M NaHCO₃. They were then incubated at 65 °C ON for reverse cross-linking. RNase A treatment at 37 °C was carried out for 1 h, then Proteinase K treatment at 55 °C for 2 h. The DNA was then extracted with phenol/chloroform, followed by ethanol precipitation. ChIP-seq library preparation was performed using MicroPlex Library Preparation kit (Diagenode) following manufacturer's instructions. Libraries were quantified using the High Sensitivity DNA Bioanalyzer kit and Kapa library quantification. Equal amounts of each indexed library were pooled for multiplex sequencing on a HiSeq 2500 Illumina system (50 bp single-end).

2.3.15 Sequencing

Sequencing was performed at the Wellcome Trust Sanger Institute and the Babraham next generation sequencing facility. Libraries were sequenced on either an Illumina MiSeq or an Illumina HiSeq using the default RTA analysis software.

2.3.16 Bioinformatics

For all experiments involving next-generation sequencing (NGS), raw sequence reads were trimmed to remove poor quality reads and adapter contamination using Trim Galore (v0.4.1). All data was analysed using SeqMonk

(www.bioinformatics.babraham.ac.uk/projects/seqmonk/).

WGBS-seq

After the raw sequences were trimmed, the remaining sequences were mapped using Bismark (v0.14.4) (Krueger and Andrews, 2011) with default parameters to the mouse reference genome NCBIM37 (mm9) using paired-end mapping mode. Following this, the reads were deduplicated and the CpG methylation calls were extracted by running the Bismark methylation extractor (v0.14.4) in paired-end mode. Data were quantitated using SeqMonk

(www.bioinformatics.babraham.ac.uk/projects/seqmonk/). Probes were defined to span 50 CpG sites and have a minimum read count of 4. The global methylation level was used for calculation of the bean plots covering genomic features defined as follows: exons (probes overlapping exons), introns (probes overlapping introns), promoters (probes overlapping 1000 bp upstream of genes), CGI promoters (promoters containing or within 250 bp of a CGI), non-CGI promoters (all other promoters), intergenic (probes not overlapping with gene bodies), LMRs (Stadler et al., 2011), H3K27ac and

H3K4me1 enhancers (Creyghton et al., 2010). The imprinted control regions were used from Tomizawa et al., 2011.

Total RNA-seq

The trimmed reads were mapped to the mouse NCBIM37 (mm9) genome assembly using TopHat v2.0.12 (Trapnell et al., 2009). Strand specific quantification of the data was quantitated overlapping mRNAs using the RNA-seq quantitation pipeline in Seq-Monk software

(www.bioinformatics.babraham.ac.uk/projects/seqmonk/). Differentially expressed (DE) genes were determined using DESeq2 (p-value 0.05, with multiple testing correction) and an intensity difference filter (p-value 0.05, with multiple testing correction) (Love et al., 2014). The DE genes with high confidence were defined as the genes which were called as significant by both statistical tests.

Global pervasive transcription, defined as genic antisense transcription, was calculated as following. The antisense transcription overlapping protein coding genes was quantitated for WT samples as well as for each time points after Dnmt1 KO, using a binomial test with a false discovery rate (FDR) of p < 0.05. Additionally, the raw antisense transcription counts for all samples was calculated and significant differential antisense expression was calculated using DESeq2 with an FDR < 0.05. The overlap of the two quantifications was used to define pervasive transcription. And the difference in antisense transcription between WT and KO samples at each time point was plotted using R.

Gene Ontology

Functional annotation enrichment analyses were performed using Panther webtools (Mi et al., 2016).

Small RNA-seq

For small RNA-seq, trimmed sequencing reads were aligned to mouse genome assembly NCBIM37 (mm9) using Bowtie2 (Langmead and Salzberg, 2012). For miRNA analysis, probes were generated overlapping CDS (coding DNA sequence) and read count enrichment was calculated normalised by total read counts (RPM = reads per million mapped reads) using Seqmonk. Differentially expressed genes were determined using DESeq2 (p-value 0.05, with multiple testing correction) and an intensity difference filter (p-value 0.05, with multiple testing correction) (Love et al., 2014). The high confidence DE genes were defined as the genes which were called as significant by both statistical tests.

For transposon analysis, small RNA-seq data was analysed using the small RNA-seq pipeline piPipes (https://github.com/bowhan/piPipes) (Han et al., 2015). Trimmed data were mapped using Bowtie against the mm9 genome. Additionally, the trimmed reads were mapped to different annotations: piRNA annotations were defined earlier (Li et al., 2013a). Repeats were defined during the analysis using the mouse repeat-masker annotation

(http://www.repeatmasker.org). The plots shown were generated as described below. The distribution of small RNAs was computed by mapping all small RNA-seq reads to the different genomic features. The length distribution was calculated taking all uniquely mapped small RNAs into account. Small RNA reads were then pre-filtered as follows: reads mapping to rRNAs and miRNAs were excluded, then reads aligning to the repeat masked mm9 genome (all annotated repeats were masked) were removed. The resulting small RNA reads were mapped to all repeats. The 5' end nucleotide composition was computed from the uniquely mapped small RNA reads. Similarly, analysis of the position of 5' to 5' overlap was performed on the mapped small RNA reads, and the length distribution and strand orientation of small RNAs shown was generated using uniquely mapped small RNA reads.

For consensus sequence mapping, small RNAs were mapped to TE consensus sequences from repeatmasker using Bowtie (v1.0.1; default parameters).

ChIP-seq

For ChIP-seq, trimmed sequencing reads were aligned to mouse genome assembly NCBIM37 (mm9) using Bowtie2 with default parameters. For calling enrichment, read count quantitation in Seqmonk was used. Probe trend plots and probe alignment plots were generated by calculating average CpG methylation levels of 1 kbp overlapping probes from 5 kbp upstream of the transcriptional start site through gene bodies (which were scaled for visualisation) to 5 kbp downstream of the transcriptional end site.

Repeats Analysis

Repeat locations for all repeat classes of interest were extracted from the pre-masked repeatmasker libraries (mouse, repeatmasker v4.0.3, library version 20130422). Instances of repeats within 2 kb of an annotated gene (defined by Ensembl) were removed to prevent mixing signals from genic expression with expression of repetitive sequences. We excluded all TEs which had less than 1000 calls in each dataset.

Bisulfite-seq libraries were processed and mapped as described above. The methylation levels at the instances of repeats were calculated as the sum of the percentages of methylated Cs over all Cs for all methylated and non-methylated calls of each repeat class. The methylation retention was plotted as scatterplot or line graph using R.

Total RNA-seq sequences were processed and mapped as described above. Non-directional overlaps were quantitated between the mapped RNA-seq reads and the instances of repeats. Subsequently, the counts of all instances of a repeat class was summed and the sum was corrected for the total length of all repeats and the size of the individual libraries to generate RPKM (reads per kilobase per million mapped reads) expression values. The heatmaps were then generated using the R pheatmap library.

For the sense/antisense transcription over repeats, the percentages of sense and antisense transcription over repeats were calculated. Additionally, the bias of sense transcription was calculated as the percentage of antisense transcription subtracted from the sense transcription. Furthermore, the overlap of TEs with genes, as defined by Ensembl, was calculated.

For ChIP-seq analysis, the sequences were processed and mapped as described above. The counts of all instances of repeat classes were summed and the sum was corrected for the total length of all repeats and the size of the individual libraries to generate RPKM expression values. The heatmaps were then generated using the R pheatmap library.

Chapter 3

Acute demethylation upon *Dnmt1* conditional KO

3.1 Introduction

During early mammalian development, DNA methylation and other repressive epigenetic marks are erased in mammalian primordial germ cells (PGCs) and the early embryo. Therefore, this represents a critical period of genome defence as transpososable elements (TEs) can potentially mobilise, presenting a danger to genome integrity. As such during this period alternative silencing pathways are needed to limit the activity of TEs in the absence of DNA methylation (Allis and Jenuwein, 2016).

A recent study has shown that during early developmental epigenetic reprogramming the DNA maintenance methylation machinery is the key driver of *in vivo* DNA demethylation (von Meyenn et al., 2016). A *Dnmt1* KO may therefore serve as a good *in vitro* system to recapitulate *in vivo* demethylation dynamics.

In this chapter, I used *Dnmt1* conditional knockout (KO) mouse ESCs to study the alternative regulatory mechanisms involved in TE regulation during DNA hypomethylation.

In order to characterise the *in vitro* system and to be able to compare it to *in vivo* epigenetic reprogramming, I analysed the effects of global demethylation upon *Dnmt1* on the whole genome using whole genome bisulfite sequencing (WGBS-seq), long and short RNA sequencing as well as chromatin profiling of repressive histone marks (Figure 3.1).

I compared differentially methylated regions (DMRs) which resisted global hypomethylation in our *in vitro* system with variably erased CGIs (VECs) during epigenetic reprogramming (Seisenberger et al., 2012).

Furthermore, as several genes and miRNAs have been described to play an important role during PGC development, I investigated whether the acute deletion of *Dnmt1* resulted into transcriptional changes of these genes.

Earlier research has shown that H3K9me2 becomes lost in PGC development, while H3K27me3 is globally enriched in the genome during the formation of the future germ cells (Hajkova et al., 2008, Seki et al., 2007). Additionally, a remodelling of H3K9me2/3 and H3K27me3 is also found in preimplantation embryos (Santos et al., 2005, Liu et al., 2016, Zheng et al., 2016). Accordingly, I looked at the distribution of repressive histone marks - H3K27me3, H3K9me2 and H3K9me3 - upon global demethylation induced by *Dnmt1* deletion.



Figure 3.1: Schematic of in vitro Dnmt1 deletion time-course in mouse ESCs.

Having characterised my Dnmt1 KO in vitro system I began to study the effects of Dnmt1 KO induced hypomethylation on TE elements. It is especially crucial during early phases of development to protect the genome from highly active TE classes as this could lead to germline mutations which can be passed on to future generations.

DNA methylation is essential for TE suppression, however this epigenetic modification is globally erased during epigenetic reprogramming (Walsh et al., 1998). Further, it has been shown that specific TE classes become upregulated during this global hypomethylation *in vivo* PGC development (Molaro et al., 2014) upon these are the young classes of ERVs - like IAPs and ETns - that still possess the ability to retrotranspose and whose transcriptional activity therefore could result into retrotransposition. Thus, I wanted to investigate whether additional epigenetic modifications play a role in TE transcriptional silencing in the absence of DNA methylation during epigenetic reprogramming.

Earlier research has discovered increased pervasive transcription over TEs in ESCs in comparison to somatic cells (Kelley and Rinn, 2012). It was shown in yeast, that low level of genome-wide pervasive transcription, antisense to genic transcription, can initiate RNAi as a defence mechanism against TEs (Cruz and Houseley, 2014).

The regulatory function of an antisense RNA, sense to a protein-coding RNA, has long been suggested (Derrien et al., 2012). Sense/antisense transcription allows the production of double stranded RNA (dsRNA), which is a target for RNAi. dsRNA viruses are controlled by RNAi in plants, insects, lower eukaryotes as well as mammalian cells (reviewed in Ding, 2010, Li et al., 2013b, Maillard et al., 2013). Furthermore, RNAi can prevent TE mobility in mammals (Yang and Kazazian, 2006, Babiarz et al., 2008). However, in neither of the studies the source of the dsRNA has been identified and therefore it remained ambiguous how the host genome defence pathway identify TE expression to control their activity.

In this study, I wanted to investigate whether pervasive genome-wide transcription, could feed into an RNAi pathway to defend against TE transcriptional activity, similarly to the mechanism that has been found in yeast. Consequently, I tested whether small RNAs were getting enriched at transcriptionally active TE classes upon global hypmethylation by *Dnmt1* KO.

In addition to DNA methylation, and small RNAs, chromatin has been studied as a mechanism to control TE activity. Knockout of the H3K9me3 histone methyltransferase ESET led to transcriptional upregulation of IAP elements in PGC development as well as in ESCs (Karimi et al., 2011, Maksakova et al., 2011). Accordingly, I investigated whether the repressive marks H3K9me3, H3K9me2 and H3K27me3 were enriched at TEs upon global hypomethylation induced by conditional *Dnmt1* KO.

This chapter presents a mechanistic study of TE regulation to unveil the complexity of TE silencing involving DNA methylation, histone modification as well as small RNAs as means of post transcriptional gene silencing (PTGS). Experiments presented in this chapter aim to elucidate the interplay of epigenetic modifications to keep TE classes under control and preserve genome integrity for the next generation.

3.2 Results

3.2.1 Acute deletion of *Dnmt1* as a model for global demethylation dynamics

To follow DNA demethylation dynamics upon acute deletion of Dnmt1 in ESCs, I used WGBS-seq, sampling DNA at different time-points after Dnmt1 KO induction (list of WGBS-seq datasets in appendix). I mapped the WGBS-seq libraries to the mouse genome (mm9) using the Bismark alignment tool (Krueger and Andrews, 2011) and did the following analysis using the Seqmonk interface. I defined 50 CpG windows and used the bisulfite pipeline in Seqmonk to analyse the methylation level of the probes in the different libraries. Confirmed knockout of Dnmt1 ($Dnmt1^{-/-}$) compared to wildtype (WT) ($Dnmt1^{fl/fl}$) ESCs led to global demethylation in all genic and intergenic features (Figure 3.2). Dnmt1 KO induced hypomethylation started at day 3 and was only completed on day 6 (Figure 3.4). In general CGIs were always hypomethylated in WT and Dnmt1 KO ESCs in comparison to the rest of the genome. Furthermore, low methylated regions (LMRs) along with enhancer regions, as defined by occupancy of H3K4me1 and H3K27ac histone modifications, followed whole genome demethylation dynamics (Figure 3.3).

I used the Seqmonk binomial test for 100 CpG regions which showed unusually high



Figure 3.2: Genome demethylation upon *Dnmt1* deletion. CpG methylation levels measured by WGBS of WT (grey) and conditional *Dnmt1* KO ESC induced for 1 day (dark red), 3 days (light red), 6 days (light pink), 9 days (light blue), 11 days (dark blue). Methylated cytosines were counted for each rolling 50 CpG window genome-wide and are expressed as percent of total cytosines per window.



Figure 3.3: Genome demethylation in LMRs and enhancer regions. CpG methylation levels measured by bisulfite sequencing of WT (grey) and conditional *Dnmt1* KO ESC induced for 1 day (dark red), 3 days (light red), 6 days (light pink), 9 days (light blue), 11 days (dark blue). Methylated cytosines were counted for each rolling 50 CpG window genome-wide and are expressed as percent of total cytosines per window. Enhancer regions are defined by H3K4me1 and H3K27ac marks.



3.2. Results

Figure 3.4: WGBS on Chromosome 2 upon *Dnmt1* KO. Methylation level across chromosome 2. Each dot represents one probe which was quantitated by 50 CpG with at least 10 reads. The scale is a hot=red to cold=blue scale, with red representing 100% methylation and blue 0% methylation.

or low methylation in all of the WT to *Dnmt1* KO comparisons. Despite global hypomethylation with similar kinetics in all genomic features, I identified 773 differentially methylated regions (DMRs), which consistently changed after *Dnmt1* KO - 20 DMRs were hypomethylated and 726 DMRs were hypermethylated (List of top 50 DMRs in appendix).

I carried out a gene ontology (GO) analysis of the DMRs resistant to demethylation, using the Panther webtool (Thomas et al., 2006). I found significant enrichment of biological terms such as neuronal development and cell differentiation (Figure 3.5). The FGF signalling pathway also retained more DNA methylation in comparison to the rest of the genome. Nevertheless, I did not find any overlap between resistant regions (VECs) in *in vivo* PGC reprogramming and the *Dnmt1* KO ESC system, other than IAPs, which resisted global demethylation as I will expand on later in this chapter.

GO Term	Cour	nt %	PValue	List Total	Pop Hits	Pop Total	Fold Enrichment
cell motion	23	6.0	1.87E-06	258	367	13588	3.30
axon guidance	12	3.1	2.35E-06	258	98	13588	6.45
neuron differentiation	23	6.0	7.23E-06	258	399	13588	3.04
neuron development	19	5.0	1.13E-05	258	292	13588	3.43
cellular component morphogenesis	21	5.5	1.18E-05	258	351	13588	3.15
cell morphogenesis	19	5.0	2.40E-05	258	309	13588	3.24
axonogenesis	13	3.4	6.28E-05	258	163	13588	4.20
neuron projection development	15	3.9	6.99E-05	258	218	13588	3.62
neuron projection morphogenesis	13	3.4	1.31E-04	258	176	13588	3.89
cell morphogenesis involved in neuron differentiation	13	3.4	1.80E-04	258	182	13588	3.76
cell morphogenesis involved in differentiation	14	3.7	2.00E-04	258	212	13588	3.48
cell projection morphogenesis	13	3.4	4.68E-04	258	202	13588	3.39

Figure 3.5: Regions resisting global demethylation upon *Dnmt1* KO. GO term analysis of all 773 regions which resisted global demethylation from day 6 after *Dnmt1* KO onwards. Only 258 genes were found with a mapped GO term.

To permit a comparison between gene expression and global demethylation, I carried out total RNA-seq analysis of the same time-points after *Dnmt1* KO, as I did for WGBS-seq (list of total RNA-seq datasets in appendix). Total RNA-seq was performed to allow for transcriptional analysis of all TE classes, also those classes that lack a poly adenylation (poly-A) tail.

3.2.2 Dnmt1 KO leads to transcriptional upregulation of imprinted loci

I did a pairwise comparison of genes upregulated between Dnmt1 KO and WT ESCs at each time-point and checked for differentially expressed genes. For this, I used the Seqmonk intensity difference filter with Benjamini and Hochberg correction for multiple testing with a p-value threshold of < 0.05 and overlapped them with the genes called differentially expressed by DESeq2 with a p-value threshold of < 0.05 and multiple testing correction (Love et al., 2014). Apart from Dnmt1 itself, none of the other genes showed any real change at the first time-point after KO induction (Day 1: 3 differentially expressed (DE) genes). However, an increased regulation with time thereafter (Day 3: 43 DE genes, Day 6: 357 DE genes, Day 9: 269 DE genes, Day 11: 50 DE genes) was found (List of differentially expressed genes upon Dnmt1 KO in appendix). Furthermore, I called significantly differentially expressed genes between WT and Dnmt1 KO from day 3 onwards. I identified 85 differentially expressed genes with a p-value of < 0.05. The genes were consistently up or down regulated in response to Dnmt1 KO.

There was a large bias for genes to be upregulated in response to the Dnmt1 induced DNA demethylation (n = 80), but there were also a small number of genes (n = 5) which were downregulated in response to global hypomethylation (Figure 3.6).

Consequent to Dnmt1 deletion, Dnmt1, Igfbp2 and Grb10 were downregulated, whereas the imprinted genes Xlr3a, Mirg and Rian were upregulated (Figure 3.7). Furthermore, genes important for PGC development - Dazl, Lefty2, Eif2z3y, Zscan4d/f - were upregulated after global DNA hypomethylation.

Figure 3.6: Genes respondent to Dnmt1 driven demethylation. Scatter plot of total RNA-seq data mapped to the whole genome at day 9 after Dnmt1 KO compared with WT ESCs. Genes significantly differentially expressed from day 6 after Dnmt1 deletion onwards (black), genes which did not change expression levels significantly (grey) across the whole time-course.





Figure 3.7: Highest respondent genes upon *Dnmt1* deletion. Bargraph of RNAseq of differentially expressed genes between WT (grey) and conditional Dnmt1 KO ESCs induced for 0 days (black), 1 day (dark red), 3 days (light red), 6 days (light pink), 9 days (light blue), 11 days (dark blue). Data shows RNA-seq data of 2 biological replicates of each time-point.

The pluripotency network *Oct4*, *Nanog*, *Klf4*, *Esrrb* and *Sox2* - was normally expressed after knockout of the DNA maintenance methylase (Figure 3.8).



Figure 3.8: The pluripotency network is expressed normally upon *Dnmt1* KO. Bargraph of RNA-seq data of pluripotency genes of WT (grey) and conditional *Dnmt1* KO ESCs induced for 0 days (black), 1 day (dark red), 3 days (light red), 6 days (light pink), 9 days (light blue). The expression levels of the two RNA-seq libraries are shown next to each other.



Figure 3.9: Expression level does not globally recapitulate methylation status. Line graph of DNA methylation level of genes which showed different expression level in WT ESCs: not expressed (RPM < 0, red), low expressed (2.5 < RPM > 0, violet), expressed (5 < RPM > 2.5, pink), high expressed (10 < RPM > 5, dark blue) or very high expressed (RPM > 10, light blue) as well as genes which were Dnmt1 dependent (orange) across the time-course after Dnmt1 deletion.

I next investigated the correlation between gene transcription and promoter methylation throughout the time-course of global DNA hypomethylation. To do this, I grouped the genes in the WT ESCs into not expressed (RPM < 0), low expressed (2.5 < RPM > 0), expressed (5 < RPM > 2.5), high expressed (10 < RPM > 5) or very high expressed (RPM > 10) and then looked at their methylation level throughout the time-course after *Dnmt1* deletion. Genes highly expressed in WT ESCs (RPM > 10) had much lower initial promoter methylation than other genes, in spite of this the expression of all other genes was independent of their promoter methylation level. Additionally, I compared the DNA methylation level of genes which were *Dnmt1* dependent, but they also demethylated with the same kinetics as the rest of the genome (Figure 3.9).

3.2.3 Small RNAs from *Dlk* and X-chromosome locus *Xlr3* become upregulated upon *Dnmt1* KO

In order to analyse whether *Dnmt1* deletion had any effect on the expression of small RNAs, I carried out small RNA-seq at each time-point after *Dnmt1* KO and compared it to the respective WT ESCs (list of small RNA-seq datasets in appendix). I mapped the small RNA-seq libraries to the whole genome, and by using the Seqmonk small RNA-seq quality control plot found that over 90% of reads mapped to miRNAs. The small RNAs were 20-24nt long (Figure 3.10A) and endogenously expressed miRNAs in ESCs were expressed throughout the whole time-course after *Dnmt1* deletion (Figure 3.11). Moreover, reads for both the 3'arm as well as the 5'arm of the mature miRNA mapped to known endogenously transcribed miRNAs in ESCs (Figure 3.10B).



Figure 3.10: Small RNA-seq libraries are 90% made up of miRNAs.(A) small RNA size distribution as well as classification of different small RNA classes in *Dnmt1* KO and WT ESCs, miRNAs (grey), rRNA (green), small nuclear RNAs (snRNAs) (violet), miscellaneous other RNAs (misc RNAs) (red), small nucleolar RNAs (snoRNA) (orange) and tRNA (light blue), (B) genic location of miRNA 200c with reads mapped in *Dnmt1* KO and WT ESCs, each line representing one read.

Next, I wanted to test whether any of the miRNAs were differentially expressed upon global hypomethylation. To address this question, I carried out pairwise comparison


Figure 3.11: Not significantly altered expression of endogenous miRNAs in ESCs. small RNA-seq of miRNA expression in WT (greys) and conditional *Dnmt1* KO ESC induced for 1 day (dark red), 3 days (light red), 6 days (light pink), 9 days (light blue), 11 days (dark blue). Data shows small RNA-seq data of 3 biological replicates of each time-point.

of Dnmt1 KO and WT samples of each time-point and looked at the differentially expressed miRNAs. To call differentially expressed miRNAs, I overlapped the differentially expressed miRNAs using DESeq2 and Seqmonk intensity difference filter with a p-value of < 0.05. Two distinct miRNA loci were upregulated in Dnmt1 KO in comparison to WT ESCs (Figure 3.12). The first locus was a miRNA cluster of the imprinted Dlk locus (Edwards et al., 2008, Charalambous et al., 2004). This locus has been previously shown to drive the expression of noncoding RNAs in ESCs (Labialle et al., 2014 Cavaillé et al., 2002 Seitz et al., 2004). The second locus was part of the imprinted Xlr3 locus on the X Chromosome (list of differentially expressed miRNAs upon Dnmt1 KO in appendix).

In order to verify the small RNA-seq results, I carried out small RNA quantitative real time PCR (qRTPCR) of some of the differentially expressed miRNAs confirming that miRNAs from the *Dlk* locus, mmu-miR-367 and mmu-miR-543, were significantly upregulated in *Dnmt1* KO ESCs in comparison to WT ESCs (Figure 3.13).



Figure 3.12: miRNAs from the *Dlk* and *Xlr3* locus are dependent on *Dnmt1*. (A) Scatter plot of total RNA-seq data mapped to the whole genome of day 9 after *Dnmt1* deletion (y-axis) compared to wildtype (x-axis). miRNAs significantly upregulated upon *Dnmt1* KO from the *Dlk* locus (black), miRNAs significantly upregulated upon *Dnmt1* KO from the *Xlr3* locus (green). (B) Bargraph of small RNA-seq of differentially expressed miRNAs - of the *Dlk* and *Xlr3* locus - between WT (grey) and conditional Dnmt1 KO ESCs induced for 1 day (dark red), 3 days (light red), 6 days (light pink), 9 days (light blue), 11 days (dark blue). Data shows small RNA-seq data of 3 biological replicates of each time-point. Differences between conditions that are statistically significant are denoted by (* p-value<0.05, ** p-value<0.005, (Student's t-test))

Figure 3.13: Upregulation of mmu-miR-367 and mmu-miR-543 upon **Dnmt1** KO. (A) small RNA qRTPCR on mature mmu-miR-367 and mmu-miR-543 at day 9 after Dnmt1 deletion (light blue) and wildtype (grey). Each qRTPCR was done in 3 technical replicates. Differences between conditions that are statistically denoted significant are by (* p-value < 0.05,** p-value<0.005, (Student's t-test))





Figure 3.14: Expression of *Dlk* locus upon *Dnmt1* KO. Chromosome view of WGBS-seq, total RNA-seq and small RNA-seq depicted as wiggle plot overlapping imprinted control regions (ICR), mRNA and small RNAs in WT and at day 9 after *Dnmt1* deletion.

In summary, WGBS-seq together with long and short RNA-seq analysis upon Dnmt1 KO allowed me to identify imprinted genes as the highest respondent genes. As an example, I analysed the Dlk locus in more detail and in the WGBS-seq the imprinted control region (ICR) became demethylated upon Dnmt1 deletion. This led to transcriptional upregulation of genes of the Dlk locus as well as upregulation of small RNAs in this imprinted region (Figure 3.14).

3.2.4 No genome-wide chromatin changes upon global demethylation in ESCs

I wanted to know whether any global chromatin rearrangements were taking place subsequently to induction of global DNA demethylation *in vitro* by *Dnmt1* KO. During global DNA demethylation in *in vivo* PGC development H3K9me2 becomes depleted while H3K27me3 becomes enriched genome-wide (Hajkova et al., 2008, Seki et al., 2007). Furthermore, a remodelling of H3K9me2/3 and H3K27me3 is also found



3.2. Results

Figure 3.15: Enrichment of repressive histone marks in the genome. Pie chart of enrichment of H3K27me3, H3K9me3 and H3K9me2 in repeats (dark violet), genic regions (light violet), promoters (dark green), CGIs (middle green), intergenic regions (light green) in wildtype ESCs.

in preimplantation embryos (Santos et al., 2005, Liu et al., 2016, Zheng et al., 2016). To study the connection between global DNA demethylation and chromatin organisation in ESCs, I performed ChIP-seq of three repressive histone marks - H3K9me3, H3K9me2 and H3K27me3.

I prepared ChIP-seq libraries from ESCs at day 4 and day 8 after *Dnmt1* KO induction, as well as the respective WT samples (list of ChIP-seq datasets in appendix). Those time-points were chosen because at day 3 post *Dnmt1* deletion global hypomethylation was not yet complete, while at day 8 the whole genome was totally hypomethylated. Simultaneously, the transcriptional changes were mostly seen at day 6 after *Dnmt1* deletion. To permit ChIP-seq analysis, I carried out a read count quantitation in Seqmonk and normalised to total read counts.

H3K27me3 was more highly abundant in the genome (n=50599 enriched sites) than H3K9me3 (n=30184) and H3K9me2 (n=2541). Additionally, H3K27me3 was mostly enriched in genic regions, CGIs, promoter sites as well as repetitive elements, whereas H3K9me3 was depleted from CGIs, intergenic regions as well as promoters and most of the enrichment of this histone mark could be found in repeats. H3K9me2 was evenly distributed across the whole genome with a slight enrichment at genes and repeats



Figure 3.16: Histone marks over gene body and TSS. Probe enrichment of H3K9me3 (green), H3K9me2 (yellow) and H3K27me3 (blue) over gene body and TSS.



Figure 3.17: Histone mark enrichment in the genome. Chromosome view of ChIP enrichment of H3K9me3 (green), H3K27me3 (blue) and H3K9me2 (yellow) over a 500kbp region in Chromosome 12. Intensity of the enrichment on the y-axis.

(Figure 3.15).

H3K27me3 was specifically enriched at the transcription start site (TSS) with a depletion at the gene body (Figure 3.16), while H3K9me2/3 were depleted in genic regions. The same trend could be seen in a specific example of a 500 kilobase (kb) region on chromosome 2: H3K9me3 was specifically enriched in the active class IAPs (IAPEZ), while H3K27me3 was highly enriched at TSSs and H3K9me2 was ubiquitously found across the whole genome (Figure 3.17).

Subsequently, I tested whether any of the histone marks changed enrichment at any regions in the genome upon *Dnmt1* deletion. However, only minor changes in H3K9me3, H3K9me2 and H3K27me3 peaks was measured globally upon DNA demethylation induced by *Dnmt1* deletion (Figure 3.18).



Figure 3.18: Minor redistribution of global histone marks upon *Dnmt1 KO*. Scatter plot of repressive histone marks overlapping genes in wildtype (y-axis) versus *Dnmt1* KO (x-axis) ESCs.

To investigate the correlation between histone marks and gene expression in ESCs, I compared the expression levels of genes in WT ESCs and overlaid them with their histone enrichment across the time-course after *Dnmt1* deletion. In WT ESCs H3K27me3 enrichment was higher over TSS of lowly expressed genes than highly expressed genes (Figure 3.19) in keeping with its repressive role (Marks et al., 2012). However, neither of the histone marks changed enrichment over genes throughout the time-course of global hypomethylation, nor did any of the repressive marks change enrichment at genes whose expression changed upon global hypomethylation (Figure 3.20).

Finally, I wanted to study the correlation between enrichment of repressive histone



Figure 3.19: H3K27me3 strong enrichment at TSSs of lowly expressed genes. Aligned probe plot of H3K27me3 enrichment surrounding 5kb of TSS.



Figure 3.20: No change of histone marks in expressed genes across Dnmt1 KO. H3K9me3, H3K27me3 and H3K9me2 read enrichment over genes which showed different expression level in WT ESCs: not expressed (RPM < 0, red), low expressed (2.5 < RPM > 0, violet), expressed (5 < RPM > 2.5, pink), high expressed (10 < RPM > 5, dark blue) or very high expressed (RPM > 10, light blue) as well as genes which were Dnmt1 dependent (orange) across the time-course after Dnmt1 deletion.



Figure 3.21: Histone marks are not enriched on regions dependent on DNA maintenance methylation. Line graph of H3K9me3, H3K27me3 and H3K9me2 read enrichment over very high (dark red) (100-75%), high (orange) (75-50%), medium (light blue)(25-50%) and non methylated (blue)(0-25%) promoter regions before and after *Dnmt1* deletion.

marks and global methylation state of promoters after Dnmt1 deletion.

3.2. Results

To approach this, I categorised the promoter methylation levels at day 11 after *Dnmt1* deletion into four categories: (1) hypermethylated (75-100% methylation), (2) methylated (75-50% methylation), (3) low methylated (25-50% methylation) and (4) hypomethylated (0-25% methylation).

This allowed me to compare the histone enrichment specifically at promoters which resisted global demethylation. But despite the fact that some promoters resisted global demethylation, and still had 75% of methylation at day 11 after Dnmt1 deletion, neither of the repressive histone marks was enriched at those promoters in comparison to the low level methylated promoters (0-25% methylated) (Figure 3.21).

In our *in vitro* system of *Dnmt1* conditional KO I was able to induce global demethylation and found mostly imprinted loci that became transcriptionally activated, while the pluripotency network continued to be expressed normally. Therefore, the conditional KO ESCs proved to be a good model to study the regulation of TEs during global demethylation dynamics *in vitro*.

3.2.5 Mapping of TEs in next generation sequencing libraries

In order to analyse DNA methylation, transcription and histone enrichment across TEs, with the help of Simon Andrews we mapped all reads to the mouse genome allowing only unique mapping of reads. Subsequently, we ignored all reads that were within 2kb of genes. TEs are frequently integrated into genes or lie within close proximity of genes. This leads to wrongly annotated 3'UTRs and therefore reads which are derived from genic expression would be wrongly ascribed to TE transcription (Figure 3.22). These reads would also be seen in consensus sequence mapping. Due to this conservative approach to mapping TEs we are likely to lose information on recently integrated TEs, as we only allow unique mapping. We will probably underestimate TE methylation and transcription levels but on the other hand we can be confident about the TE transcription to be real. For small RNAs, we did consensus sequence mapping to TE classes, as not many genic reads are expected in small RNA-seq libraries.



Figure 3.22: Mapping of TEs in WGBS-seq, RNA-seq and ChIP-seq libraries. Sequencing reads overlapping genes and transposons sitting within 2kb of genes.

3.2.6 TE classes become demethylated upon Dnmt1 KO in ESCs

The *Dnmt1* conditional KO system allowed me to study DNA methylation, small RNA contribution as well as chromatin changes as potential regulators of TE transcriptional activity. First I analysed the demethylation of TEs upon *Dnmt1* KO.

Acute *Dnmt1* KO led to hypomethylation of all transposon classes (Figure 3.23). Nevertheless, some TEs had altered demethylation kinetics compared to the rest of the genome (Figure 3.24).

MMERVK10C always remained more methylated than the rest of the genome and IAP



Figure 3.23: TE family demethylation upon *Dnmt1* **deletion.** CpG methylation levels measured by WGBS of wild type (grey) and conditional *Dnmt1* knockout ESC induced for 1 day (dark red), 3 days (light red), 6 days (light pink), 9 days (light blue), 11 days (dark blue). Methylated cytosines were counted for each rolling 50 CpG window genome-wide and are expressed as percent of total cytosines per window.

elements resisted demethylation until day 6 and then followed genome demethylation kinetics (Figure 3.24). In regard to IAP elements this recapitulates *in vivo* demethylation dynamics, as IAPs are also resistant to global demethylation during primordial germ cell development (Seisenberger et al., 2012).

3.2.7 Dnmt1 KO leads to transcriptional activation of ERV elements

To analyse whether any classes of TEs were dependent on DNA maintenance methylation for their transcription, I mapped the total RNA-seq data uniquely to the genome and checked for transcriptional activation of TEs. Although all TEs became demethylated only a small number of TE classes were transcriptionally upregulated (Figure 3.25). I grouped the TE elements into 3 categories: (1) members of the "white" category were constantly expressed, (2) members of the "black" category increased in transcription upon *Dnmt1* KO and (3) members of the "grey" category were transcriptionally upregulated and subsequently resilenced (in the absence of DNA methylation). In summary, regardless of global TE hypomethylation, with the aforementioned exceptions,



Figure 3.24: Resistance to demethylation of some TEs. (left) Scatter plot of DNA demethylation at day 11 after *Dnmt1* deletion with red dot indicated TE family methylation level, (right) genome background model of demethylation dynamics set to zero (grey line) and demethylation dynamics of TE classes (black line) above.

only a very specific class of endogenous retroviruses (ERVs) became transcriptionally upregulated upon Dnmt1 KO (white category not shown) (Figure 3.25).



Figure 3.25: TE classes upregulated upon *Dnmt1* **KO.** Heatmap of RNA-seq data mapped to TE classes. Only TE classes which were intergenic and within 2kb of genes were considered. Mean of TE classes with at least 1000 integration sites were regarded. RNA-seq was done in duplicates at each time-point with day 0 being the compiled mean of the control RNA-seq datasets.



Figure 3.26: Unique TE elements upregulated upon Dnmt1 KO. RNA-seq data of day 0, day 6 and day 9 upon Dnmt1 KO. Average transcription is shown as read line. Differences between conditions that are statistically significant are denoted by (* p-value<0.05, ** p-value<0.005, (Student's t-test))

LINE and SINE elements were constantly expressed irrespective of global hypomethylation and fell into the "white" category. MMERVK10C increased in expression upon DNA hypomethylation and made up the "black" category. Members of all 3 families of ERVs were expressed and resilenced upon *Dnmt1* KO, and were therefore members of the "grey" category: ERV1 elements - ORR1B1, RLTR1B, RLTR9E, RLTR45, ERV2 elements - IAP-d, IAPLTR3, IAPEZ, ETnERV2, MMERVK9C and ERV3 elements - MERVL.

I examined single intragenic TE insertions of each category - 1: L1MdGf; 2: MMERVK10C; and 3: IAPs and MERVL; and found that their transcriptional profiles recapitulated the mean profiles of the respective classes, which confirmed that transcriptional upregulation was not driven by a minority of elements but was rather a class wide response Figure 3.26).

3.2.8 Sense/antisense transcription of repeat families feed into the RNAi pathway

Earlier studies in yeast have shown that low level of genome-wide pervasive transcription, antisense to genic transcription, can initiate RNAi as a defence mechanism against TEs (Cruz and Houseley, 2014).

With the help of Simon Andrews I performed an analysis of antisense transcription in the genome. In general, there was increased pervasive transcription - defined as genic antisense transcription - upon global hypomethylation induced by Dnmt1 deletion (Figure 3.27). Upon further investigation, it transpired that the pervasive genic transcripts that we detected were in fact produced by transcription of TEs that had integrated in antisense direction to these genes. Indeed, 94.9 % of the genic antisense transcripts overlapped with TEs which were upregulated upon Dnmt1 KO (Figure 3.28). A genome-wide study of genic integration of TE elements confirmed this trend, as TEs were generally depleted in genes; however when overlapping they were inserted in the antisense direction to genes (Figure 3.29). To investigate this further, we analysed my total RNA-seq data to determine whether we could observe both sense and antisense transcription at sites of TE integration. Figure 3.27: Increased genomewide pervasive transcription upon *Dnmt1* KO. Quantification of genic antisense transcription increases upon *Dnmt1* KO. The difference in *Dnmt1* KO versus WT ESCs was calculated at each time point after induced *Dnmt1* KO.





Figure 3.28: Pervasive transcription overlaps with TEs. Chromosome view of genic transcription with an integrated TE insertion. Sense reads (red), antisense reads (blue).



Figure 3.29: TE orientation antisense to genes. Quantifying TE insertion sense (red) or antisense (blue) to genes as well as percentage of TEs overlapping genes. The distribution and percentage of all repeats served as a control.



Figure 3.30: Sense/antisense transcription in TE families. Barplots of sense/antisense strands of RNA-seq data mapped to TE families in conditional *Dnmt1* knock-out ESC induced for 0 days (black), 1 day (dark red), 3 days (light red), 6 days (light pink), 9 days (light blue), 11 days (dark blue), 17 days (dark green) and 25 days (light green). Sense (filled), Antisense (hatched).

To do this we mapped both sense and antisense reads to the different TE classes and showed that TE antisense transcription was found in all of the TE classes, while only the sense transcripts of members of the "grey" TE category were upregulated (Figure 3.30).

As discussed previously, TE antisense transcription is a result of sense transcription of the genes in which the TEs are integrated. As such we were interested in whether genic sense transcription, could actually be serving as a trap for TE transcriptional activity, in order to control expression of TEs. This system may function as follows: the production of sense and antisense transcripts over TEs leads to the production of dsRNAs, which subsequently feed into the RNAi pathway thus silencing TEs post transcriptionally (Figure 3.32). To test this hypothesis, I performed small RNA-seq at time-points after Dnmt1 deletion.



Figure 3.31: Model of TE regulation by RNAi. Sense and antisense transcription of TE elements with genic sense transcription working as a trap of TE sense transcriptional upregulation. This leads to the production of dsRNA and allows for the production of small RNAs which can in turn regulate transcription of TEs.

3.2.9 Small RNAs are produced from TEs upon loss of Dnmt1

To evaluate whether small RNAs played a role in resilencing of the TE elements in our system, I carried out small RNA-seq at different time-points after *Dnmt1* deletion. I mapped the small RNA-seq libraries to the consensus sequences of TE classes defined by repeatmasker (www.repeatmasker.org/). A significant increase of small RNAs that mapped to IAPLTR1a, MERVL, MMERVK10C and ETn elements upon *Dnmt1* KO was detected. No increased amounts of small RNAs were detected mapping to any of the other classes of TEs, such as L1MdGf and RLTR45, upon KO of the DNA maintenance methylase *Dnmt1* (Figure 3.32).

In order to better characterise the nature of the small RNAs which mapped to TEs, I concentrated on small RNAs which mapped to the IAP consensus sequence. The small RNAs mapped across the full length of the IAP transcript. There was no specific small RNA enrichment at the 3'UTR, a characteristic for miRNAs (Figure 3.33). Consequently, I wanted to next check whether or not the small RNAs were *bona fide* players of the RNAi pathway.



Figure 3.32: Small RNAs map to TEs. Bargraph of Small RNA-seq libraries in wild type (grey) and conditional Dnmt1 knockout ESC induced for 1 day (dark red), 3 days (light red), 6 days (light pink), 9 days (light blue), 11 days (dark blue), 25 days (light green). Small RNA-seq libraries were done in 3 biological replicates for each time-point. Each qRTPCR was done in 3 technical replicates. Differences between conditions that are statistically significant are denoted by (* p-value<0.05, ** p-value<0.005, (Student's t-test))



Figure 3.33: Small RNAs map to the IAP consensus sequence. Chromosome view of small RNA-seq library reads of *Dnmt1* KO and WT ESCs at each time-point mapped to the IAPEZ consensus sequence. Each line represents one read.

3.2.10 Small RNAs are actively loaded into the RNAi machinery

An essential component of the RNA induced silencing complex (RISC) is one of the four ARGONAUTE proteins (AGO). While all AGOs can bind small RNAs, only AGO2 has the ability to complete endonucleolytic slicing of the target mRNA, which is characteristic of endosiRNAs (Meister et al., 2004, Liu et al., 2004, Song et al., 2004, Okamura et al., 2004, Rand et al., 2005, Matranga et al., 2005).

In order to test whether the small RNAs mapping to IAPs are *bona fide* siRNAs, I carried out an AGO2 immunoprecipitation (IP) and analysed the small RNA content of the pulldown by small RNA-seq (Figure 3.35A).

I mapped the small RNAs to the whole genome as well as repeat classes using the piPipes small RNA-seq pipeline (Han et al., 2015).



Figure 3.34: AGO2 small RNA-seq libraries. (A) Pie chart of AGO2 libraries mapped to genome and repeatome. miRNAs (black), 5'UTR bound small RNAs (light green), 3'UTR bound smallRNAs (yellow), intron bound small RNAs (dark blue), repeats bound small RNAs (dark green), rRNA and tRNA (grey), unannotated small RNAs (white) .(B) Bar graph of size distribution of sense and antisense reads mapped to all repeats.

The AGO2 pulldown was successful, as 90% of small RNAs mapped to the whole genome were miRNAs for both KO and WT sample (data not shown). In the *Dnmt1* KO ESCs at day 9 after induction, 40% of the remaining small RNAs mapped to repeats (Figure 3.34A, WT data not shown).

The small RNAs that mapped to repeats were 22 nucleotides (nt) long and mapped to



Figure 3.35: AGO2 siRNAs map to TE classes. (A) Graphical depiction of AGO2 pulldown of small RNAs. Ago2 antibody is used to pulldown AGO2 protein with small RNAs bound. These small RNAs are released subsequently and subjected to small RNA-seq, (B) Bargraph of small RNA-seq libraries mapped to TE classes. AGO2 pulldown were done in 4 biological replicates at day 9 after Dnmt1 KO. Differences between conditions that are statistically significant are denoted by (* p-value<0.05, ** p-value<0.005, (Student's t-test))

both sense and antisense strands of all repeat classes (Figure 3.34B).

I found small RNAs that mapped to all TEs in WT ESCs, but they significantly increased upon *Dnmt1* KO in IAPEZ, ETn, MMERVK10C and L1MdGf (Figure 3.35B). The small RNAs mapping to TEs had the specific characteristics of endosiRNAs, as they were 22 nt long, had a 5'T overhang at nucleotide 20 and formed 5'-5' overlaps to each other at nucleotide 20 (Figure 3.36). Small RNAs mapped sense and antisense to all TE elements (Figure 3.37). There was no strand bias for small RNAs mapping to LINE-1 elements, while for IAPs and ETn elements there was a sense strand bias.

3.2.11 EndosiRNAs and not miRNAs play a critical role in resilencing of IAPs

To address the question whether endosiRNAs or miRNAs were playing the primary role in IAP resilencing I did siRNA knock down (KD) of different components of the RNAi pathway in *Dnmt1* KO ESCs. DGCR8 works upstream of DICER in the miRNA pathway, but not of the siRNA pathway, and siRNAs are exclusively bound by endonucleolytically active AGO2. I did a KD of *Ago2*, *Dicer* as well as *Dgcr8* in the *Dnmt1*



Figure 3.36: Analysis of AGO2 bound small RNAs that mapped to repeats. (A) Bargraph of 5'5' overhang of small RNAs mapped to repeats, (B) Bargraph of nucleotide contribution along the stretch of the small RNAs mapped to repeats.



Figure 3.37: Sense/antisense small RNAs map to TE families. Bargraph of sense (blue) and antisense (red) small RNA reads mapping to TE families at day 9 after Dnmt1 deletion with WT in transparent. Mapping of unique reads to repeatmasker repeat elements using piPipes small RNAseq pipeline.

KO at day 12 after KO induction and checked for IAP expression to identify whether the miRNA or endosiRNA pathway played the primary role in IAP silencing.



Figure 3.38: IAP is upregulated upon Dicer and Ago2 KD. (A) Graphical depiction of knockdown study, IAP expression (red), small RNA expression (grey). (B) qRTPCR of IAP expression after siRNA knockdown (KD) of *Dicer*, *Ago2*, *Dgcr8* in comparison to non-targeting control. Every siRNA KD was done in 3 technical replicates. Differences between conditions that are statistically significant are denoted by (* p-value<0.05, ** p-value<0.005, (Student's t-test))



Figure 3.39: siRNA knockdown of RNAi pathway. qRTPCR of *Dicer*, Ago2 and Dgcr8 with (+) or without (-) siRNA. Every siRNA KD was done in 3 technical replicates. Differences between conditions are statistically significant are denoted (* p-value<0.05 (Student's t-test)).

I performed the RNAi experiment at day 9 after Dnmt1 KO to allow for transcriptional upregulation of IAPs. The expression of Dgcr8, Ago2 and Dicer was downregulated to 20% of their normal expression level by siRNAs (Figure 3.39). I measured the

expression of IAPs by qRTPCR at day 12 after *Dnmt1* KO (Figure 3.38A). RNAi against *Dgcr8* and the scrambled non-targeting control siRNA resulted in a resilencing of the IAPs, whereas in the Dicer and Ago2 KDs resilencing of IAPs was impaired (Figure 3.38 B).

I therefore propose that endosiRNAs are primarily involved in IAP resilencing.

3.2.12 Histone modifications may account for different behaviour of TE families upon acute *Dnmt1* deletion

Figure 3.40: H3K9me3 enrichment in IAPEZ elements. Wiggle plot over IAPEZ elements in *Dnmt1* KO at day 4 (dark red) and day 8 (dark blue) in comparison to WT ESCs (grey). IAP gene body in black and LTRs in grey. Intensity of enrichment on y-axis.



In ESCs IAPs are tightly guarded by H3K9me3 methylation (Mikkelsen et al., 2007). The KO of the histone methyltransferase ESET in PGCs leads to a loss of H3K9me3 on ERV elements - specifically IAPs (Leung et al., 2014). H3K27me3 has been shown to be highly enriched in MERVL elements (Hayashi et al., 2016). Global demethylation by transition from serum to 2i cultured ESCs has shown that repressive histone marks were redistributed to potentially silence TEs (von Meyenn et al., 2016, Walter et al., 2016).

In order to check whether this redistribution of repressive histone marks also played a role in resilencing TE classes in our system, I carried out ChIP-seq of H3K9me3, H3K9me2 and H3K27me3 before and after *Dnmt1* deletion (Day 4 and Day 8). H3K9me3 peaks were constantly found over IAPs, and H3K27me3 was enriched at MERVL elements at day 8 after *Dnmt1* deletion in comparison to day 4 (Figure 3.40, 3.41).

Redistribution of repressive histone marks H3K9me2/3 and H3K27me3 could indeed



Figure 3.41: H3K27me3 and H3K9me3 enrichment over MERVL elements. Scatter plot of H3K27me3 and H3K9me3 enrichment over all MERVL elements in the genome at day 4 (dark red) and day 8 (dark blue) *Dnmt1* KO ESCs.

partially explain the resilencing of certain TE classes, as follows. The TEs could be subdivided into the ones which acquired H3K27me3 or H3K9me2, while some had enrichment of both marks after day 8 of *Dnmt1* deletion. MMERVK10C of the "black" category showed high enrichment of all three histone marks, which could explain the low expression level throughout the course of Dnmt1 deletion, as these ERVs are known to be highly dependent on H3K9me3 (Maksakova et al., 2011). TE elements of the "grey" category showed varied enrichment of repressive histone marks, as follows. MERVL showed high enrichment of H3K27me3 but depletion of H3K9me2 upon Dnmt1 KO, which could explain the transcriptional upregulation and resilencing of those ERV elements upon global demethylation (Figure 3.42). H3K27me3 enrichment but loss of H3K9me3 occurred at IAPLTR3 and IAP-d elements. H3K9me3 and H3K27me3 were lost over ORR1B and RLTR14 and those elements were expressed until day 11 after Dnmt1 deletion, when potentially other mechanisms are being put in place for long term silencing. H3K9me2 switched to H3K9me3 over RLTR1B upon Dnmt1 KO, while H3K9me2 became depleted and H3K9me3 became enriched at RLTR45, RLTR9E, IAPEZ and MMERVK9C upon Dnmt1 KO. Differing repressive histone marks also controlled TEs of the "white" category. While LTRSI2 and RMER16 did show increased enrichment of H3K9me2 and me3 upon Dnmt1 KO, L1MdGf and ERVB4 showed high enrichment of H3K27me3, RLTR14 lost H3K27me3 and H3K9me2 but gained H3K9me3 modifications.



Figure 3.42: Repressive histone marks control TEs upon Dnmt1 KO. Heatmap of ChipSeq data of H3K9me3, H3K9me2 and H3K27me3 mapped to TE families at day 4 and day 8 after Dnmt1 deletion. Read count enrichment. Each library was done with no replicate. Enrichment of histone marks was calculated relative to WT enrichment.

Altogether, the repressive histone marks I analysed after Dnmt1 deletion may explain the repression of some TE classes after transcriptional upregulation. But also TEs, which were already targeted by small RNAs, got enriched by repressive histone marks after Dnmt1 deletion.

In summary, I found demethylation of nearly all TE classes after Dnmt1 deletion. Subsequently, I classified TE elements dependent on their expression following Dnmt1 deletion.

I found that small RNAs of the endosiRNA class mapped to certain TE classes, irrespective of their transcriptional upregulation after *Dnmt1* KO. Additionally, I found repressive histone marks being targeted to specific demethylated TE classes with the potential to silence them, some with no major contribution of endosiRNAs, whilst others had abundant amounts of endosiRNAs. Hence it is possible that the small RNA silencing pathway is independent of histone modification pathways in the control of TEs during global demethylation.

3.3 Discussion

In this chapter, I first presented and evaluated the *Dnmt1* KO ESCs as a mechanistic system appropriate to dissect *in vivo* demethylation dynamics. I used this hypothesis in an *in vitro* system by acute deletion of *Dnmt1* in mouse ESCs, carried out genome-wide methylome and transcriptome analysis at specific time-points after deletion. Furthermore, I studied the dynamics of repressive chromatin changes during the induced global hypomethylation.

The analysis of genomic demethylation upon *Dnmt1* deletion by WGBS-seq showed that most genomic elements became demethylated at a similar rate, nevertheless I was able to identify DMRs that were relatively resistant to demethylation and which overlapped loci implicated in FGF signalling as well as neuronal development. The resistance to hypomethylation of genes important for neuronal development could indicate that these genes need to be kept silenced in order to prevent transcriptional activity and hence differentiation of the ESCs into neuronal lineages.

I found that a minority of transcripts became upregulated upon *Dnmt1* deletion; the genes that were consistently upregulated from day 6 onwards were mostly imprinted loci. Reassuringly, genes important for PGC development also became upregulated.

Interestingly, I could only confirm a general link between transcription and methylation levels for genes that were highly expressed in ESCs, as their promoters also had the lowest methylation level in WT ESCs, while no further correlation between transcription and DNA methylation was found.

Most known miRNAs (which are also expressed during PGC development) were uneffected by Dnmt1 deletion. I found two imprinted loci harbouring clusters of miRNAs which were upregulated upon global hypomethylation. More detailed analysis of the Dlk locus confirmed that the demethylation of the ICR was associated with the transcriptional upregulation of genes and miRNAs in this locus.

Repressive histone modifications (H3K9me2 and H3K27me3) are significantly remodelled during PGC development and in preimplantation embryos (Hajkova et al., 2008, Seki et al., 2007). However, H3K9me2/3 and H3K27me3 becomes enriched in preimplantation embryos (Santos et al., 2005, Liu et al., 2016, Zheng et al., 2016). In order to ascertain if there was also global loss of H3K9me2 and enrichment of H3K27me3 in hypomethylated ESCs it would be necessary to check the levels of these histone marks by western blotting. Nevertheless, I saw that the enrichment of these marks throughout the genome confirmed earlier studies (Mikkelsen et al., 2007, Marks et al., 2012).

The enrichment of H3K27me3, H3K9me3 and H3K9me2 did not change upon Dnmt1 deletion. This also explains the missing connection between repressive histone marks and transcription of genes in ESCs. At the same time, regions which resisted global hypomethylation throughout the time-course of Dnmt1 deletion did not change in histone mark occupancy. I consider it most likely that resistance to demethylate reflects targeting of these regions by *de novo* methyltransferases DNMT3a and DNMT3b.

After I confirmed that the *Dnmt1* conditional KO in ESCs served as a good system to recapitulate global demethylation dynamics *in vivo*, I tested the hypothesis that pervasive transcription, antisense to genes, across TEs may work as a trap of transcriptional activation of TEs.

The key conclusion from this work is that global demethylation in ESCs leads to sense strand transcriptional activation of certain TE classes, while the antisense transcription across TE classes was constant. I demonstrated that overlapping sense/antisense transcription was indeed found in TEs and that this can feed into an endosiRNA pathway working through AGO2 to potentially silence TE classes by PTGS.

The *Dnmt1* KO system enabled me to uncover TE classes potentially controlled by endosiRNAs and repressive histone marks as well as other TE classes, that were solely enriched in repressive histone marks to potentially re-silence TEs during phases of transcriptional activation through acute DNA demethylation.

In more detail, I found specific ERV elements, IAPs and MMERVK10C, RLTR45 and RLTR1B, which partially resisted global demethylation. In the case of IAP elements, this recapitulates demethylation kinetics *in vivo* in preimplantation embryos and in PGC development (Seisenberger et al., 2012, Kobayashi et al., 2012).

Additionally, the *Dnmt1* conditional KO system allowed me to dissect the transcriptional response of TE classes after acute DNA demethylation and test my hypothesis that gene derived pervasive transcripts can sense TE transcription.

Interestingly, only a subset of TEs became transcriptionally upregulated upon *Dnmt1* deletion, among which were IAPs and MERVL elements; these TEs have been reported before to become transcriptionally active *in vivo* during PGC development and in preimplantation embryos (Molaro et al., 2014, Friedli et al., 2014).

Furthermore, I found the same classes of TE elements upregulated in my system as during global waves of demethylation during the transition from serum to 2i grown ESCs (Walter et al., 2016). However, while in serum to 2i transition also LINE elements became transcriptionally upregulated in my global hypomethylation system by deletion of *Dnmt1* LINE elements were expressed independent of DNMT1 across the whole time-course. Yet, with my system I created a KO of the DNA maintenance methylase machinery, which recapitulates more closely *in vivo* demethylation dynamics (von Meyenn et al., 2016).

Sense/antisense transcription has been described to occur in TEs (Svoboda et al., 2004) and in the oocyte pseudogenes provide the antisense strand to TEs to feed into an RNAi pathway (Tam et al., 2008).

In the case of ERV elements, the insertion of forward and reverse promoter sequences (LTRs) can produce sense/antisense reads (Svoboda et al., 2004). Here I identified pervasive genic transcription antisense to TEs as a trap of TE sense transcriptional activation.

Interestingly, earlier studies have shown that TEs are mostly found in intergenic regions but if found in genic locations, they are integrated in an antisense direction to the genes (Medstrand et al., 2002; Zhang et al., 2008). This suggests a strong negative selection bias by the genome as the integration of TEs is random (Brady et al., 2009, Lander et al., 2001).

In this study, I present a model that the antisense insertion of TEs into genes may represent a mechanism used by the genome to control the TE transcriptional upregulation. I suggest that genic sense transcription in antisense direction to TEs may work as a trap of TE sense transcription, this would increase the production of dsRNAs and feed into an endosiRNAs pathway to control TEs post-transcriptionally.

The small RNAs that mapped to different TE elements were shown to be bound by AGO2. Although, I could detect sense and antisense reads which mapped to all repeat classes, I found a sense bias in IAP and ETn mapped small RNAs.

This may be explained by the action of AGO2, which endonucleolytically cleaves its target, and could thereby lead to TE RNA degradation products. However, I cannot rule out the possibility that there could be a mixed class of small RNAs that play a role in TE silencing. Furthermore, the increased amount of AGO2 bound small RNAs mapping to IAPs, ETnERV2 and MMERVK10C upon *Dnmt1* KO could be explained by the higher abundance of IAP transcript produced upon global hypomethylation. This interpretation does not mean that the small RNAs I find to be bound by AGO2 are not important but rather that their enrichment upon *Dnmt1* KO could be overestimated.

In chapter 4, I will address the role of these small RNAs in TE silencing upon *Dnmt1* KO in more detail.

To date, several histone modifications have been implicated a role in TE silencing. Knockout studies in ESCs of H3K9 histone methyltransferases ESET and G9a have shown that members of the ERV families are highly repressed by H3K9me2/3 (Maksakova et al., 2013). H3K9me3 is highly enriched on IAPs in ESCs potentially to prevent IAPs from expression upon DNA hypomethylation (Mikkelsen et al., 2007, Hutnick et al., 2010). The knockout of ESET in PGCs leads to a loss of H3K9me3 on ERV elements - specifically IAPs (Leung et al., 2014, Hutnick et al., 2010). Additionally, H3K27me3 has also been found to be enriched in TEs during epigenetic reprogramming in PGCs at E13.5 (Ng et al., 2013) but in PGCs at E11.5 no specific enrichment of H3K27me3 has been found on IAPs (Kim et al., 2014). Therefore, H3K27me3 could play a role in certain TE classes but potentially not in the silencing of IAP elements.

My observations show a similar redistribution of repressive marks as reported in ear-

lier studies during demethylation induced by 2i (Walter et al., 2016). H3K27me3 was also identified in our system to be a key potential driver of TE silencing of MERVL elements, as previously reported (Walter et al., 2016).

The present study shows that MMERVK10C, as the only element, had enrichment of H3K9me3 upon Dnmt1 deletion. It is also notable that this element is the one most resistant to Dnmt1 deletion induced demethylation consistent with H3K9me2/3 marked regions being more resistant to global demethylation (von Meyenn et al., 2016).

Sharif et al., (2016) proposed a model of UHRF1 control of IAP activity through a combinatorial effect of DNA methylation together with H3K9me3 (Sharif et al., 2016), however, in my system upon *Dnmt1* deletion H3K9me3 became depleted at IAPs, with no subsequent enrichment, in agreement with a recent report using ESCs serum to 2i transition (Walter et al., 2016). Nevertheless, I cannot exclude the fact that potentially H3K9me3 might play a role at later stages after *Dnmt1* KO, as I only looked at day 8 after deletion.

I suggest a model in which endosiRNAs play a key role as a first response after transcriptional activation of IAPs. It could well be the case that the small RNAs target the histone marks to the TE classes - in order to test this KOs of histone methyltransferases in the background of *Dnmt1* deletion would need to be carried out.

In summary, this study presents a mechanistic dissection of epigenetic modifications on silencing of TE elements. Previous studies have shown that TEs are silenced by DNA methylation (Walsh et al., 1998, Arand et al., 2012, Sharif et al., 2016), histone marks (Karimi et al., 2011, Walter et al., 2016, Leung et al., 2014) as well as small RNAs (Flemr et al., 2013,Svoboda et al., 2004, Tam et al., 2008, Aravin et al., 2007). This study took the unique approach in analysing all three epigenetic marks together as means to preserve genome integrity by controlling TE activity. As such, I was able to identify endosiRNAs as key players of TE silencing as well as the importance of several epigenetic modifications to work in parallel in order to keep the genome intact during global hypomethylation in ESCs.



Figure 3.43: Schematic of endosiRNAs becoming expressed to control TEs upon *Dnmt1* deletion in ESCs. TEs in mouse ESCs are methylated (filled lollipops) and get demethylated (open lollipops) upon *Dnmt1* KO. Sense transcription of TEs (red) is inhibited by DNA methylation and gets upregulated upon *Dnmt1* KO. Genic sense transcription (grey) is constant antisense to TEs. This results in the production of dsRNA and allows the production of endosiRNAs bound by AGO2 (black) which silence TEs via post transcriptional gene silencing. Additionally, histone marks (blue hexagons) become enriched at TEs which get transcriptionally upregulated and may be involved in the subsequent transcriptional silencing.

Chapter 4

TE regulation upon *Dicer* KO and *Dicer/Dnmt1* KO

4.1 Introduction

The transcriptional activity of TEs induced by *Dnmt1* KO potentially exposes an achilles heel for TEs, as it activates small RNA-based mechanisms that may serve to restrain the TEs. Indeed small RNAs have been described to play a role in TE control with miRNAs as well as endosiRNAs having been found to be involved in TE silencing in the oocyte (Heras et al., 2013, Flemr et al., 2013, Stein et al., 2015).

In the previous chapter I showed that endosiRNAs bound to TEs became enriched at TEs during induced global demethylation in mouse ESCs.

In this chapter, I examine whether these endosiRNAs play a role in TE repression after TE transcriptional activation upon *Dnmt1* KO. Additionally, I aim to test the hypothesis that lowly expressed LINE-1 elements throughout the time-course of *Dnmt1* deletion are post transcriptionally repressed by small RNAs, similarly to LINE element repression by endosiRNAs in human cancer cell lines (Goodier and Kazazian, 2008). I constructed *Dicer* constitutive and conditional double KOs in the background of *Dnmt1* KO ESCs - (DKO) and (cDKO) respectively, to compare the long term effect of *Dicer* KO with an acute deletion of this cellular endonuclease.

Eukaryotic small RNAs are generated through the RNAse-III activity of DICER, which is the central player of the RNA interference (RNAi) network mediating gene silencing on a transcriptional or post transcriptional level, TGS or PTGS, respectively (Bartel, 2004, Hammond, 2005). RNAi plays a role in early development as the knockout of crucial players of the RNAi pathway - *Ago2*, *Dicer* and *Dgcr8* - lead to early embryonic lethality (Bernstein et al., 2003). Nevertheless, it is possible to obtain *Dicer* KO mouse ESCs that show normal stem cell properties but an inability to differentiate *in vitro* and *in vivo* (Kanellopoulou et al., 2005). *Dicer* KO in mouse ESCs leads to a loss of the RNAi machinery by complete depletion of miRNAs (Kanellopoulou et al., 2005, Murchison et al., 2005). Further to the depletion of small RNAs, *Dicer* KO in ESCs has been shown to also have an effect on DNA methylation levels by controlling Dnmt expression through the miR-290 cluster (Sinkkonen et al., 2008, Benetti et al., 2008). In this chapter, I characterise the transcriptome of my *Dicer/Dnmt1* double KO (DKO) and compare them to previously studied *Dicer* KO ESCs.

After this genome-wide characterisation of my *Dicer/Dnmt1* double KO (DKO), I examine the role of DICER in TE transcriptional re-silencing, after global hypomethylation.

Dicer knockdown in mouse preimplantation embryos leads to transcriptional upregulation of MERVL and IAP elements through the production of dsRNAs (Svoboda et al., 2004). EndosiRNAs have been found to regulate TE transcription in the growing oocyte (Tam et al., 2008, Watanabe et al., 2006) and *Dicer* KO in human cancer cell lines leads to upregulation of LINE elements (Goodier and Kazazian, 2008). I follow TE transcriptional activation by total RNA-seq and qRTPCR in the *Dicer/Dnmt1* double KO (DKO) and conditional double KO (cDKO). To understand whether TE transcriptional re-silencing, after global hypomethylation, is achieved by repressive histone marks or by PTGS through small RNAs I also perform ChIP-seq of H3K27me3, H3K9me2 and H3K9me3 in *Dicer/Dnmt1* DKO ESCs. Recent work has shown that miRNAs as well as endosiRNAs can play a major role in of TE silencing in mouse ESCs (Flemr et al., 2013) and I present an interesting miRNA binding to IAPEZ and investigate the potential for translational gene silencing of TEs in ESCs.

This chapter sheds light onto the interplay of small RNAs and DNA methylation in ESCs. Further it uncovers the role of DNA methylation and small RNAs in TE regulation in constant *Dicer/Dnmt1* DKO and acute cDKO ESCs.

4.2 Results

4.2.1 Constitutive KO of *Dicer* by CRISPR-Cas9

To assess whether AGO2 bound small RNAs had a mechanistic role in TE silencing during global hypomethylation in Dnmt1 KO ESCs, I generated a Dicer KO via CRISPR-Cas9 in the $Dnmt1^{fl/fl}$ ESCs. From this, I was able to create Dicer KO as well as Dicer/Dnmt DKO ESCs.

The *Dicer* KO was constructed with Cas9 guide RNAs targeting Exon 23 and 24 - encoding the RNAse III domain and the RNA binding domain, respectively (Bernstein et al., 2003). The resulting genomic deletion resulted in a catalytically inactive protein (Figure 4.1).

Dicer mRNA levels in *Dicer* KO ESCs ($Dicer^{-/-}/Dnmt1^{fl/fl}$) were reduced compared to WT ($Dicer^{+/+}/Dnmt1^{fl/fl}$) ESC (Figure 4.2 left). To confirm the loss of enzymatic activity in the *Dicer* KO ESCs, I checked for the expression of endogenously expressed miRNAs in *Dicer* KO compared to WT ESCs, by small RNA specific qRTPCR.

The expression of mmu-miR93 in *Dicer* KO compared to WT ESCs was abolished. However, small nuclear RNAs (snoRNAs), which are DICER independent were still being expressed (Figure 4.2 right).

After I confirmed that the engineered *Dicer* KO genes encoded for an enzymatically inactive DICER protein, I characterised the transcriptome of those ESCs.



Figure 4.1: CRISPR-Cas9 KO of *Dicer* in $Dnmt1^{fl/fl}$ ESCs. KO strategy for *Dicer* in mouse ES cells constructing gRNAs against Exon 23 and 24 of Dicer mRNA. gRNA protospacer adjacent motif (PAM) sequences (red).





4.2.2 Transcriptional changes in *Dicer* KO and *Dicer/Dnmt1* double KO ESCs



Figure 4.3: Differentially expressed genes between *Dicer* KO and WT ESCs. Scatter plot of RNA-seq data of *Dicer* KO (y-axis) versus WT (x-axis) ESCs. Differentially expressed genes were called by intensity difference of Seqmonk (black), all other genes are depicted in grey. RNA-seq libraries (n=1).

To study the effect of DICER on the mRNA transcription, I carried out total RNA-seq of Dicer KO ECSs (list of total RNA-seq datasets in appendix). Additionally, I performed RNA-seq in the *Dicer* KO at day 1 and day 11 after CRE mediated deletion of Dnmt1, to analyse the effect of depletion of small RNAs as well as global hypomethylation on self-renewal properties of ESCs. First, I compared the transcriptional profiles of *Dicer* KO and WT ESCs I called differentially expressed genes using the Seqmonk intensity difference filter with Benjamini and Hochberg for multiple testing correction with a p-value threshold of < 0.05. Most of the differentially expressed genes were transcriptionally silenced upon deletion of *Dicer* (n = 90). The downregulated genes comprised *Lin28* as well as *Dnmt3L*. Furthermore, I found a small number of genes (n = 7) which were upregulated upon *Dicer* KO, among these was the developmentally important gene Lefty1 (Figure 4.3) (list of differentially expressed miRNAs upon Dicer KO in appendix). To finish the analysis of the Dicer KO ESCs, I carried out a GO Term analysis, using Panther webtool, of the differentially expressed genes between KO and WT ESCs. Biological processes which were mostly affected by the KO of *Dicer* were involved in organism development and angiogenesis (Figure 4.4).

GO biological process complete	Pop Total(22322)	Count(51)	list (expected) Fold Enrichment		PValue
blood vessel development	484	15	1.11	13.56	1.32E-09
asculature development	510	15	1.17	12.87	2.77E-09
cardiovascular system development	523	15	1.19	12.55	3.96E-09
circulatory system development	822	17	1.88	9.05	1.47E-08
angiogenesis	289	12	0.66	18.17	1.73E-08
blood vessel morphogenesis	383	13	0.88	14.86	2.27E-08
collagen fibril organization	38	7	0.09	80.63	3.52E-08
extracellular matrix organization	185	10	0.42	23.66	1.12E-07
extracellular structure organization	186	10	0.42	23.53	1.18E-07
single-multicellular organism process	5010	33	11.45	2.88	1.02E-06
nulticellular organism development	4353	30	9.95	3.02	5.66E-06
single-organism developmental process	5001	32	11.43	2.8	5.95E-06
developmental process	5045	32	11.53	2.78	7.52E-06
anatomical structure development	4735	31	10.82	2.87	8.24E-06
anatomical structure morphogenesis	2106	21	4.81	4.36	1.57E-05
wound healing	254	9	0.58	15.51	4.96E-05
esponse to wounding	311	9	0.71	12.67	2.78E-04

Figure 4.4: GO term analysis of differentially expressed genes between *Dicer* KO and WT ESCs. Significantly expressed genes were analysed by Panther GO Term analysis.



Figure 4.5: Differentially exgenes between pressed Dicer KO, Dicer/Dnmt1 DKO and WT ESCs. Heatmap of hierarchical clustering of genes differentially expressed between Dicer KO, Dicer/Dnmt1 DKO Per probe normalised reads ESCs. are depicted of all RNA-seq datasets quantified as RPM. Each dataset is sequenced as n=1. red = not expressed, blue = highly expressed.
Subsequently, I carried out a differential expression analysis between *Dicer* KO, *Dicer/Dnmt1* double KO (DKO) with a deletion of the DNA methyltransferase *Dnmt1* induced demethylation for 1 day as well as 11 days and WT ESCs as control. Also for this analysis, I used the Seqmonk intensity difference filter with Benjamini and Hochberg for multiple testing correction with a p-value threshold of < 0.05.



Figure 4.6: Expression of *Lin28*, *Dnmt3l*, *Fbln2* and *Oct4* in *Dicer* KO, *Dicer/Dnmt1* DKO ESCs. Chromosome view of read count quantitation across the 4 genes *Lin28*, *Dnmt3l*, *Fbln2* and *Oct4*. High bars indicated high expression, low bars indicate low expression. Every bar overlaps at least 1 read.

I found 258 significantly differentially expressed genes between the different conditions. Genes in cluster I and IV were up- or downregulated solely upon *Dicer* KO (Figure 4.5). Genes in cluster II were upregulated solely upon *Dnmt1* KO. Genes in cluster III showed upregulation upon *Dicer/Dnmt1* DKO (Figure 4.5). These genes were upregulated in an additive way upon knockout of both genes, they were not expressed in WT and day 1 after *Dnmt1* KO, but were upregulated at day 11 after *Dnmt1* deletion as well as in *Dicer* KO, with a further upregulation in the *Dicer/Dnmt1* DKO - like *Fbln2* in Figure 4.6. This cluster comprised a lot of cell adhesion genes, indicating a change in cell morphology upon KO of both *Dicer* and *Dnmt1* (list of genes in Cluster III in appendix).

The pluripotency genes - *Sox2*, *Klf4*, *Esrrb*, *Oct4* and *Nanog* - were not significantly differentially expressed in the *Dicer* KO as well as the *Dicer/Dnmt1* DKO ESCs (Figure



Figure 4.7: Expression of the pluripotency genes in *Dicer* KO, *Dicer/Dnmt1* DKO ESCs. Bargraph of RNA-seq data of Dnmt genes in WT (grey), *Dicer/Dnmt1* DKO induced for 1 day (green) and 11 days (light green), *Dicer* KO uninduced for 1 day (orange) and 11 days (yellow), conditional *Dnmt1* KO ESCs induced for 1 day (dark red) and 11 days (dark blue). RNA-seq libraries (n=1).

4.6, 4.7), therefore the self-renewal of *Dicer* KO ESCs was not affected. As reported before, the *de novo* methyltransferases *Dnmt3a* and *Dnmt3b* were downregulated upon *Dicer* KO (Sinkkonen et al., 2008, Benetti et al., 2008). However, I observed the strongest down-regulation of expression on *Dnmt3l*. *Dnmt1* was normally expressed in the *Dicer* KO ESCs (Figure 4.8).



Figure 4.8: Expression of the Dnmts in *Dicer* KO, *Dicer/Dnmt1* DKO ESCs. Bargraph of RNA-seq data of Dnmt genes in WT (grey), *Dicer/Dnmt1* DKO induced for 1 day (green) and 11 days (light green), *Dicer* KO uninduced for 1 day (orange) and 11 days (yellow), conditional *Dnmt1* KO ESCs induced for 1 day (dark red) and 11 days (dark blue). RNA-seq libraries (n=1).

In summary, in the *Dicer* KO ESCs I recapitulated the same transcriptional changes that have been reported before (Sinkkonen et al., 2008, Benetti et al., 2008). I found a downregulation of the *de novo* methyltransfereases, furthermore, I found a repression of organism developmental genes, which could lead to the inability of *Dicer* KO ESCs to differentiate. Additionally, I was able to identify a gene groups, that were dependent on DICER as well as DNMT1 and showed an additive transcriptional effect in the *Dicer/Dnmt1* DKO.

4.2.3 Genome-wide histone marks upon *Dicer* KO and *Dnmt1/Dicer* DKO

After an analysis of the transcriptome I analysed the chromatin changes upon *Dicer* KO as well as *Dicer/Dnmt1* DKO ESCs. I tested for three repressive histone marks - H3K9me3, H3K9me2 and H3K27me3 in the different KO ESCs (list of ChIP-seq datasets in appendix), because during *in vivo* epigenetic reprogramming these repressive histone marks are also largely exchanged (Hajkova et al., 2008, Ancelin et al., 2006, Liu et al., 2016, Zheng et al., 2016). I aimed to discover whether the acute demethylation by *Dnmt1* KO and additionally deletion of the small RNAs by deletion of *Dicer* KO would also have an effect on the overall chromatin structure in ESCs.



Figure 4.9: Enrichment of repressive histone marks at genomic features in *Dicer* KO and *Dicer/Dnmt1* DKO. Bar graph of enrichment of H3K27me3, H3K9me3 and H3K9me2 in repeats (dark violet), genic regions (light violet), promoters (dark green), CGIs (middle green), intergenic regions (light green) in WT ESCs, *Dnmt1* KO, *Dicer* KO and *Dicer/Dnmt1* DKO



Figure 4.10: Histone mark enrichment in the genome upon *Dicer* KO and *Dicer/Dnmt1* DKO. Wiggle plot of ChIP-seq enrichment of H3K9me3, H3K27me3 and H3K9me2 at three genomic loci in *Dicer/Dnmt1* DKO at day 11 (light green), *Dicer* KO (orange), *Dnmt1* KO at day 11 (dark blue) and WT (grey). Enrichment intensity shown on y-axis.

At the outset, I analysed the general enrichment of the three histone marks across different genomic locations. As expected, H3K27me3 was enriched in genic regions, in CGIs as well as promoters. Additionally, this repressive histone mark was enriched on repeats. H3K9me3 was depleted in CGIs, promoters and genes but highly enriched over repeats. H3K9me2 was generally distributed across the whole genome with some enrichment in repeats and genic regions. I did not observe any changes of histone distribution for either of the repressive histone marks upon *Dicer* KO or *Dicer/Dnmt1* DKO in ESCs. The same regions were still enriched and I did not find any significant redistribution of histone marks upon deletion of the DNA maintenance methylase or DICER derived small RNAs (Figure 4.9 and 4.10).

As a final analysis, I overlaid transcriptional levels of genes in *Dicer* KO ESCs with chromatin changes upon the single and DKOs of *Dicer* and *Dnmt1*. I could not find any correlation between genes expressed in WT ESCs and chromatin enrichment in the different KOs (Figure 4.11).

The analysis of the transcriptome, as well as chromatin changes upon deletion of *Dicer* in the single KO as well as in combination on *Dnmt1* KO in the *Dicer/Dnmt1* DKO, showed that there was an additive effect of transcriptional changes in the DKO ESCs. Though I could not find any synthetic redistribution of the three histone marks I



Figure 4.11: Transcriptional profile in *Dicer* KO is not explained by histone mark enrichment. Line graph of histone enrichment of genes which showed different expression level in *Dicer* KO ESCs: not expressed (RPM < 0, red), low expressed (2.5 < RPM > 0, violet), expressed (5 < RPM > 2.5, pink), high expressed (10 < RPM > 5, dark blue) or very high expressed (RPM > 10, light blue) as well as genes which were *Dicer* dependent (orange).

studied at the whole genome level.

4.2.4 LINE-1 and major satellite expression in *Dicer* constitutive KO ESCs

In order to address the involvement of *Dicer* derived small RNAs on TE regulation, I conducted a detailed analysis of TEs in the *Dicer* KO and *Dnmt1/Dicer* DKOs. To test whether endosiRNAs were playing an active role in TE repression, I repeated the *Dnmt1* deletion experiments in the *Dicer* KO background. Over the time-course of *Dnmt1* KO in the *Dicer* KO ESCs, while expression of LINE-1 and major satellites were detectable by qRTPCR, surprisingly no expression of IAPs and ETns was found (Figure 4.12).

This result showed that DICER regulated small RNAs were involved in the silencing of LINE-1 elements and major satellites. Major satellites became transcriptionally upregulated following *Dnmt1* deletion and therefore showed dependency on both endosiRNA and DNA methylation. Despite that, LINE-1 elements were highly expressed upon *Dicer* KO, albeit showed no additional transcriptional activation upon *Dnmt1*



Figure 4.12: LINE-1 and major satellites are upregulated in a *Dicer* constitutive KO. qRTPCR analysis of TE classes in *Dicer* KO and *Dicer* KO with conditional *Dnmt1* induced for 1 day (dark red), 3 days (light red), 6 days (light pink), 8 days (light blue), 10 days (dark blue), and 12 days (black). Error bars represent standard deviation of 2 biological replicates with 2 technical replicates. Values were normalized to Atp5b, Hspcb and major satellites to U1. Differences between conditions that are statistically significant are denoted by (* p-value<0.05, ** p-value<0.005, *** p-value<0.0005 (Student's t-test))

KO.

Surprisingly, although IAPs and ETns were Dnmt1 KO dependant and became upregulated upon global demethylation, both TEs were transcriptionally silenced in the Dicer/Dnmt1 DKO ESCs. Nevertheless, upon Dicer/Dnmt1 DKO IAPEZ and ETn both got transcriptionally expressed to a much lower amount than in the single Dnmt1KO.

Earlier studies have shown that IAP elements are specifically silenced by H3K9me3 marks in *Dnmt1* constitutive KO ESCs (Hutnick et al., 2010, Mikkelsen et al., 2007). I wanted to test whether any repressive chromatin enrichment could explain the transcriptional silencing of ETns and IAPs in my *Dnmt1/Dicer* DKO ESCs.



Figure 4.13: *Dicer* dependent TE upregulation. Heatmap of RNA-seq data mapped to TE classes. Only TE classes which were intergenic and within 2kb away from genes were considered. Mean of TE classes with at least 1000 integration sites were regarded and z-score was calculated between KO and WT ESCs. RNA-seq (n=1).

With the help of Simon Andrews, we undertook an unbiased approach to find the TE classes that were solely upregulated in the *Dicer* constitutive KOs in our RNA-seq data (Figure 4.13). For the analysis, we followed the mapping of TEs as described in chapter 3, and then analysed only the TE classes which were differentially expressed with a threshold of p < 0.05 in *Dicer* KO compared to WT ESCs. Only seven TE classes were differentially upregulated upon *Dicer* KO - and therefore defined as DICER dependent TEs. Two classes of IAP, IAPLTR3 and IAP-d, had already been identified in chapter 3 as being dependent on *Dnmt1* and also showed a dependency on *Dicer* KO.

4.2.5 Histone modifications may explain different behaviours of TEs to acute *Dnmt1* KO



Figure 4.14: H3K9me3 enrichment over TEs dependent on *Dicer* and *Dnmt1*. Heatmap of ChIP-seq data of H3K9me3 mapped to TE families at day 4 and day 8 after *Dnmt1* KO, *Dicer* KO and *Dnmt1/Dicer* DKO in comparison to WT ESCs. Read count enrichment depicted as z-score. ChIP-seq libraries (n=1).

In early mammalian development chromatin marks become remodelled during global epigenetic reprogramming (Hajkova et al., 2008, Ancelin et al., 2006, Liu et al., 2016, Zheng et al., 2016). IAPs are highly guarded by H3K9me3 and MERVL elements by H3K27me3, respectively (Hutnick et al., 2010, Maksakova et al., 2013). H3K9me2 follows DNA methylation dynamics and gets lost upon global demethylation in PGC reprogramming (Seki et al., 2005). The same dynamics have been seen globally in the serum to 2i transition in mouse ES cells (von Meyenn et al., 2016, Walter et al., 2016). Therefore, I wanted to test if repressive histone marks are enriched at TEs specifically upon KO of *Dicer* as well as in the DKO of *Dicer/Dnmt1*.

I performed ChIP-seq experiments of H3K9me3, H3K9me2 and H3K27me3 in the Dicer

KO and *Dicer/Dnmt1* DKO ESCs with *Dnmt1* KO and WT ESCs as a control. I analysed the enrichment of these marks on TE classes dependent on DNMT1 as defined in chapter 3 (Figure 3.25) and DICER as defined earlier in this chapter (Figure 4.13).



Figure 4.15: H3K9me2 enrichment over TEs dependent on *Dicer* and *Dnmt1*. Heatmap of ChIP-seq data of H3K9me2 mapped to TE families at day 4 and day 8 after *Dnmt1* KO, *Dicer* KO and *Dnmt1/Dicer* DKO in comparison to WT ESCs. Read count enrichment depicted as z-score. ChIP-seq libraries (n=1).

H3K9me3 was depleted upon *Dicer* KO in most TE classes, with a higher depletion in the DKO ESCs. MERVL elements showed a 2-fold enrichment of H3K9me3 in the *Dicer* KO ESCs than in WT ESCs, but were less enriched than in the *Dnmt1* KO ESCs. L1MdGf elements showed higher enrichment of H3K9me3 in the *Dicer* KO ESCs than in *Dnmt1* KO or WT ESCs, but a depletion of this mark in the DKO ESCs. ORR1B1 and RLTR14 showed both an enrichment in *Dicer* KO and *Dicer/Dnmt1* DKO ESCs.

H3K9me2 showed the exact opposite pattern and was generally enriched in all TEs upon *Dicer* KO and *Dicer/Dnmt1* DKO. L1MdGf elements were depleted of H3K9me2 in the *Dicer* KO and even more depleted in the *Dicer/Dnmt1* DKO. MERVL elements



Figure 4.16: H3K27me3 enrichment over TEs dependent on *Dicer* and *Dnmt1*. Heatmap of ChIP-seq data of H3K27me3 mapped to TE families at day 4 and day 8 after *Dnmt1* KO, *Dicer* KO and *Dnmt1/Dicer* DKO in comparison to WT ESCs. Read count enrichment depicted as z-score. ChIP-seq libraries (n=1).

had lost enrichment of H3K9me2 in *Dnmt1* KO to the same level as in the *Dicer* KO but were totally depleted of this repressive histone mark in the *Dicer/Dnmt1* DKO ESCs. Interestingly, IAPs and MMERVK10C lost H3K9me2 specifically upon *Dnmt1* deletion with highest enrichment of this mark in single *Dicer* KO ESCs.

H3K27me3 became enriched in all TE classes upon *Dicer* KO and *Dicer/Dnmt1* KO. While L1mdGf and RLTR45 elements had the highest enrichment of H3K27me3 upon *Dnmt1* KO in comparison to *Dicer* KO and WT ESCs, in IAP-d, MERVL, IAPLTR3 and ERVB4IB the enrichment of H3K27me3 was higher than in WT ESCs in all three KO conditions. Interestingly, IAPEZ, ETnERV2 and MMERVK10C had enrichment of H3K27me3 from WT to *Dnmt1* KO to *Dicer* KO with highest enrichment in the *Dicer/Dnmt1* DKO.

This is a compelling result, as it suggests that in the *Dicer/Dnmt1* DKO IAPEZ, ETns as well as MMERVK10C elements are enriched for the H3K27me3 repressive histone mark, which is likely to prevent those TEs from transcriptional upregulation in the *Dicer* constitutive KO ESCs.

4.2.6 TE response after double conditional KO of *Dicer* and *Dnmt1*

I aimed to investigate whether the unexpected transcriptional repression of IAPs and ETn elements was due to long term *Dicer* deletion, therefore I constructed *Dicer*^{fl/fl} conditional double KO (cDKO) ESCs in the $Dnmt1^{fl/fl}$ background ESCs. The loxP sites were introduced in intron 15-16 and intron 20-21 of the *Dicer* gene (Figure 4.17). The activation of the CRE recombinase by induction with tamoxifen (4OHT) led to recombination between the loxP sequences, which in consequence led to a downregulation of the Dicer mRNA (Figure 4.18A) as well as abrogation of its enzymatic activity shown by depletion of endogenous mmu-miR93 expression (Figure 4.18B).

Dnmt1 deletion experiments were repeated in the $Dicer^{fl/fl}$ background. qRTPCR of time-points after acute Dicer/Dnmt1 cDKO showed significantly higher transcriptional

Figure 4.17: Construction of $Dicer^{fl/fl}$ $Dnmt1^{fl/fl}$ \mathbf{in} **cDKO ESCs.** Knockout strategy for Dicer in mouse ES cells introducing loxP sites in Intron 15-16 and intron 20-21 of Dicer. Agarose gel of PCR to screen for genomic recombination of 2 Dicer cDKO clones after addition of tamoxin for 3 days. Recombination of Intron 15-16 was tested with primer set 1, recombination of intron 20-21 was tested with primer set 2 and recombination of both introns was tested with primer set 3, LD = 1000bp DNA ladder.

4.2. Results





Figure 4.18: Dicer mRNA and miRNAs expression in *Dicer/Dnmt1* cDKO. (A) Dicer mRNA is getting downregulated upon CRE recombination induced by tamoxifen in 1 clone, (B) endogenously expressed miRNAs in mouse ES cells are downregulated upon *Dicer* KO controlled by snoRNA expression. Each experiment was done in 3 technical replicates. Differences between conditions that are statistically significant are denoted by (* p-value<0.05, ** p-value<0.005, (Student's t-test))

activation of IAPs on day 10 (Figure 4.19). Additionally, on-going expression of IAPs was found at later time-points (Day 12) with no significant resilencing of the IAPs. This indicates that the small RNAs produced were DICER dependent and played a role in IAP silencing upon *Dnmt1* KO.

MERVL elements showed the same transcriptional kinetics upon *Dicer/Dnmt1* cDKO as for *Dnmt1* KO and became transcriptionally activated and subsequently resilenced. While ETn elements were upregulated and subsequently resilenced upon *Dnmt1* KO,

these elements were not transcriptionally activated in the *Dicer/Dnmt1* DKO, nor in the *Dicer/Dnmt1* cDKO ESCs. This may be due to their resilencing by histone modifications, but could also be explained by Dicer independent small RNAs, as I found an increase of small RNAs mapping to the ETn elements in the third chapter.



Figure 4.19: TE expression in the $Dnmt1^{fl/fl}$ and $Dicer^{fl/fl}/Dnmt1^{fl/fl}$ cDKO ESCs. qRTPCR analysis of TE classes in the $Dnmt1^{fl/fl}$ and $Dicer^{fl/fl}/Dnmt1^{fl/fl}$ cDKO ESCs induced for 1 day (dark red), 3 days (light red), 6 days (light pink), 8 days (light blue), 10 days (dark blue), and 12 days (black). Error bars represent standard deviation of 2 biological replicates with 2 technical replicates. Values were normalised to Atp5b, Hspcb and major satellites to U1. Differences between conditions that are statistically significant are denoted by (* p-value<0.05, ** p-value<0.005, (Student's t-test))

MMERVK10C were only upregulated in *Dnmt1* KO ESCs 25 days after KO induction, therefore it is not surprising that they were not upregulated in the *Dicer/Dnmt1* cDKO ESCs until day 12. Major satellites as well as LINE elements were not dependent on knockout of the maintenance methylation machinery alone but in the *Dicer/Dnmt1* cDKO they became transcriptionally upregulated. In order to check whether major satellites have the ability to produce endosiRNAs, I mapped the total RNA-seq dataset to major satellite consensus sequences (Figure 4.20A). I observed sense and

antisense transcripts in WT as well as *Dnmt1* KO ESCs at day 9 after induced CRE recombination (Figure 4.20B). Additionally, I pulled down AGO2 bound small RNAs and mapped them back to the major satellite consensus sequence (Figure 4.20C).



Figure 4.20: Long and short small RNAs map to major satellites. (A) major satellite (GSAT element) was analysed by unique mapping in the genome and a general expression of this elements, red bar depicts the mean expression level. (B) sense (red) and antisense (blue) reads mapped to major satellite consensus sequence by piPipes small RNA-seq analysis tool. (C) Bargraph of small RNAs pulled down by AGO2 mapped to major satellite consensus sequence, in WT ESCs = grey and Dnmt1 KO ESCs at day 9 = blue. Experiments were performed as 4 biological replicates.

In summary, I was able to identify DICER dependent small RNAs, that controlled LINE-1 elements and major satellites, as well as IAP elements during acute demethylation in ESCs.

Furthermore, I found histone mark enrichment at RNAi independent TE classes and concluded that upon genome hypomethylation, histone modifications and RNAi play a major role to keep TE elements transcriptionally silenced.

4.2.7 miRNA production against IAPs suggest multiple levels of gene regulation

In chapter 3, I identified small RNAs perfectly mapping to the IAP transcript, but also some which imperfectly mapped to the consensus sequence (Figure 4.21). I aimed to investigate whether these smallRNAs could be miRNAs, that regulate IAP expression at the level of translation.

Mmu-miR-7081 mapped to the IAP consensus sequence independent of Dnmt1 deletion



Figure 4.21: miRNAs map to IAPEZ consensus sequence independent of DNA demethylation. Consensus sequence mapping of small RNA-seq data to IAPEZ consensus sequences allowing 2 mismatches, sense reads = red, antisense reads = blue, arrow indicating miRNA.

(Figure 4.21). As the miRNA is expressed independent of DNA methylation, the potential mechanism of translational gene silencing by this miRNA is already in place before transcriptional activation of IAPs. The mmu-mir-7081 mapped to the 3'LTR of the polymerase of IAPs. As a *bona fide* miRNA it mapped with its 5'seed region totally complementary to the IAP 3'LTR followed by a bulge sequence and additional posterior binding (Figure 4.22). The pre-miRNA gene of mmu-mir-7081 is encoded in an intron of *Dnmt1* and is only conserved in the murine lineage (Figure 4.23). This is curious, as also IAPs are also specific to the rodent lineage. I suggest that the miRNAs could have coevolved with IAPs to control this highly active TE in the mouse germline.



Figure 4.22: Mmu-miR-7081 complementary to IAP consensus sequence. miRNA 7081 binds with 5' seed region to IAP sequence at nucleotide 6450. The seed region is binding followed by a 3 nt bulge sequence, additionally there is some 3' binding with additional bulges. Local alignment was done using EMBOSS Matcher alignment tool with default parameters.

I confirmed the constant expression of pre-mmu-miR-7081 in the small RNA-seq libraries across the whole time course after *Dnmt1* deletion (Figure 4.24B) by targeted small RNA qRTPCR at day 9 after *Dnmt1* and detected mmu-miR-7081 expression throughout the whole time-course (Figure 4.24A).



Figure 4.23: Mmu-mirR-7081 is conserved in the murine lineage. Evolutionary conservation of mmu-miR-7081 across different species. UCSC genome browser.

Figure 4.24: Expression of mmu-miR-7081 upon *Dnmt1* KO.(A) small RNA qRTPCR on mature mmu-miR-7081 at day 9 after Dnmt1 deletion (light blue) and wildtype (grey), (B) small RNA-seq of miRNA expression in wild type (grey) and conditional *Dnmt1* knockout ESC induced for 1 day (dark red), 3 days (light red), 6 days (light pink), 9 days (light blue), 11 days (dark blue).



4.2.8 IAP protein production upon *Dicer* KO

In order to study whether transcriptional activation of IAPs leads to production of IAP protein we undertook immunofluorescence (IF) staining of $Dicer^{-/-}$ ESCs. Only in the double KO of Dicer and Dnmt1 IAP protein was detected by confocal microscopy (Figure 4.25). In the single Dnmt1 KO and Dicer KO as well as the WT ESCs no IAP protein was observed. This indicates that Dicer dependent small RNAs are needed for translational resilencing of IAPs and opens the possibility for miRNA-7081 to play an important role.

Altogether, this could suggest that there are two ways of TE silencing through small RNAs in ESCs. The first one is through endosiRNA dependent post-transcriptional gene silencing of TE transcripts and the second one is through translational repression and depletion of the TE protein.



Chapter 4. TE regulation upon Dicer KO and Dicer/Dnmt1 KO 4.3. Discussion

Figure 4.25: IAP protein production upon *Dicer* KO in *Dnmt1* KO ESCs. Immunofluorescent staining of *Dnmt1* KO, *Dicer* KO as well as *Dicer/Dnmt1* DKO ESCs. Nuclei are stained with DAPI in blue and IAP is stained in red. scale bar = 50μ m.

4.3 Discussion

In chapter 3 of this thesis, I found AGO2 bound small RNAs mapping to TEs. Additionally, my KD study in chapter 3 suggested that IAP transcription is dependent on DICER. I concluded that endosiRNAs are likely to be the major RNAs involved in TE regulation in ESCs.

In this chapter, I aimed to test whether DICER derived small RNAs are regulators of TE resilencing after transcriptional upregulation during global hypomethylation.

I constructed the *Dicer* constitutive KO ESCs by a deletion of the enzymatic activity and could confirm the loss of enzymatic activity by loss of miRNAs normally expressed in WT ESCs.

Due to differing results of phenotypes and transcriptional changes in different *Dicer* KO ESCs from other labs (Kanellopoulou et al., 2005, Murchison et al., 2005), I wanted

to characterise my KOs carefully and carried out a transcriptional analysis through RNA-seq. GO term analysis of the differentially expressed genes showed that genes important for organism development were substantially downregulated in *Dicer* KO ESCs in comparison to WT ESCs, which suggests that the KO ESCs could have a differentiation defect. Indeed, earlier studies have found a defect in differentiation of *Dicer* KO ESC cells *in vitro* and *in vivo* (Kanellopoulou et al., 2005, Murchison et al., 2005).

As reported before, *Dnmt3a* and *Dnmt3b* were downregulated in our *Dicer* KO ESCs, while *Dnmt1* was normally expressed (Sinkkonen et al., 2010, Murchison et al., 2005). Interestingly, *Dnmt3l* was the most downregulated Dnmt, with no expression in *Dicer* KO ESCs.

Additionally, I found an intriguing additive effect in the *Dicer/Dnmt1* DKO ESC. After day 1 of induction of *Dnmt1* KO a cluster of genes (Figure 4.25) was much more lowly expressed than at day 11 after induction of *Dnmt1* KO and the expression increased in the *Dicer* KO compared to WT ESCs. This could be attributed to the enzymatic effect of DICER and DNMT1 - and therefore to the activity of small RNAs and DNA methylation - respectively. However, this could also be the effect of DICER on Dnmt expression through small RNAs, as has been suggested before (Murchison et al., 2005).

As the RNA-seq results were all done in one *Dicer* KO clone with no replicate, it is important to prepare another replicate and repeat the results before coming to any conclusions. However, if the additive effect is confirmed, it would be possible to do RNA-seq of the cDKO ESCs of *Dicer/Dnmt1* at different time points to study the dynamics of DNA methylation and small RNAs on transcription in mouse ESCs. After the transcriptional analysis I also studied the chromatin changes upon *Dicer* KO. No genome-wide redistribution of H3K9me3, H3K27me3 or H3K9me2 was found upon DNA hypomethylation and depletion of Dicer dependent small RNAs.

After the analysis of transcriptome as well as chromatin changes upon *Dicer* KO in ESCs, I studied the TE response in the *Dicer/Dnmt1* DKO and cDKO ESCs.

However, I could not find any transcriptional upregulation of IAPs nor ETn elements in the *Dicer/Dnmt1* DKO ESCs, while LINE elements and major satellites were highly upregulated in the DKO ESCs.

I suggest that that upon *Dicer* constitutive KO active TEs were silenced through repressive histone marks. In order to test this hypothesis I performed histone ChIP-seq of repressive histone marks H3K9me3, H3K9me2 and H3K27me3 in the *Dicer* KO ESCs compared to *Dicer/Dnmt1* DKO, *Dnmt1* KO and WT ESCs.

Interestingly, I saw high enrichment of H3K27me3 on IAPs as well as ETh elements upon *Dicer* KO. This enrichment was even higher in the *Dicer/Dnmt1* double KO ESCs. This could mean that H3K27me3 histone marks are influenced by the existence of small RNAs, and their depletion could result in higher H3K27me3 redistribution in the genome. To assess a relationship between the H3K27me3 enrichment in the *Dicer* KO ESCs, it would be necessary to test the specificity of the enrichment by deleting the PRC2 complex in the *Dicer* and *Dicer/Dnmt1* KO ESCs.

After this, I constructed the conditional *Dicer/Dnmt1* cDKO ESCs and tested for TE activity. I could only find lack of IAP resilencing in the double conditional KO of both *Dicer* and *Dnmt1*. This is an intriguing finding, as it suggests that in the DKO in the state of hypomethylation as well as depletion of small RNAs, targeted chromatin changes on very active TE classes, such as IAPEZ, may prevent these elements from transcriptional expression.

The transcriptional upregulation of IAPs could not be followed for longer than 12 days, as the ESCs started to die afterwards. This could be due to uncontrolled retrotransposition events of IAPs. In order to test this hypothesis it is possible to measure IAP protein levels or to carry out retrotransposition assay of this TE. Another way of testing for retrotransposition would be to inverstigate chimeric transcripts in RNA-seq data of *Dicer/Dnmt1* cDKO ESCs over time.

To my surprise, ETn elements were still guarded by unknown mechanisms in the cDKO ESCs throughout the time-course of deletion. ETn elements are highly active in the mouse germline and guarded by DNA methylation; earlier I found AGO2 bound small RNAs mapping to ETns. These TEs were transcriptionally silenced in the *Dicer* KO as well as in the *Dicer/Dnmt1* cDKO ESCs. We found high enrichment of H3K9me2 repressive histone marks upon *Dicer/Dnmt1* DKO on ETn elements and therefore suggest that ETn elements are not resilenced by endosiRNAs primarily but rather by transcriptional gene silencing through repressive histone marks. However ETns could also be guarded by additional repressive histone marks, which I have not tested for, as their histone enrichment after hypomethylation matched with that of IAP elements.

Furthermore, in this study I only dissected the role of DICER dependent small RNAs on TE silencing, although DICER independent AGO2 bound small RNAs could also play a role in TE silencing. DICER independent small RNAs are highly abundant in mouse ESCs (Babiarz et al., 2008) and are important for major satellite silencing (Cao et al., 2009). DICER independent small RNAs could also explain the repression of ETn elements, which had increasing amounts of AGO2 bound small RNAs mapped to them, but were not responsive upon KO of this cellular endonuclease, therefore the produced small RNAs that mapped to ETns might be DICER independent.

Transcription of LINE-1 elements and major satellites is not dependent on DNA methylation, as both repeats were expressed in ESCs with or without *Dnmt1* deletion (Chapter 3). On both elements AGO2 bound small RNAs with an endosiRNA signature were found. *Dicer/Dnmt1* DKO, as well as *Dicer/Dnmt1* cDKO led to transcriptional upregulation of both elements. This suggests that in mESCs the low transcriptional activity of LINEs and major satellites is due to the activity of endosiRNAs keeping them repressed all the time. However, LINEs as well as the pericentric repeat class are known to be silenced by repressive histone marks (Saksouk et al., 2014, Tsumura et al., 2006, Ip et al., 2012, Goodier and Kazazian, 2008) and I also found increased enrichment of histones upon in upon KO of *Dicer* and *Dnmt1*.

The differential silencing of LINE-1 and major satellites primarily by endosiRNA while ERV elements are highly guarded by DNA methylation, could be due to the fact that the older TEs are already domesticated in the mouse genome while ERV elements are still highly active in the male germline. IAP elements are then in the second instance primarily guarded by endosiRNAs while the ETn elements are guarded by repressive histone marks. I also found a potential role of translational gene silencing of IAP elements, through an miRNA which is transcribed in an intron of *Dnmt1*. This is very intriguing, as the pre-miRNA of the mature miRNA as well as the IAPs are both confined to the murine lineage. This could suggest that the miRNA mechanism of IAP control has coevolved with IAPs in order to guard the genome from activity of this specific ERVK elements. To test this hypothesis, it is possible to deplete ESCs of this specific miRNA by locked nucleic acids (LNAs) and measure translation of IAP protein by western blot. To reveal whether this specific miRNA also plays a role in the first response to feed into an endosiRNA pathway, as it is the case in plants, it would be necessary to perform the miRNA knockdown at the time of transcriptional expression and check by small RNA-seq, whether endosiRNAs still map to IAP transcripts.



Figure 4.26: Schematic of DICER generated endosiRNAs control TEs upon *Dnmt1* deletion in ESCs. TEs in mouse ESCs are methylated (filled lollipops) and get demethylated (open lollipops) upon *Dnmt1* KO. Sense transcription of TEs (red) is inhibited by DNA methylation and gets upregulated upon *Dnmt1* KO. Genic sense transcription (grey) is constant antisense to TEs. This results in the production of dsRNA and allows the production of DICER dependent (white) endosiRNAs bound by AGO2 (black) which silence TEs via post transcriptional gene silencing. Additionally, histone marks (blue hexagons) become enriched at TEs which get transcriptionally upregulated and may be involved in the subsequent transcriptional silencing.

In summary, I analysed the role of small RNAs during the dynamics of global demethylation in mouse ES cells and were able to show that LINEs, major satellites as well as IAP element expression are dependent on DICER derived endosiRNAs among DNA methylation and histone marks. For other TE elements histone marks and DNA methylation seem to play the primary role in silencing.

Additionally, I also found miRNAs to potentially play a role in IAP silencing.

This could expand the way we think about epigenetics as major regulators of TE silencing during early mammalian development. DNA methylation is in place at regions which need to be remodelled. Accompanying this methylome remodelling are major chromatin changes. Potentially small RNAs - such as siRNAs, piRNAs and miRNAs - are involved in keeping TE elements under control to preserve genome integrity for future generations.

Chapter 5

Small RNAs as suppressors of TEs upon demethylation

5.1 Introduction

In chapter 3, I found that during acute demethylation by *Dnmt1* deletion endosiRNAs bound by AGO2 map to IAPs, LINEs and major satellites. In chapter 4, I presented evidence that these small RNAs are going through an DICER mechanism of repeat silencing.

In this chapter, I examine whether small RNAs could potentially play a role in TE regulation in other systems of global genome demethylation. Accordingly, I compare the small RNAs mapping to TEs in the the transition from (1) serum to 2i grown ESCs, (2) early stages of induced pluripotent stem cell (iPSC) reprogramming as well as (3) *in vitro* and *in vivo* reprogramming in PGC development.

Likewise, I study the expression of miRNAs in the three demethylation systems. I compare the miRNAs that influence the pluripotency of ESCs and are indispensable during PGC development across the (1) serum to 2i grown ESCs, (2) early stages of induced pluripotent stem cell (iPSC) reprogramming as well as (3) *in vitro* and *in*

vivo reprogramming in PGC development (Zhao et al., 2014, Hadjimichael et al., 2016, Medeiros et al., 2011, Takada et al., 2009, Hayashi et al., 2008, Jia et al., 2013, Bhin et al., 2015).

miRNAs have been reported to control germ cell development. Knock-out of *Dicer* solely in PGCs, results in poor germ cell proliferation (Hayashi et al., 2008). Additionally, the depletion of the whole miR290-295 cluster in mice leads to embryonic lethality through the impairement of PGC development and premature failure in the ovary (Medeiros et al., 2011). The miR290-295 cluster as well as the miR302-367 are highly abundant in ESCs and as both clusters are cell cycle miRNAs, they are invaluable for ESC self-renewal. The miR16 and miR191 clusters are highly expressed in ESCs and target Smad2 and thereby they inhibit mesoderm differentiation. At the same time, miR23 is able to inhibit the endodermal and ectodermal differentiation. However, miRNAs can also promote differentiation - miR421 inhibits BMP signalling and suppresses the pluripotency factor Oct4 (Zhao et al., 2014, Hadjimichael et al., 2016).

(1) Mouse ESCs have the ability to generate all somatic and germline cells *in vitro* as well as *in vivo* in chimeric embryos (Smith, 2001). To preserve stem cell identity, pluripotency factors have to be expressed (Nichols et al., 1998, Hochedlinger et al., 2005, Hough et al., 2006, Takahashi and Yamanaka, 2006). However, ESCs express these pluripotency factors heterogeneously in culture (Chambers et al., 2007, Toyooka et al., 2008, Niwa et al., 2009).

This could be due to undefined culturing conditions of ESCs in serum. Since 2008, it is possible to culture ESCs in serum-free conditions in the presence of two inhibitors of the FGF and GSK signalling pathways (2i inhibitors, PD0325901 and CHIR99021) (Ying et al., 2008, Kunath et al., 2007, Wray et al., 2010). ESCs grown in 2i culture conditions adopt a ground state of pluripotency, with loss of mosaic expression levels of pluripotency factors (Wray et al., 2010, Ficz et al., 2013, Marks et al., 2012). A key feature of 2i grown ESCs compared to serum grown ESCs is the global hypomethylation of the whole genome. However, IAPs as well as major satellites partially escape this demethylation (Ficz et al., 2013). Additionally, global histone rearrangements have

been found in the more naïve state ESCs in comparison to serum grown ESCs (Marks et al., 2012, von Meyenn et al., 2016). Earlier studies have shown that this global hypomethylation leads to transcriptional activation of IAPs, MERVL as well as LINE elements, and subsequent resilencing of these TEs by histone redistribution (Walter et al., 2016).

I aim to investigate whether next to histone marks, small RNAs silence TE classes during global hypomethylation during the transition from serum to 2i grown ESCs.

(2) Four transcription factors - *cMyc*, *Klf4*, *Oct4* and *Sox2* - have been identified to reprogram somatic cells into iPSCs (Takahashi and Yamanaka, 2006, Takahashi et al., 2007, Park et al., 2008). In iPSC generation the epigenetic landscape has to be erased to establish a pluripotent state (Maherali et al., 2007, Mikkelsen et al., 2008, Onder et al., 2012, Polo et al., 2012, Papp and Plath, 2013, Lee et al., 2014a). During the generation of iPSCs there is a wave of demethylation (Milagre et al, in preparation). Additionally, it was shown that in iPSC reprogramming endogenous retroviruses (ERVs) become up-regulated (Wissing et al., 2012, Friedli et al., 2014). I generated small RNA-seq libraries during different stages of reprogramming and mapped them to TE classes to investigate whether any small RNAs could be found to control TEs during this event of global hypomethylation.

(3) The precursor of the future gametes - oocyte and sperm - are primodial germ cells (PGCs). A small number of PGCs (n = 40) is first detectable in the epiplast at embryonic day (E)7.25. Afterwards, PGCs migrate to the genital ridges (E8-10.5) and proliferate thereafter extensively. To establish the prerequisite for pluripotency, PGCs have to reset the epigenetic memory in a phase of global epigenetic reprogramming (von Meyenn and Reik, 2015, Reik and Surani, 2015).

Due to the limited cell number of PGCs *in vivo*, *in vitro* systems to study PGC differentiation was in demand. It has been possible to isolate cells in embryoid bodies (EBs) that express germ cell markers and use them as a proxy for *in vitro* generation of PGCs (Daley, 2007, Saitou and Yamaji, 2010). However, to properly study PGC development more faithful *in vitro* systems were demonstrated more recently (Hayashi et al., 2011). In these PGC-like cells (PGCLC) the start of global demethylation was observed (von Meyenn et al., 2016, *accepted*).

During this global demethylation in PGCs, TEs become transcriptionally activated (Molaro et al., 2014). In this chapter, I perform small RNA-seq of PGCLC and *in vitro* PGCs of male and female PGCs at E13.5, E14.5 and E15.5, to study whether I can observe any small RNAs controlling TEs during PGC development. First, I aim to find piwi interacting small RNAs (piRNAs) in the PGCLC. piRNAs are small RNAs, which control TEs in the male germline (Aravin et al., 2007). Second, I investigate whether endosiRNAs control TEs *in vivo* during PGC development, in the same way as endosiRNAs are important for TE silencing in our *in vitro* demethylation model by deletion of *Dnmt1* in ESCs. This study sheds light on the potential role of endosiRNAs in TE silencing during global hypomethylation.

5.2 Results

5.2.1 small RNAs in ESCs during serum/2i transition

I was interested to learn whether the global hypomethylation, that has been reported in the transition from serum to 2i grown ESCs, could lead to an increase of small RNAs that mapped to TEs.

Earlier studies have shown, that the global demethylation happened 72h after switching ESCs serum medium to serum-free 2i medium. Additionally, certain TE classes become transcriptionally up-regulated during serum to 2i transition (Walter et al., 2016). Therefore, I sampled RNA at 24h, 48h and 72h after exchanging the culture conditions, from serum to serum-free 2i medium, and prepared small RNA-seq libraries. I analysed the small RNA composition of the small RNA-seq libraries of the three timepoints after adding 2i medium to the ESC using the piPipes small RNA-seq pipeline (Han et al., 2015). Most of the small RNAs were miRNAs (64% miRNAs in 24h 2i, 55% miRNAs in 48h 2i and 42% miRNAs in 72h 2i) (Figure 5.1).

I observed a compelling increase of small RNAs that mapped to repeats, upon transition of serum to 2i culturing conditions - 7% in 24h 2i, 11% in 48h 2i and 13% in 72h 2i (Figure 5.1). As a consequence, I characterised the small RNA fraction that mapped to repeats further.



Figure 5.1: Classification of small RNAs mapped to 2i grown ESCs. Pie chart of small RNA-seq libraries mapped to 2i grown ESCs for 24h (left), 48h (middle) and 72h (right). miRNAs (black), 5'UTR bound small RNAs (light green), 3'UTR bound smallRNAs (yellow), intron bound small RNAs (dark blue), repeats bound small RNAs (dark green), rRNA and tRNA (grey), unannotated small RNAs (white).



Figure 5.2: 21nt small RNAs in 2i grown ESCs. Bar graph of size distribution of small RNA-seq libraries depleted of small RNAs that mapped to miRNAs in 2i grown ESCs at 24h (yellow), 48h (dark green) and 72h (light green) in comparison to WT ESCs (grey), day 9 *Dnmt1* KO ESCs (light blue).

I studied the size distribution of the small RNA-seq libraries, without the miRNA

fraction, in ESCs at 24h, 48h and 72h after changing into 2i culturing conditions, to analyse whether any distinct size fraction of small RNAs would be present. Indeed, at 24h culturing the ESCs in 2i medium I detected similarly high amounts of 21nt small RNAs as in *Dnmt1* KO ESCs at 9 days after KO induction, while the WT ESCs had much less 21nt small RNAs (Figure 5.2). After culturing the ESCs for longer time in 2i - 48h and 72h - the 21nt peak went down.

This is an intriguing finding, as the global demethylation in the transition from serum to 2i has been reported to happen from 48h 2i onwards (von Meyenn et al., 2016, Ficz et al., 2013).



Figure 5.3: Small RNAs map to SINE elements in 2i medium. Bar graph of proportion of small RNAs without the miRNA fraction that mapped to DNA transposons (black), SINEs (dark grey), LINEs (grey), LTRs (light grey).

I wanted to know whether the 21nt small RNAs were enriched at a specific TE classes. Therefore, I mapped the 21-24nt fraction of small RNAs, depleted of miRNAs, to LINE, SINE, LTR and DNA transposons. I measured an increase in small RNAs that mapped to SINE elements in 2i grown ESCs. SINE bound siRNAs have been found in mouse ESCs, nonetheless no increase of SINE transcription was observed in earlier studies in 2i ESCs (Babiarz et al., 2008, Walter et al., 2016).

Despite the small RNAs that mapped to SINE elements, the amount of small RNAs that mapped to the other TE families were not change in 2i grown ESCs compared to

serum grown ESCs.

This result differs from the response to acute demethylation through *Dnmt1* KO ESCs, where a bigger proportion of small RNAs mapped to ERVs.

To understand whether a change in miRNA profiles was found after the change from serum to 2i medium, I analysed the small RNAs that mapped uniquely to the genome. Reassuringly, I found mostly miRNAs with a size of 21-22nt in the genome in the small RNA-seq quality control plot of Seqmonk (Figure 5.4).



Figure 5.4: miRNAs are the most abundant class of small RNAs mapped to the genome in 2i grown ESCs. Size distribution and classification of small RNAs mapped uniquely to the mouse genome of 2i grown ESCs for 24h, 48h or 72h. miRNAs (grey), miscRNA (red), rRNA (green), tRNA (blue), snoRNA (orange), snRNA(violet).

I compared the expression of known miRNAs, important in PGC development and endogenously expressed in ESCs, across the serum to 2i transition by analysing the miRNAs mapped to the whole genome using Seqmonk (Medeiros et al., 2011, Takada et al., 2009, Hayashi et al., 2008, Zhao et al., 2014 Hadjimichael et al., 2016).

Then, I carried out a hierarchical clustering of the selected miRNAs across the early stages of 2i transition. The clustering was done in Seqmonk using per probe normalised counts and therefore the heatmap depicts changes in miRNA expression across the different time points. miRNAs in cluster I were high expressed at 24h and 48h 2i but became down-regulated in 72h 2i. Cluster I encompassed the miR302 family, which is known to control ESC self renewal by controlling the cell cycle. miRNAs in cluster II were highly expressed in 24h 2i, but down-regulated thereafter. This cluster contained miRNAs, highly expressed in ESCs, like miR200, miR23a, miR421. The third cluster (III) was low expressed at both 24 and 48h 2i, but miRNAs increased expression at

72h 2i. Among these miRNAs were miR-let7g - normally expressed in differentiated cells (Figure 5.5).



Figure 5.5: Heatmap of miRNAs expression in 2i grown ESCs. Heatmap of hierachical clustering of miRNAs expressed in ESCs as well as important in PGC development were compared over the time-course of serum to 2i transition. Per probe normalised.

5.2.2 small RNAs in iPSCs reprogramming

In the lab Inês Milagre performed reprogramming of mouse MEFs to iPSCs and assessed the global methylation levels at certain time-points after reprogramming. The global demethylation in iPSCs reprogramming was confirmed on all genic regions. The demethylation went down to 45% with the biggest hypomethylation at passage 12 (data not shown). This global demethylation was also found in all TEs - with resistance to demethylation in IAPs (Seisenberger et al., 2012, Kobayashi et al., 2012) - which opened the possibility for TE transcriptional upregulation and thus another system where small



RNAs could play a role in TE repression (Milagre et al., in preparation).

Figure 5.6: Classification of small RNAs in iPSCs reprogramming. Pie chart of small RNA-seq libraries mapped to MEFs (left), passage 12 (middle) and 17 (right) iPSCs. miRNAs (black), 5'UTR bound small RNAs (light green), 3'UTR bound smallRNAs (yellow), intron bound small RNAs (dark blue), repeats bound small RNAs (dark green), rRNA and tRNA (grey), unannotated small RNAs (white).



Figure 5.7: No specific fraction of siRNAs in early stages of iPSCs reprogramming. Bar graph of size distribution of small RNA-seq libraries without miRNAs, mapped to MEFs (yellow), passage 12 (orange) and passage 17 (green) iPSCs.

For this reason, I produced small RNA-seq libraries at four time-points during reprogramming. For the classification I used the piPipes small RNA-seq pipeline (Han et al., 2015). First of all, I classified the small RNAs. 72% of the small RNAs in MEFs were miRNAs, with only 5% repeat mapped small RNAs. However during reprogramming in passage 12 iPSCs, the miRNA fraction went down to 14% miRNAs, while the fraction of repeats bound small RNAs increased to 28%. In the iPSCs at passage 17 the miR-NAs made up 40% of the total small RNAs with 17% mapping to repeats. I wanted to test whether the small RNAs at passage 12 were siRNAs (Figure 5.6). The intriguing increase of small RNAs that mapped to repeats during iPSCs reprogramming motivated me to study the small RNAs depleted of miRNAs further. However, I only found 22nt small RNAs in MEFs, while during the iPSCs reprogramming no discrete size of small RNAs was detected. I concluded that no siRNAs are mapping to TEs during iPSCs reprogramming at early stages of reprogramming (Figure 5.7).



Figure 5.8: Classification of small RNAs of iPSCs mapped to the genome. Bar graph of size distribution and classification of small RNA-seq libraries at different passages of iPSCs during reprogramming in comparison to MEFs. miRNAs (grey), miscRNA (red), rRNA (green), tRNA (blue), snoRNA (orange), snRNA(violet).

Even though, the miRNA fraction decreased during iPSCs formation, I observed miR-NAs with a mean size of 21-22nt in MEFs, iPSCs passage 12, 17 as well as passage 31 in the small RNA-seq quality control plot of Seqmonk (Figure 5.7) after mapping the small RNA-seq reads to the whole genome. I analysed the expression of miRNAs, known to be important during early mammalian development and essential for ESC pluripotency, throuhgout iPSCs reprograming (Medeiros et al., 2011, Takada et al., 2009, Hayashi et al., 2008, Zhao et al., 2014 Hadjimichael et al., 2016).

I performed a hierarchical clustering with per probe normalisation in Seqmonk. This resulted in a division of miRNAs into into two main cluster: (cluster I) miRNAs high expressed in MEFs and down-regulated thereafter, (cluster II) miRNAs not expressed in MEFs and up-regulated thereafter. Cluster I was reassuringly comprised of the miR-let7 family, which is known to be highly expressed in differentiated tissues but lowly

expressed in pluripotent cells. Cluster II contained the miR290-295 family, which is also known to be important for proper PGC development, as well as the miR200 family which is highly expressed in ESCs (Figure 5.9). Interestingly, the expression of cluster II miRNAs was downregulated again at passage 31.



Figure 5.9: Heatmap of miRNAs expression in iPSCs during reprogramming. Heatmap of hierarchical clustering of miRNAs expressed in ESCs as well as important in PGC development were compared during iPSCs generation.

5.2.3 small RNAs in mouse PGC like cells (PGCLC) and PGCs

piRNAs in vivo and in vitro PGCs

miRNAs as well as piRNAs have been studied in male PGCs at E16.5 *in vivo* PGCs (Hayashi et al., 2008, Molaro et al., 2014). Here, we carried out a comprehensive study

of small RNAs in E13.5 to E15.5 male and female PGCs. Additionally, Ferdinand von Meyenn setup a protocol to study *in vitro* mouse PGC like cells (PGCLC) (von Meyenn et al, 2016, *accepted*) and I generated small RNA-seq libraries of PGCLCs from two different mouse ESC strains - E14 and C57BL/6 (B6) - and compared them to WT E14 ESCs small RNA-seq libraries, as a control.

I found 54% small RNAs that mapped to repeats with only 9% small RNAs mapping to miRNAs in the PGCLC in comparison to mESCs, where 60% of the library was made up of miRNAs and only 15% of small RNAs mapped to repeats (Figure 5.10). The small RNAs of the PGCLC, that mapped to repeats were 24-32nt long and had a thymidine overhang at the 5' ends. Additionally, these small RNAs showed the piRNA profile of a 5'-5' overlap at nucleotide 10 with reads mapping to sense and antisense strands of repeats (Figure 5.11).



Figure 5.10: Classification of small RNAs mapping to *in vitro* PGCs. (A) Pie chart of small RNA-seq libraries mapped to male E15.5 PGCs in comparison to E14 mouse ESCs. miRNAs (black), 5'UTR bound small RNAs (light green), 3'UTR bound smallRNAs (yellow), intron bound small RNAs (dark blue), repeats bound small RNAs (dark green), rRNA and tRNA (grey).

I wanted to compare the PGCLC small RNA-seq libraries to the *in vivo* PGCs. Wendy Dean isolated *in vivo* PGCs of male and female at E13.5, E14.5 and E15.5 and I prepared small RNA-seq libraries of them. In comparison to PGCLC, I found similar classes of small RNAs in the small RNA-seq libraries of male and female *in vivo* PGCs, at each developmental stage assessed. I measured 54% small RNAs mapping to repeats and only 9% mapped to miRNAs (Figure 5.12). The size distribution of all of the *in vivo* PGC small RNAs mapping to repeats, showed that they were comprised of a class of 20-24nt and a class of 24-32nt small RNAs, while in ESCs I only detected a peak of 20-24nt small RNAs and no signature of piRNAs (Figure 5.13A, B).



Figure 5.11: piRNAs from PGCLCs mapped to TEs. (A) Bargraph of nucleotide contribution along the stretch of the small RNAs mapped to repeats, (B) Bargraph of 5'5' overhang of small RNAs mapped to repeats, (C) Bargraph of size distribution of small RNA-seq libraries mapped to TE classes of sense (red) and antisense (blue) reads.



Figure 5.12: Classification of small RNAs mapping to *in vivo* PGCs. (A) Pie chart of small RNA-seq libraries mapped to male E15.5 PGCs in comparison to E14 mouse ESCs. miRNAs (black), 5'UTR bound small RNAs (light green), 3'UTR bound smallRNAs (yellow), intron bound small RNAs (dark blue), repeats bound small RNAs (dark green), rRNA and tRNA (grey).

To analyse whether piRNAs map to active TEs in the mouse germline, I mapped the small RNA-seq data of PGCLCs and *in vivo* PGCs to the consensus sequence of the IAPEZ element. I detected mostly 22nt small RNAs in ESCs (at very low level) and a high number of small RNAs of size 24-29nt mapping to IAPs in PGCLCs. A small number of small RNAs of the size of 22nt could be found in PGCLCs. I mapped the small RNA-seq libraries of *in vivo* PGCs to IAPEZ consensus sequence and found small RNAs mainly of the size of piRNAs mapping to IAPs. Intriguingly, in E13.5 PGCs I found small amount of piRNAs mostly from male PGCs mapping to IAPs, while at

E14.5 mostly female small RNAs were mapping to IAPEZ consensus sequence, while at E15.5 the piRNAs of the male PGCs are getting very enriched over the female piRNAs (Figure 5.14).



Figure 5.13: TE bound small RNAs *in vivo* **PGCs and mouse ESCs.** (A) Bargraph of size distribution of small RNA-seq libraries mapped to the whole genome of ESCs and PGCs, miRNAs/siRNAs (light grey), piRNAs (dark grey), (B) Bargraph of nucleotide contribution along the stretch of the small RNAs mapped to repeats in PGCs and ESCs.

A small number of small RNAs of the size of 22nt could be found in PGCLCs as well as during *in vivo* PGC development. This class of 22nt small RNAs was of interest, as it could represent a class of endosiRNAs acting *in vivo* during PGC development. For this reason, I examined whether these 22nt small RNAs would also map to other TE classes.

endosiRNAs in vivo and in vitro PGCs map to TEs

In order to see whether the 20-24nt small RNAs mapping to all repeats *in vivo* PGCs, were also endosiRNAs, working potentially through a similar mechanism of TE silencing I presented during global hypomethylation in mouse ESCs (Chapter 3 and 4), I


Figure 5.14: IAP bound small RNAs in *in vitro* and *in vivo* PGCs. Bar graph of size distribution of small RNA-seq libraries mapped to the consensus sequence of IAPEZ in ESCs (orange), PGCLCs (turquoise) and E13.5, E14.5, E15.5 male(black) and female (white) *in vivo* PGCs.

only mapped 20-24nt small RNAs to all repeats. While most of these small RNAs were miRNAs, I found 10% repeat bound small RNAs (Figure 5.15A). These small RNAs mapped to both sense and antisense of all repeats and had a peak at 21nt (Figure 5.15B).



Figure 5.15: Contribution of 20-24nt small RNAs mapped to repeats. (A) Pie chart of small RNA-seq with the size of 20-23nt of male E13.5 libraries mapped to all repeats. miRNAs (black), 5'UTR bound small RNAs (light green), 3'UTR bound smallRNAs (yellow), intron bound small RNAs (dark blue), repeats bound small RNAs (dark green), rRNA and tRNA (grey), unannotated small RNAs (white), (B) Bargraph of size distribution of small RNA-seq libraries mapped to TE classes of sense (red) and antisense (blue) reads,



Figure 5.16: In vivo PGC small RNAs mapped to TEs. Bar graph 20-24nt small RNAs of smallRNA-seq data mapped to IAPs, MMERVK10C, MERVL, L1MdGf and ETnERV2 TEs in E13.5 and E14.5 male and female PGCs.

I mapped the 20-24nt small RNA fraction of the *n vivo* E13.5 and E14.5 PGC libraries to all the repeat classes dependent on *Dnmt1* and *Dicer* defined in Chapter 3 and 4. I measured an increase of small RNAs that mapped to IAPs as well as L1MdGf in E13.5 and E14.5 males, while no small RNAs mapped at MERVL, ETnERV2 and MMERVK10C (Figure 5.16).

miRNAs in vivo and in vitro PGCs

To finish the characterisation of small RNAs *in vitro* PGCLC as well as *in vivo* PGCs, I investigated the small RNAs that mapped uniquely to the genome.

I analysed the expression pattern of miRNAs expressed during PGC development, as well as miRNAs which are confined to the ES cell lineage, in the *in vivo* and *in vitro* PGCs small RNA-seq libraries. I performed hierarchical clustering per probe normalised in Seqmonk. Most miRNAs fell into the same cluster. While in *in vitro* PGCs all of the miRNAs were lowly expressed, all of them increased in expression *in vivo* PGCs. Despite this the miRNAs did not form apparent clusters. I could recapitulate the effect of remethylation of male PGCs earlier than female PGCs from E14.5 onwards, as most of the miRNAs were highly expressed in male E13.5 PGCs but already down-regulated at E14.5 and E15.5 male PGCs. miR-let7g was highly expressed in all *in vivo* samples but not in the *in vitro* PGCLCs.



Figure 5.17: Heatmap of miRNAs expression *in vitro* and *in vivo* PGCs. Heatmap of hierachical clustering of miRNAs expressed in ESCs as well as during PGC development in our *in vivo* PGCs cells at E13.5, E14.4 and E15.5 in male as well as female, and *in vitro* PGCLCs.

In conclusion, not in all demethylation systems I found small RNAs mapping to TEs. Although in our demethylation system with *Dnmt1* deletion, as well as during *in vitro* and *in vivo* PGCs and in the serum/2i transition, we detected endosiRNAs next to piRNAs that map to TE elements, in iPSCs we could not find small RNAs of a distinct size that mapped to TEs.

During the miRNA analysis I confirmed the expression of miR290-295 cluster in pluripotent states, while the let-7 class was higher expressed in differentiated cells like MEFs (Zhao et al., 2014, Hadjimichael et al., 2016, Medeiros et al., 2011, Takada et al., 2009, Hayashi et al., 2008, Jia et al., 2013, Bhin et al., 2015).

5.3 Discussion

In the earlier two chapters, I proposed the role of DICER small RNAs in TE silencing during acute demethylation in ESCs.

In this chapter I wanted to test the hypothesis, that the suggested mechanism could also act in other systems of global hypomethylation.

To begin, I investigated the existence of endosiRNAs in a transition from serum to 2i grown ESCs. The global hypomethylation induced through the addition of FGF and GSK signalling inhibitors, leads to transcriptional upregulation of TE classes (Walter et al., 2016). The highest abundance of 21nt small RNAs were present at 24h after 2i transition. However the DNA demethylation is only apparent after 48h after 2i transition(Ficz et al., 2012, von Meyenn et al., 2016). This suggests, that the small RNAs were responding very quickly to the change from serum to 2i conditions in ESCs, irrespective of DNA methylation. To verify whether 21nt RNAs are increased prior to 24h 2i, small RNA libraries from earlier time-points should be sequenced.

Most of the 21nt small RNAs were mapping to SINE elements and while SINE siRNAs have been found before in mESCs, no SINE up-regulation was detected in serum to 2i transition (Barbierz et al., 2008, Walter et al., 2016). This could mean that the SINE RNAs are guarded by endosiRNAs during serum/2i transition, but are dispensable of DNA methylation, similarly to LINE elements in the acute demethylation system. However, the nature of the 21nt small RNA fraction should be verified by AGO2 IP small RNA-seq.

As a second system of global demethylation, I examined stages of mouse iPSCs reprogramming. I measured 28% and 17% small RNAs binding to repeats in passage 12 and 17 iPSCs - respectively. However, I could not observe any distinct size classes of small RNAs and therefore concluded that at the stages we sampled siRNAs were not mapping to TEs in iPSCs.

This could be because I had very limited amount of RNA (10ng of total RNA) of input material for passage 12 iPSCs. Small RNA-seq library preparation for small input libraries are now available, therefore it would be possible to sample these early time-points during iPSCs reprogramming in a future study.

As a third system, I used *in vitro* and *in vivo* PGCs. In the lab we were able to generate PGCLCs as an *in vitro* system for PGC development. I observed 1.4% of small RNAs that mapped to piRNA clusters as well as 55% small RNAs that mapped to repeats and showed piRNA properties, the small RNA contributions of PGCLS was comparable to the amounts of piRNAs and repeat mapped small RNAs *in vivo* male PGC at E15.5. We were able to present the first *in vitro* model to study mammalian piRNAs in PGCLCs (von Meyenn et al, 2016, *accepted*). In addition, I produced small RNAs of the size of 24-29nt length mapping to IAPEZ in the E14.5 female PGCs. We also found the typical piRNA features in the fraction of small RNAs that mapped to repeats.

Further to the TE bound small RNAs, I examined the expression of miRNAs in the different demethylation systems. I found miRNAs expressed in naïve ESCs, in iPSCs as well as in PGC development. In the serum to 2i transition miR200, a highly abundant miRNA in ESCs, as well as mir23a, known to suppress ectoderm and endoderm development, were down-regulated in 72h 2i conditions. In addition to this, I measured an increase of the miR-let7 family - known to be expressed in differentiated cell lines in the 72h 2i samples. This suggests that in the early phases of 2i transition, small RNAs were very respondent and changed expression levels quickly. However, it is very com-

pelling that miRNAs, that are known to prevent differentiation, were down-regulated after culturing the ESCs for 72h in 2i, as 2i grown ESCs have a transcriptome more closely to ICM stage and have been also described as naïve ESCs. To reveal further insights later time points might explain whether the miRNA profile of 2i ESCs is still changing after 72h of culturing ESCs in 2i medium.

miRNA expression in the iPSCs reprogramming verified the reprogramming of the iP-SCs, as miRNAs known to be expressed in differentiated tissues were highly expressed in MEFs but down-regulated thereafter.

Also in PGC development our data supported earlier studies, as we found all the miR-NAs important for PGC development to be expressed. Interestingly, I found that miRNAs at E14.5 and E15.5 in the male PGCs were down-regulated more then the same stage female PGCs, which could be explained by the global remethylation of the male PGCs, while the female PGCs are still hypomethylated.

Noteworthy, the presence of endosiRNAs mapping to TE classes during *in vivo* PGC reprogramming at E14.5 opens the possibility that our *in vitro* system of *Dnmt1* induced global hypomethylation - presented in chapter 3 and 4 - presents a model also for *in vivo* TE regulation.

During global hypomethylation the activation of TE elements presents a risk for the genome integrity. In order to prevent retrotransposition events, molecular mechanisms need to be in place to suppress TE activity. Here, I present a mechanism of endosiR-NAs, that can silence TEs and protect the genome during acute demethylation.

Chapter 6

Conclusions and Outlook

Transposable elements (TEs) refer to DNA sequences which can move from one location in the genome to another; the result of these movements or retrotranspositions can be either detrimental or advantageous to the genome.

On one hand, TEs activity can lead to chromosomal breakage, large-scale genomic rearrangements as well as disruption of protein coding genes and alteration of transcriptional regulatory networks (McCLINTOCK, 1951). In addition, TEs are responsible for 10-15% mutations in mice (Kazazian and Moran, 1998, Maksakova et al., 2006). On the other hand however, TE insertions may also lead to the derivation of proteincoding genes with new and potentially beneficial functions for the host genome. A key example of this is the domestication of the *Rag1* gene, which is invaluable for the adaptive immune system and that evolved through TE endogenisation. Another examples is the MERVL transcriptional network which, mostly driven by "solo" LTRs, seems to be implicated in activation of the transcriptional network at the two-cell stage of mouse embryonic development.

These observations lead to the intriguing question of the apparently mutualistic relationship between host genome and TEs. This is sometimes referred to as a "friend or foe?" relationship: while transcriptional activity of TEs is needed to drive evolution, this activity must be tightly controlled by the host genome, since transposons can also disrupt protein-coding genes. Thus, in order to preserve genome integrity, the cellular machinery must guard the host DNA from the activity of TEs. This presents major challenges for the host, firstly as the genome has to distinguish TEs from protein-coding genes and secondly because a plethora of TE families exist in the genome as they arise through vertical gene transfer. This leads to an "arms race" between genome and TEs: while the genome tries to find ways to control TE activity to preserve genome integrity TEs continuously evolve into different varieties to escape these control mechanisms.

Several epigenetic marks have been implicated in TE control. It has been suggested that DNA methylation actually evolved in order to protect the genome from TE activity and the repressive histone marks H3K9me3 and H3K9me2 are known to be involved in ERV silencing in ESCs. Additionally, small RNAs have been described to play a role in TE control.

While, TE mobilisation is undoubtedly advantageous to TEs, it also presents a disadvantage as this transcriptional activity can activate small RNA-based mechanisms that serve as "guardians of the genome". For example, TEs that retrotranspose into piRNA clusters are silenced through piRNA pathways during spermatogenesis and feed into this small RNA based immune system. miRNAs and endosiRNAs have also been found to be involved in TE silencing in the oocyte (Flemr et al., 2013, Stein et al., 2015).

It is especially during the global hypomethylation stage of epigenetic reprogramming in the mammalian genome, that genome integrity is in danger. Global hypomethylation allows for transcriptional activation of LINEs and ERVs in oocyte and blastocyst and a burst of IAP transcriptional activity has been found to occur during PGC reprogramming (Molaro et al., 2014). As such the genome is particularly vulnerable for retrotransposition events, with the added risk of germline mutations occurring, and thus it is possible that silencing mechanisms other than DNA methylation have to be put in place to control TE mobilisation at this time.

In this thesis I performed a mechanistic study to follow TE activity after acute demethylation with a system that strongly recapitulates *in vivo* epigenetic reprogramming: I used *Dnmt1* conditional KO ESCs and followed demethylation dynamics after KO induction, by WGBS-seq, total RNA-seq, small RNA-seq and ChIP-seq. In chapter 3, I show that *Dnmt1* conditional KO leads to genome wide hypomethylation and transcriptional upregulation of ERV elements. The mouse genome presents the perfect model to study TE regulatory mechanisms, as mouse TEs are younger than predominantly inert human TEs, most of them having integrated in the last 25 Myrs containing around 1000 active LINEs and ERVs. Therefore, the mouse genome is still in the process finding ways to protect itself from the mobility of TEs.

I also uncovered a new potential genomic mechanism to identify and target TE transcriptional activity. I observed that in areas where there was genic pervasive transcription, a TE had integrated and that these transcripts can as such serve as an antisense strand to TE sense transcription and feed into an endosiRNA pathway of TE silencing. During a phase of evolutionary expansion TEs do not have a preference for integration in any genomic location, however, TEs are most commonly domesticated in intergenic regions. This may occur through positive selection in order to prevent TE insertions that could disrupt gene expression. For example it has been shown that ERVs are selected to mostly persist intergenically and that, if ERVs are integrated into genic regions they are manifested in antisense direction to the genes (Medstrand et al., 2002). In chapter 3 I present evidence that the presence of complementary TE sense and antisense transcripts could demonstrate a way by which the genome could control the activity of TEs during TE mobilisation, as the genic regions works to constantly produce antisense strand to the TE's sense transcript. The genic sense strand may thus be regarded as a signal of TE transcription. The TE sense transcript and the genic sense transcript can produce dsRNAs and can feed into an endosiRNA pathway and controling retrotransposition. This model is supported by findings in the yeast genome (Cruz and Houseley, 2014). The produced endosiRNA are going through an AGO2 dependent mechanism and are potentially involved in TE regulation subsequent to their initial transcriptional activation.

In chapter 4, I performed *Dicer* KO in the *Dnmt1* KO ESCs. This mechanistic study allowed me to identify DICER dependent pathways that are essential for transcriptional resilencing of certain TE classes. Interestingly, the TEs which were dependent upon DICER activity were IAP elements, which are one of the most active classes of TEs in the mouse germline.

I also identified an apparently cooperative effect in the defence mechanisms working to silence TE elements. I found IAPs and ETn elements, which were largely dependent on DNMT1 as shown in chapter 3, to be transcriptionally silenced in the double KO of *Dicer/Dnmt1* and, as such, by an unknown mechanism. Subsequently, analysis of ChIP-seq data showed an enrichment of H3K27me3 at these sites upon double KO of *Dicer/Dnmt1*. This is a very intriguing finding, as it could mean that PRC mediated pathways of H3K27me3 deposition may be activated, to preserve TE silencing, upon global hypomethylation in the absence of small RNAs. Future research, possibly with histone methyltransferase knockouts and knockouts of players of the PRC complexes, would allow further insight into this.

Additionally, I found indications that miRNAs might play a role in IAP silencing. The miRNA in the intron of *Dnmt1* is only conserved in the mouse genome and thus has potentially evolved in conjunction with the murine-lineage-specific-IAP elements and may play a role in the control of IAPs. miRNAs might play similar roles to silence young TE elements in other lineages. Future experiments involving depletion of this miRNA will allow more insight into this intriguing finding.

In chapter 5, I looked for small RNA based silencing mechanisms that control TEs during the three incidences of global hypomethylation: serum to 2i culturing of ESCs, iPSC reprogramming and *in vitro* and *in vivo* PGCs.

In the serum to 2i transition of ESCs I found increased small RNAs binding to SINE elements. This is a very interesting preliminary result and could shed light onto other systems and other TE classes that may become silenced through endosiRNA activity.

Knowledge of TE activity is particularly important in iPSCs as it could result in deleterious effects on the genome and iPSCs are used for regenerative medicine. In chapter 5 I assessed endosiRNAs as a potential mean to control mobility of TEs during iPSC reprogramming. I was unable to find any small RNAs that control TE activity in iPSCs at passage 12, 17 and 31, however future experiments should concentrate on additional time points during iPSC reprogramming, to clarify this result. Intriguingly, *in vivo*, endosiRNAs were found to bind to TE classes during PGC development. As these small RNAs had characteristics of endosiRNA they suggest the *in vitro* mechanism I uncovered in mouse ESCs upon global hypomethylation to potentially be active during PGC development in the germline.

Finally, I was able to detect piRNAs in *in vitro* PGCs, which presents the possibility to mechanistically study those small RNAs and their putative role in control of TEs in greater detail.

In summary, I have presented endosiRNAs to be guardians of the genome, they act to silence TEs by restricting TE expression during acute hypomethylation conditions. This study has also suggested a possible connection between small RNA pathways and histone modifications working in conjunction to keep TEs silenced. My research highlights young classes of TEs, which are still retrotransposition active, to be targeted by several histone modifications and endosiRNA pathways.

The different ways of TEs silencing presented in this study indicate the challenge the host genome has to preserve genome integrity by preventing TE retrotransposition events. Several ingenious silencing mechanisms have had to evolve and must be engaged at different stages during early development to allow sufficient protection of the germline DNA from TEs.

Chapter 6. Conclusions and Outlook

Bibliography

- Aapola, U., R. Lyle, K. Krohn, S. E. Antonarakis, and P. Peterson 2001. Isolation and initial characterization of the mouse Dnmt3l gene. *Cytogenetics* and Cell Genetics, 92(1-2):122–126.
- Ahl, V., H. Keller, S. Schmidt, and O. Weichenrieder 2015. Retrotransposition and Crystal Structure of an Alu RNP in the Ribosome-Stalling Conformation. *Molecular Cell*, 60(5):715–727.
- Allfrey, V. G., R. Faulkner, and A. E. Mirsky 1964. ACETYLATION AND METHYLATION OF HISTONES AND THEIR POS-SIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS*. Proceedings of the National Academy of Sciences of the United States of America, 51(5):786–794.
- Allis, C. D. and T. Jenuwein
- 2016. The molecular hallmarks of epigenetic control. *Nature Reviews Genetics*, 17(8):487–500.
- Ameres, S. L. and P. D. Zamore
- 2013. Diversifying microRNA sequence and function. Nature Reviews Molecular Cell Biology, 14(8):475–488.
- Ancelin, K., U. C. Lange, P. Hajkova, R. Schneider, A. J. Bannister, T. Kouzarides, and M. A. Surani

2006. Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. *Nature Cell Biology*, 8(6):623–630.

Anderson, R., T. K. Copeland, H. Schöler, J. Heasman, and C. Wylie 2000. The onset of germ cell migration in the mouse embryo. *Mechanisms of Development*, 91(1-2):61–68.

- Angermueller, C., S. J. Clark, H. J. Lee, I. C. Macaulay, M. J. Teng, T. X. Hu, F. Krueger, S. A. Smallwood, C. P. Ponting, T. Voet, G. Kelsey, O. Stegle, and W. Reik
 - 2016. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nature Methods*, 13(3):229–232.
- Anxolabéhère, D., M. G. Kidwell, and G. Periquet
- 1988. Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of Drosophila melanogaster by mobile P elements. *Molecular Biology and Evolution*, 5(3):252–269.
- Arand, J., D. Spieler, T. Karius, M. R. Branco, D. Meilinger, A. Meissner, T. Jenuwein,
 G. Xu, H. Leonhardt, V. Wolf, and J. Walter
 2012. In Vivo Control of CpG and Non-CpG DNA Methylation by DNA Methyltransferases. *PLoS Genetics*, 8(6).
- Aravin, A., D. Gaidatzis, S. Pfeffer, M. Lagos-Quintana, P. Landgraf, N. Iovino,
 P. Morris, M. J. Brownstein, S. Kuramochi-Miyagawa, T. Nakano, M. Chien, J. J.
 Russo, J. Ju, R. Sheridan, C. Sander, M. Zavolan, and T. Tuschl
 2006. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature*, 442(7099):203–207.
- Aravin, A. A., R. Sachidanandam, D. Bourc'his, C. Schaefer, D. Pezic, K. F. Toth,T. Bestor, and G. J. Hannon2008. A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo
- DNA Methylation in Mice. Molecular Cell, 31(6):785–799.Aravin, A. A., R. Sachidanandam, A. Girard, K. Fejes-Toth, and G. J. Hannon2007. Developmentally Regulated piRNA Clusters Implicate MILI in Transposon
- Arita, K., M. Ariyoshi, H. Tochio, Y. Nakamura, and M. Shirakawa 2008. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a baseflipping mechanism. *Nature*, 455(7214):818–821.
- Arnaudo, A. M. and B. A. Garcia

Control. Science, 316(5825):744-747.

2013. Proteomic characterization of novel histone post-translational modifications. Epigenetics & Chromatin, 6:24. Avery, O. T., C. M. MacLeod, and M. McCarty

- 1944. STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUC-ING TRANSFORMATION OF PNEUMOCOCCAL TYPES. *The Journal of Experimental Medicine*, 79(2):137–158.
- Avvakumov, G. V., J. R. Walker, S. Xue, Y. Li, S. Duan, C. Bronner, C. H. Arrowsmith, and S. Dhe-Paganon

2008. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature*, 455(7214):822–825.

Azuara, V.

2006. Chromatin signatures of pluripotent cell lines. Nat. Cell Biol., 8:532–538.

Babiarz, J. E. and R. Blelloch

2009. Small RNAs – their biogenesis, regulation and function in embryonic stem cells. In *StemBook*. Cambridge (MA): Harvard Stem Cell Institute.

Babiarz, J. E., J. G. Ruby, Y. Wang, D. P. Bartel, and R. Blelloch

2008. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes & Development*, 22(20):2773–2785.

Babushok, D. V. and H. H. Kazazian

2007. Progress in understanding the biology of the human mutagen LINE-1. *Human Mutation*, 28(6):527–539.

- Babushok, D. V., E. M. Ostertag, C. E. Courtney, J. M. Choi, and H. H. Kazazian 2006. L1 integration in a transgenic mouse model. *Genome Research*, 16(2):240–250.
- Baek, D., J. Villén, C. Shin, F. D. Camargo, S. P. Gygi, and D. P. Bartel 2008. The impact of microRNAs on protein output. *Nature*, 455(7209):64–71.

Bagci, H. and A. G. Fisher

2013. DNA Demethylation in Pluripotency and Reprogramming: The Role of Tet Proteins and Cell Division. *Cell Stem Cell*, 13(3):265–269.

Bailey, J. A., G. Liu, and E. E. Eichler

2003. An Alu transposition model for the origin and expansion of human segmental duplications. *American Journal of Human Genetics*, 73(4):823–834.

Bannister, A. J.

2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*, 410:120–124.

Bannister, A. J. and T. Kouzarides

2011. Regulation of chromatin by histone modifications. *Cell Research*, 21(3):381–395.

Barau, J., A. Teissandier, N. Zamudio, S. Roy, V. Nalesso, Y. Hérault, F. Guillou, and D. Bourc'his

2016. The DNA methyltransferase DNMT3c protects male germ cells from transposon activity. *Science (New York, N.Y.)*, 354(6314):909–912.

Barski, A.

2007. High-resolution profiling of histone methylations in the human genome. *Cell*, 129:823–837.

Bartel, D. P.

2004. MicroRNAs. Cell, 116(2):281–297.

Bartel, D. P.

2009. MicroRNAs: Target Recognition and Regulatory Functions. *Cell*, 136(2):215–233.

Bartolomei, M. S.

2009. Genomic imprinting: employing and avoiding epigenetic processes. Genes & Development, 23(18):2124-2133.

Baubec, T., D. F. Colombo, C. Wirbelauer, J. Schmidt, L. Burger, A. R. Krebs, A. Akalin, and D. Schübeler

2015. Genomic profiling of DNA methyltransferases reveals a role for DNMT3b in genic methylation. *Nature*, 520(7546):243–247.

- Baust, C., L. Gagnier, G. J. Baillie, M. J. Harris, D. M. Juriloff, and D. L. Mager 2003. Structure and Expression of Mobile ETnII Retroelements and Their Coding-Competent MusD Relatives in the Mouse. *Journal of Virology*, 77(21):11448–11458.
- Benetti, R., S. Gonzalo, I. Jaco, P. Muñoz, S. Gonzalez, S. Schoeftner, E. Murchison, T. Andl, T. Chen, P. Klatt, E. Li, M. Serrano, S. Millar, G. Hannon, and M. A. Blasco
 - 2008. A mammalian microRNA cluster controls DNA methylation and telomere

recombination via Rbl2-dependent regulation of DNA methyltransferases. *Nature* structural & molecular biology, 15(3):268–279.

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

2006a. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell*, 125(2):315–326.

- Bernstein, E., A. A. Caudy, S. M. Hammond, and G. J. Hannon 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, 409(6818):363–366.
- Bernstein, E., E. M. Duncan, O. Masui, J. Gil, E. Heard, and C. D. Allis 2006b. Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Molecular and Cellular Biol*ogy, 26(7):2560–2569.
- Bernstein, E., S. Y. Kim, M. A. Carmell, E. P. Murchison, H. Alcorn, M. Z. Li, A. A. Mills, S. J. Elledge, K. V. Anderson, and G. J. Hannon 2003. Dicer is essential for mouse development. *Nature Genetics*, 35(3):287–287.
- Bestor, T., A. Laudano, R. Mattaliano, and V. Ingram 1988. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. *Journal of Molecular Biology*, 203(4):971–983.

- 2000. The DNA methyltransferases of mammals. *Human Molecular Genetics*, 9(16):2395–2402.
- Bestor, T. H. and V. M. Ingram
- 1983. Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proc. Natl Acad. Sci. USA*, 80:5559–5563.
- Bhin, J., H.-S. Jeong, J. S. Kim, J. O. Shin, K. S. Hong, H.-S. Jung, C. Kim, D. Hwang, and K.-S. Kim

2015. PGC-Enriched miRNAs Control Germ Cell Development. *Molecules and Cells*, 38(10):895–903.

Bestor, T. H.

- Billy, E., V. Brondani, H. Zhang, U. Müller, and W. Filipowicz 2001. Specific interference with gene expression induced by long, double-stranded RNA in mouse embryonal teratocarcinoma cell lines. *Proceedings of the National Academy of Sciences of the United States of America*, 98(25):14428–14433.
- Bird, A., M. Taggart, M. Frommer, O. J. Miller, and D. Macleod 1985. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*, 40(1):91–99.
- Biterge, B. and R. Schneider
 - 2014. Histone variants: key players of chromatin. *Cell and Tissue Research*, 356(3):457–466.
- Boeke, J. D. and J. P. Stoye

1997. Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements. In *Retroviruses*, J. M. Coffin, S. H. Hughes, and H. E. Varmus, eds. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

- Bohnsack, M. T., K. Czaplinski, and D. Gorlich 2004. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. RNA (New York, N.Y.), 10(2):185–191.
- Borgel, J., S. Guibert, Y. Li, H. Chiba, D. Schübeler, H. Sasaki, T. Forné, and M. Weber 2010. Targets and dynamics of promoter DNA methylation during early mouse development. *Nature Genetics*, 42(12):1093–1100.
- Bostick, M., J. K. Kim, P.-O. Estève, A. Clark, S. Pradhan, and S. E. Jacobsen 2007. UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science*, 317(5845):1760–1764.
- Bourc'his, D. and T. H. Bestor
 - 2004. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3l. *Nature*, 431(7004):96–99.

Bourc'his, D., G.-L. Xu, C.-S. Lin, B. Bollman, and T. H. Bestor

2001. Dnmt3l and the Establishment of Maternal Genomic Imprints. *Science*, 294(5551):2536–2539.

Boyer, L. A.

2006. Polycomb complexes repress developmental regulators in murine embryonic

stem cells. Nature, 441:349–353.

- Boyer, L. A., T. I. Lee, M. F. Cole, S. E. Johnstone, S. S. Levine, J. P. Zucker, M. G. Guenther, R. M. Kumar, H. L. Murray, R. G. Jenner, D. K. Gifford, D. A. Melton, R. Jaenisch, and R. A. Young 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 122(6):947–956.
- Bracken, A. P., N. Dietrich, D. Pasini, K. H. Hansen, and K. Helin 2006. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes & Development*, 20(9):1123–1136.
- Brady, T., Y. N. Lee, K. Ronen, N. Malani, C. C. Berry, P. D. Bieniasz, and F. D.
 Bushman
 2009. Integration target site selection by a resurrected human endogenous retrovirus.
 Genes & Development, 23(5):633-642.
- Brannan, C. I., E. C. Dees, R. S. Ingram, and S. M. Tilghman 1990. The product of the H19 gene may function as an RNA. *Molecular and Cellular Biology*, 10(1):28–36.
- Brennecke, J., A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, and G. J. Hannon

2007. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell*, 128(6):1089–1103.

- Brinkman, A. B., H. Gu, S. J. J. Bartels, Y. Zhang, F. Matarese, F. Simmer, H. Marks,
 C. Bock, A. Gnirke, A. Meissner, and H. G. Stunnenberg
 2012. Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Research*, 22(6):1128–1138.
- Bromham, L., F. Clark, and J. J. McKee
- 2001. Discovery of a novel murine type C retrovirus by data mining. *Journal of Virology*, 75(6):3053–3057.
- Brouha, B., J. Schustak, R. M. Badge, S. Lutz-Prigge, A. H. Farley, J. V. Moran, and H. H. Kazazian

2003. Hot L1s account for the bulk of retrotransposition in the human population. Proceedings of the National Academy of Sciences of the United States of America, 100(9):5280-5285.

- Brown, C. J., A. Ballabio, J. L. Rupert, R. G. Lafreniere, M. Grompe, R. Tonlorenzi, and H. F. Willard
 - 1991. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*, 349(6304):38–44.
- Brown, S. W.

1966. Heterochromatin. Science (New York, N.Y.), 151(3709):417-425.

Brownell, J. E.

1996. Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell*, 84:843–851.

- Bruno, I. G., R. Karam, L. Huang, A. Bhardwaj, C. H. Lou, E. Y. Shum, H.-W. Song,
 M. A. Corbett, W. D. Gifford, J. Gecz, S. L. Pfaff, and M. F. Wilkinson
 2011. Identification of a MicroRNA that Activates Gene Expression by Repressing
 Nonsense-Mediated RNA Decay. *Molecular Cell*, 42(4):500–510.
- Bulut-Karslioglu, A., I. De La Rosa-Velázquez, F. Ramirez, M. Barenboim, M. Onishi-Seebacher, J. Arand, C. Galán, G. Winter, B. Engist, B. Gerle, R. O'Sullivan, J. A. Martens, J. Walter, T. Manke, M. Lachner, and T. Jenuwein
 2014. Suv39h-Dependent H3k9me3 Marks Intact Retrotransposons and Silences LINE Elements in Mouse Embryonic Stem Cells. *Molecular Cell*, 55(2):277–290.
- Bénit, L., N. De Parseval, J. F. Casella, I. Callebaut, A. Cordonnier, and T. Heidmann 1997. Cloning of a new murine endogenous retrovirus, MuERV-L, with strong similarity to the human HERV-L element and with a gag coding sequence closely related to the Fv1 restriction gene. *Journal of Virology*, 71(7):5652–5657.
- Callinan, P. A. and M. A. Batzer
- 2006. Retrotransposable elements and human disease. *Genome Dynamics*, 1:104–115.
- Cammas, F., M. Mark, P. Dollé, A. Dierich, P. Chambon, and R. Losson 2000. Mice lacking the transcriptional corepressor TIF1beta are defective in early postimplantation development. *Development (Cambridge, England)*, 127(13):2955– 2963.
- Cao, F., X. Li, S. Hiew, H. Brady, Y. Liu, and Y. Dou

2009. Dicer independent small RNAs associate with telomeric heterochromatin. RNA, 15(7):1274–1281.

Cao, R.

2002. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*, 298:1039–1043.

Capy, P., R. Vitalis, T. Langin, D. Higuet, and C. Bazin

- 1996. Relationships between transposable elements based upon the integrasetransposase domains: Is there a common ancestor? *Journal of Molecular Evolution*, 42(3):359–368.
- Carmell, M. A., A. Girard, H. J. van de Kant, D. Bourc'his, T. H. Bestor, D. G. de Rooij, and G. J. Hannon 2007. MIWI2 Is Essential for Spermatogenesis and Repression of Transposons in the Mouse Male Germline. *Developmental Cell*, 12(4):503–514.
- Castañeda, J., P. Genzor, G. W. v. d. Heijden, A. Sarkeshik, J. R. Yates, N. T. Ingolia, and A. Bortvin

2014. Reduced pachytene piRNAs and translation underlie spermiogenic arrest in Maelstrom mutant mice. *The EMBO Journal*, 33(18):1999–2019.

- Cavaillé, J., H. Seitz, M. Paulsen, A. C. Ferguson-Smith, and J.-P. Bachellerie 2002. Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Human molecular genetics*, 11(13):1527–1538.
- Chambers, I., J. Silva, D. Colby, J. Nichols, B. Nijmeijer, M. Robertson, J. Vrana, K. Jones, L. Grotewold, and A. Smith

2007. Nanog safeguards pluripotency and mediates germline development. Nature, 450(7173):1230–1234.

Chambers, S. M., C. A. Fasano, E. P. Papapetrou, M. Tomishima, M. Sadelain, and L. Studer

2009. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature Biotechnology*, 27(3):275–280.

Charalambous, M., T. R. Menheniott, W. R. Bennett, S. M. Kelly, G. Dell, L. Dandolo,

and A. Ward

2004. An enhancer element at the Igf2/H19 locus drives gene expression in both imprinted and non-imprinted tissues. *Developmental Biology*, 271(2):488–497.

Charlesworth, B. and D. Charlesworth

1983. The population dynamics of transposable elements. *Genetics Research*, 42(1):1–27.

Chen, C., T. Ara, and D. Gautheret

2009. Using Alu elements as polyadenylation sites: A case of retroposon exaptation. Molecular Biology and Evolution, 26(2):327–334.

Chen, T., Y. Ueda, J. E. Dodge, Z. Wang, and E. Li

2003. Establishment and Maintenance of Genomic Methylation Patterns in Mouse
Embryonic Stem Cells by Dnmt3a and Dnmt3b. *Molecular and Cellular Biology*,
23(16):5594–5605.

- Chen, Z.-X., J. R. Mann, C.-L. Hsieh, A. D. Riggs, and F. Chédin 2005. Physical and functional interactions between the human DNMT3l protein and members of the de novo methyltransferase family. *Journal of Cellular Biochemistry*, 95(5):902–917.
- Chin, H. G., P.-O. Estève, M. Pradhan, J. Benner, D. Patnaik, M. F. Carey, and S. Pradhan

2007. Automethylation of G9a and its implication in wider substrate specificity and HP1 binding. *Nucleic Acids Research*, 35(21):7313–7323.

Chinwalla, A. T., L. L. Cook, K. D. Delehaunty, G. A. Fewell, L. A. Fulton, R. S. Fulton, T. A. Graves, L. W. Hillier, E. R. Mardis, J. D. McPherson, T. L. Miner, W. E. Nash, J. O. Nelson, M. N. Nhan, K. H. Pepin, C. S. Pohl, T. C. Ponce, B. Schultz, J. Thompson, E. Trevaskis, R. H. Waterston, M. C. Wendl, R. K. Wilson, S.-P. Yang, P. An, E. Berry, B. Birren, T. Bloom, D. G. Brown, J. Butler, M. Daly, R. David, J. Deri, S. Dodge, K. Foley, D. Gage, S. Gnerre, T. Holzer, D. B. Jaffe, M. Kamal, E. K. Karlsson, C. Kells, A. Kirby, E. J. Kulbokas, E. S. Lander, T. Landers, J. P. Leger, R. Levine, K. Lindblad-Toh, E. Mauceli, J. H. Mayer, M. McCarthy, J. Meldrim, J. P. Mesirov, R. Nicol, C. Nusbaum, S. Seaman, T. Sharpe, A. Sheridan, J. B. Singer, R. Santos, B. Spencer, N. Stange-Thomann, J. P. Vinson, C. M. Wade,

J. Wierzbowski, D. Wyman, M. C. Zody, E. Birney, N. Goldman, A. Kasprzyk, E. Mongin, A. G. Rust, G. Slater, A. Stabenau, A. Ureta-Vidal, S. Whelan, R. Ainscough, J. Attwood, J. Bailey, K. Barlow, S. Beck, J. Burton, M. Clamp, C. Clee, A. Coulson, J. Cuff, V. Curwen, T. Cutts, J. Davies, E. Eyras, D. Grafham, S. Gregory, T. Hubbard, A. Hunt, M. Jones, A. Joy, S. Leonard, C. Lloyd, L. Matthews, S. McLaren, K. McLay, B. Meredith, J. C. Mullikin, Z. Ning, K. Oliver, E. Overton-Larty, R. Plumb, S. Potter, M. Quail, J. Rogers, C. Scott, S. Searle, R. Shownkeen, S. Sims, M. Wall, A. P. West, D. Willey, S. Williams, J. F. Abril, R. Guigó, G. Parra, P. Agarwal, R. Agarwala, D. M. Church, W. Hlavina, D. R. Maglott, V. Sapojnikov, M. Alexandersson, L. Pachter, S. E. Antonarakis, E. T. Dermitzakis, A. Reymond, C. Ucla, R. Baertsch, M. Diekhans, T. S. Furey, A. Hinrichs, F. Hsu, D. Karolchik, W. J. Kent, K. M. Roskin, M. S. Schwartz, C. Sugnet, R. J. Weber, P. Bork, I. Letunic, M. Suyama, D. Torrents, E. M. Zdobnov, M. Botcherby, S. D. Brown, R. D. Campbell, I. Jackson, N. Bray, O. Couronne, I. Dubchak, A. Poliakov, E. M. Rubin, M. R. Brent, P. Flicek, E. Keibler, I. Korf, S. Batalov, C. Bult, W. N. Frankel, P. Carninci, Y. Hayashizaki, J. Kawai, Y. Okazaki, S. Cawley, D. Kulp, R. Wheeler, F. Chiaromonte, F. S. Collins, A. Felsenfeld, M. Guyer, J. Peterson, K. Wetterstrand, R. R. Copley, R. Mott, C. Dewey, N. J. Dickens, R. D. Emes, L. Goodstadt, C. P. Ponting, E. Winter, D. M. Dunn, A. C. v. Niederhausern, R. B. Weiss, S. R. Eddy, L. S. Johnson, T. A. Jones, L. Elnitski, D. L. Kolbe, P. Eswara, W. Miller, M. J. O'Connor, S. Schwartz, R. A. Gibbs, D. M. Muzny, G. Glusman, A. Smit, E. D. Green, R. C. Hardison, S. Yang, D. Haussler, A. Hua, B. A. Roe, R. S. Kucherlapati, K. T. Montgomery, J. Li, M. Li, S. Lucas, B. Ma, W. R. McCombie, M. Morgan, P. Pevzner, G. Tesler, J. Schultz, D. R. Smith, J. Tromp, K. C. Worley, and E. D. Green

2002. Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420(6915):520–562.

Choo

1997. The Centromere.

Chuong, E. B., N. C. Elde, and C. Feschotte

2016. Regulatory activities of transposable elements: from conflicts to benefits.

Nature Reviews Genetics, advance online publication.

Ciccone, D. N., H. Su, S. Hevi, F. Gay, H. Lei, J. Bajko, G. Xu, E. Li, and T. Chen 2009. KDM1b is a histone H3k4 demethylase required to establish maternal genomic imprints. *Nature*, 461(7262):415–418.

Cohen, C. J., W. M. Lock, and D. L. Mager2009. Endogenous retroviral LTRs as promoters for human genes: a critical assess-

Comings, D. E.

ment. Gene, 448(2):105–114.

1980. Arrangement of chromatin in the nucleus. Human Genetics, 53(2):131–143.

- Cooper, S., M. Dienstbier, R. Hassan, L. Schermelleh, J. Sharif, N. P. Blackledge,
 V. De Marco, S. Elderkin, H. Koseki, R. Klose, A. Heger, and N. Brockdorff
 2014. Targeting polycomb to pericentric heterochromatin in embryonic stem cells
 reveals a role for H2ak119u1 in PRC2 recruitment. *Cell Reports*, 7(5):1456–1470.
- Cordonnier, A., J. F. Casella, and T. Heidmann 1995. Isolation of novel human endogenous retrovirus-like elements with foamy virusrelated pol sequence. *Journal of Virology*, 69(9):5890–5897.
- Cornelis, G., C. Vernochet, S. Malicorne, S. Souquere, A. C. Tzika, S. M. Goodman, F. Catzeflis, T. J. Robinson, M. C. Milinkovitch, G. Pierron, O. Heidmann, A. Dupressoir, and T. Heidmann
 - 2014. Retroviral envelope syncytin capture in an ancestrally diverged mammalian clade for placentation in the primitive Afrotherian tenrecs. *Proceedings of the National Academy of Sciences of the United States of America*, 111(41):E4332–4341.
- Cost, G. J., Q. Feng, A. Jacquier, and J. D. Boeke 2002. Human L1 element target-primed reverse transcription in vitro. *The EMBO Journal*, 21(21):5899–5910.

Costa, F. F.

2007. Non-coding RNAs: Lost in translation? Gene, 386(1-2):1-10.

Cox, D. N., A. Chao, J. Baker, L. Chang, D. Qiao, and H. Lin 1998. A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes & Development*, 12(23):3715–3727.

Craig, N. L.

2015. Mobile DNA III. American Society of Microbiology.

Creyghton, M. P., A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young, and R. Jaenisch

2010. Histone H3k27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States* of America, 107(50):21931–21936.

Cruz, C. and J. Houseley

2014. Endogenous RNA interference is driven by copy number. *eLife*, 3.

- Czech, B., C. D. Malone, R. Zhou, A. Stark, C. Schlingeheyde, M. Dus, N. Perrimon, M. Kellis, J. A. Wohlschlegel, R. Sachidanandam, G. J. Hannon, and J. Brennecke 2008. An endogenous small interfering RNA pathway in Drosophila. *Nature*, 453(7196):798–802.
- Dai, Q., A. Ren, J. O. Westholm, A. A. Serganov, D. J. Patel, and E. C. Lai 2013. The BEN domain is a novel sequence-specific DNA-binding domain conserved in neural transcriptional repressors. *Genes & Development*, 27(6):602–614.

Daley, G. Q.

2007. Gametes from Embryonic Stem Cells: A Cup Half Empty or Half Full? *Science*, 316(5823):409–410.

- Dalton, A. J., M. Potter, and R. M. Merwin 1961. Some ultrastructural characteristics of a series of primary and transplanted plasma-cell tumors of the mouse. *Journal of the National Cancer Institute*, 26:1221– 1267.
- Davis, E., F. Caiment, X. Tordoir, J. Cavaillé, A. Ferguson-Smith, N. Cockett, M. Georges, and C. Charlier

2005. RNAi-Mediated Allelic trans-Interaction at the Imprinted Rtl1/Peg11 Locus. Current Biology, 15(8):743–749.

Davis, T. L., J. M. Trasler, S. B. Moss, G. J. Yang, and M. S. Bartolomei 1999. Acquisition of theH19methylation Imprint Occurs Differentially on the Parental Alleles during Spermatogenesis. *Genomics*, 58(1):18–28.

Davis, T. L., G. J. Yang, J. R. McCarrey, and M. S. Bartolomei

- 2000. The H19 methylation imprint is erased and re-established differentially on the parental alleles during male germ cell development. *Human Molecular Genetics*, 9(19):2885–2894.
- De Fazio, S., N. Bartonicek, M. Di Giacomo, C. Abreu-Goodger, A. Sankar, C. Funaya, C. Antony, P. N. Moreira, A. J. Enright, and D. O'Carroll 2011. The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature*, 480(7376):259–263.
- De Harven, E. and C. Friend
 - 1958. Electron microscope study of a cell-free induced leukemia of the mouse: a preliminary report. *The Journal of Biophysical and Biochemical Cytology*, 4(2):151–156.
- de Koning, A. P. J., W. Gu, T. A. Castoe, M. A. Batzer, and D. D. Pollock 2011. Repetitive elements may comprise over two-thirds of the human genome. *PLoS* genetics, 7(12):e1002384.
- de la Rica, L., . Deniz, K. C. L. Cheng, C. D. Todd, C. Cruz, J. Houseley, and M. R. Branco

2016. TET-dependent regulation of retrotransposable elements in mouse embryonic stem cells. *Genome Biology*, 17:234.

Deaton, A. M., S. Webb, A. R. W. Kerr, R. S. Illingworth, J. Guy, R. Andrews, and A. Bird

2011. Cell type–specific DNA methylation at intragenic CpG islands in the immune system. *Genome Research*, 21(7):1074–1086.

DeBerardinis, R. J. and H. H. Kazazian

1999. Analysis of the promoter from an expanding mouse retrotransposon subfamily. Genomics, 56(3):317–323.

del Mar Lorente, M., C. Marcos-Gutiérrez, C. Pérez, J. Schoorlemmer, A. Ramírez, T. Magin, and M. Vidal

2000. Loss- and gain-of-function mutations show a polycomb group function for Ring1a in mice. *Development (Cambridge, England)*, 127(23):5093–5100.

Denli, A. M., B. B. J. Tops, R. H. A. Plasterk, R. F. Ketting, and G. J. Hannon 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature*, 432(7014):231-235.

Derrien, T., R. Johnson, G. Bussotti, A. Tanzer, S. Djebali, H. Tilgner, G. Guernec, D. Martin, A. Merkel, D. G. Knowles, J. Lagarde, L. Veeravalli, X. Ruan, Y. Ruan, T. Lassmann, P. Carninci, J. B. Brown, L. Lipovich, J. M. Gonzalez, M. Thomas, C. A. Davis, R. Shiekhattar, T. R. Gingeras, T. J. Hubbard, C. Notredame, J. Harrow, and R. Guigó

2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Research*, 22(9):1775–1789.

Dewannieux, M., C. Esnault, and T. Heidmann

2003. LINE-mediated retrotransposition of marked Alu sequences. *Nature Genetics*, 35(1):41–48.

Dhayalan, A., A. Rajavelu, P. Rathert, R. Tamas, R. Z. Jurkowska, S. Ragozin, and A. Jeltsch

2010. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. The Journal of Biological Chemistry, 285(34):26114–26120. Ding, S.-W.

2010. RNA-based antiviral immunity. Nature Reviews. Immunology, 10(9):632–644.

- Dixon, J. R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–380.
- Djebali, S., C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G. K. Marinov, J. Khatun, B. A. Williams, C. Zaleski, J. Rozowsky, M. Röder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, N. S. Bar, P. Batut, K. Bell, I. Bell, S. Chakrabortty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, E. Falconnet, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, O. J. Luo, E. Park, K. Persaud, J. B. Preall, P. Ribeca, B. Risk, D. Robyr, M. Sammeth, L. Schaffer, L.-H. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, X. Ruan,

Y. Hayashizaki, J. Harrow, M. Gerstein, T. Hubbard, A. Reymond, S. E. Antonarakis, G. Hannon, M. C. Giddings, Y. Ruan, B. Wold, P. Carninci, R. Guigó, and T. R. Gingeras

2012. Landscape of transcription in human cells. Nature, 489(7414):101–108.

- Dodge, J. E., Y.-K. Kang, H. Beppu, H. Lei, and E. Li 2004. Histone H3-K9 Methyltransferase ESET Is Essential for Early Development. *Molecular and Cellular Biology*, 24(6):2478–2486.
- Dong, K. B., I. A. Maksakova, F. Mohn, D. Leung, R. Appanah, S. Lee, H. W. Yang, L. L. Lam, D. L. Mager, D. Schübeler, M. Tachibana, Y. Shinkai, and M. C. Lorincz 2008. DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *The EMBO journal*, 27(20):2691–2701.
- Doolittle, W. F. and C. Sapienza
- 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature, 284(5757):601-603.
- Doucet, A. J., J. E. Wilusz, T. Miyoshi, Y. Liu, and J. V. Moran 2015. A 3 poly(A) tract is required for LINE-1 retrotransposition. *Molecular cell*, 60(5):728–741.

Druker, R. and E. Whitelaw

- 2004. Retrotransposon-derived elements in the mammalian genome: a potential source of disease. *Journal of Inherited Metabolic Disease*, 27(3):319–330.
- Du, J., L. M. Johnson, S. E. Jacobsen, and D. J. Patel 2015. DNA methylation pathways and their crosstalk with histone methylation. *Nature Reviews Molecular Cell Biology*, 16(9):519–532.
- Ecco, G., M. Cassano, A. Kauzlaric, J. Duc, A. Coluccio, S. Offner, M. Imbeault, H. M. Rowe, P. Turelli, and D. Trono

2016. Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues. *Developmental cell*, 36(6):611–623.

- Edwards, C. A., A. J. Mungall, L. Matthews, E. Ryder, D. J. Gray, A. J. Pask, G. Shaw, J. A. Graves, J. Rogers, I. Dunham, M. B. Renfree, A. C. Ferguson-Smith, and the SAVOIR consortium
 - 2008. The Evolution of the DLK1-DIO3 Imprinted Domain in Mammals. PLoS Biol,

6(6):e135.

Elbarbary, R. A., B. A. Lucas, and L. E. Maquat

2016. Retrotransposons as regulators of gene expression. Science, 351(6274):aac7247.

- Endoh, M., T. A. Endo, T. Endoh, K.-i. Isono, J. Sharif, O. Ohara, T. Toyoda, T. Ito,
 R. Eskeland, W. A. Bickmore, M. Vidal, B. E. Bernstein, and H. Koseki
 2012. Histone H2a Mono-Ubiquitination Is a Crucial Step to Mediate PRC1Dependent Repression of Developmental Genes to Maintain ES Cell Identity. *PLOS Genet*, 8(7):e1002774.
- Esnault, C., J. Maestre, and T. Heidmann 2000. Human LINE retrotransposons generate processed pseudogenes. *Nature Genetics*, 24(4):363–367.
- Estève, P.-O., H. G. Chin, A. Smallwood, G. R. Feehery, O. Gangisetty, A. R. Karpf,
 M. F. Carey, and S. Pradhan
 2006. Direct interaction between DNMT1 and G9a coordinates DNA and histone
 methylation during replication. *Genes & Development*, 20(22):3089–3103.
- Evans, M. J. and M. H. Kaufman 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature*, 292(5819):154–156.
- Fadloun, A., S. Le Gras, B. Jost, C. Ziegler-Birling, H. Takahashi, E. Gorab, P. Carninci, and M.-E. Torres-Padilla

2013. Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nature Structural & Molecular Biology*, 20(3):332–338.

Fatemi, M., A. Hermann, S. Pradhan, and A. Jeltsch

2001. The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA1. *Journal of Molecular Biology*, 309(5):1189–1199.

Fedoroff, N. V.

2012a. McClintock's challenge in the 21st century. *Proceedings of the National* Academy of Sciences of the United States of America, 109(50):20200–20203. 2012b. Transposable Elements, Epigenetics, and Genome Evolution. *Science*, 338(6108):758–767.

Feng, Q., J. V. Moran, H. H. Kazazian, and J. D. Boeke

1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, 87(5):905–916.

Ferguson-Smith, A. C.

2011. Genomic imprinting: the emergence of an epigenetic paradigm. *Nature Reviews Genetics*, 12(8):565–575.

Feschotte, C., N. Jiang, and S. R. Wessler 2002. Plant transposable elements: where genetics meets genomics. *Nature Reviews Genetics*, 3(5):329–341.

Feschotte, C. and E. J. Pritham 2007. DNA Transposons and the Evolution of Eukaryotic Genomes. Annual review of genetics, 41:331–368.

Ficz, G., M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews, and W. Reik 2011. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature*, 473(7347):398–402.

Ficz, G., T. A. Hore, F. Santos, H. J. Lee, W. Dean, J. Arand, F. Krueger, D. Oxley,
Y.-L. Paul, J. Walter, S. J. Cook, S. Andrews, M. R. Branco, and W. Reik
2013. FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to
the epigenetic ground state of pluripotency. *Cell Stem Cell*, 13(3):351–359.

Fischle, W., Y. Wang, and C. D. Allis

2003. Histone and chromatin cross-talk. *Current Opinion in Cell Biology*, 15(2):172–183.

Flemming, W.

1882. Zellsubstanz, Kern und Zelltheilung. Vogel. Google-Books-ID: JCsZA-AAAYAAJ.

Flemr, M., R. Malik, V. Franke, J. Nejepinska, R. Sedlacek, K. Vlahovicek, and P. Svo-

Fedoroff, N. V.

boda

2013. A Retrotransposon-Driven Dicer Isoform Directs Endogenous Small Interfering RNA Production in Mouse Oocytes. *Cell*, 155(4):807–816.

Fodor, B. D.

2006. Jmjd2b antagonizes H3k9 trimethylation at pericentric heterochromatin in mammalian cells. *Genes Dev.*, 20:1557–1562.

- Forman, J. J. and H. A. Coller
- 2010. The code within the code: microRNAs target coding regions. *Cell Cycle* (*Georgetown, Tex.*), 9(8):1533–1541.
- Fort, A., K. Hashimoto, D. Yamada, M. Salimullah, C. A. Keya, A. Saxena, A. Bonetti, I. Voineagu, N. Bertin, A. Kratz, Y. Noro, C.-H. Wong, M. de Hoon, R. Andersson, A. Sandelin, H. Suzuki, C.-L. Wei, H. Koseki, The FANTOM Consortium, Y. Hasegawa, A. R. R. Forrest, and P. Carninci

2014. Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nature Genetics*, 46(6):558–566.

- French, N. S. and J. D. Norton 1997. Structure and functional properties of mouse VL30 retrotransposons. *Biochimica Et Biophysica Acta*, 1352(1):33–47.
- Friedli, M., P. Turelli, A. Kapopoulou, B. Rauwel, N. Castro-Díaz, H. M. Rowe, G. Ecco, C. Unzu, E. Planet, A. Lombardo, B. Mangeat, B. E. Wildhaber, L. Naldini, and D. Trono

2014. Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. *Genome Research*, 24(8):1251–1259.

Friedman, J. R., W. J. Fredericks, D. E. Jensen, D. W. Speicher, X. P. Huang, E. G. Neilson, and F. J. Rauscher

1996. KAP-1, a novel corepressor for the highly conserved KRAB repression domain. Genes & Development, 10(16):2067–2078.

Friedman, R. C., K. K.-H. Farh, C. B. Burge, and D. P. Bartel 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, 19(1):92–105. Frost, R. J. A., F. K. Hamra, J. A. Richardson, X. Qi, R. Bassel-Duby, and E. N. Olson 2010. MOV10l1 is necessary for protection of spermatocytes against retrotransposons by Piwi-interacting RNAs. *Proceedings of the National Academy of Sciences of the* United States of America, 107(26):11847–11852.

- Förstemann, K., Y. Tomari, T. Du, V. V. Vagin, A. M. Denli, D. P. Bratu, C. Klattenhoff, W. E. Theurkauf, and P. D. Zamore
 - 2005. Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS biology*, 3(7):e236.
- Galupa, R. and E. Heard
 - 2015. X-chromosome inactivation: new insights into cis and trans regulation. Current Opinion in Genetics & Development, 31:57–66.
- Gao, Z., J. Zhang, R. Bonasio, F. Strino, A. Sawai, F. Parisi, Y. Kluger, and D. Reinberg

2012. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Molecular Cell*, 45(3):344–356.

- Garcia-Perez, J. L., A. J. Doucet, A. Bucheton, J. V. Moran, and N. Gilbert 2007. Distinct mechanisms for trans-mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase. *Genome Research*, 17(5):602–611.
- Gibbs, R. A., M. L. Metzker, D. M. Muzny, E. J. Sodergren, S. Scherer, G. Scott,
 D. Steffen, K. C. Worley, P. E. Burch, G. Okwuonu, S. Hines, L. Lewis, C. DeRamo, O. Delgado, S. Dugan-Rocha, G. Miner, M. Morgan, A. Hawes, R. Gill,
 Celera, R. A. H. P. Investigator), M. D. Adams, P. G. Amanatides, H. BadenTillson, M. Barnstead, S. Chin, C. A. Evans, S. Ferriera, C. Fosler, A. Glodek,
 Z. Gu, D. Jennings, C. L. Kraft, T. Nguyen, C. M. Pfannkoch, C. Sitter, G. G.
 Sutton, J. C. Venter, T. Woodage, D. S. P. Investigator), H.-M. Lee, E. Gustafson,
 P. Cahill, A. Kana, L. Doucette-Stamm, K. Weinstock, K. Fechtel, R. B. W. P.
 Investigator), D. M. Dunn, E. D. Green, R. W. Blakesley, G. G. Bouffard, P. J. d.

Fu, X.-D.

^{2014.} Non-coding RNA: a new frontier in regulatory biology. *National science review*,1(2):190–204.

J. P. Investigator), K. Osoegawa, B. Zhu, M. M. P. Investigator), J. S. P. Investigator), I. Bosdet, C. Fjell, S. Jones, M. Krzywinski, C. Mathewson, A. Siddiqui, N. Wye, J. McPherson, S. Z. P. Investigator), C. M. Fraser, J. Shetty, S. Shatsman, K. Geer, Y. Chen, S. Abramzon, W. C. Nierman, R. A. G. P. Investigator), P. H. Havlak, R. Chen, K. J. Durbin, A. Egan, Y. Ren, X.-Z. Song, B. Li, Y. Liu, X. Qin, S. Cawley, G. M. W. (Coordinator), K. C. W. O. Coordinator), A. J. Cooney, R. A. Gibbs, L. M. D'Souza, K. Martin, J. Q. Wu, M. L. Gonzalez-Garay, A. R. Jackson, K. J. Kalafus, M. P. McLeod, A. Milosavljevic, D. Virk, A. Volkov, D. A. Wheeler, Z. Zhang, J. A. Bailey, E. E. Eichler, E. Tuzun, E. Birney, E. Mongin, A. Ureta-Vidal, C. Woodwark, E. Zdobnov, P. Bork, M. Suyama, D. Torrents, M. Alexandersson, B. J. Trask, J. M. Young, H. Huang, H. Wang, H. Xing, S. Daniels, D. Gietzen, J. Schmidt, K. Stevens, U. Vitt, J. Wingrove, F. Camara, M. M. Albà, J. F. Abril, R. Guigo, A. Smit, I. Dubchak, E. M. Rubin, O. Couronne, A. Poliakov, N. Hübner, D. Ganten, C. Goesele, O. Hummel, T. Kreitler, Y.-A. Lee, J. Monti, H. Schulz, H. Zimdahl, H. Himmelbauer, H. Lehrach, H. J. J. P. Investigator), S. Bromberg, J. Gullings-Handley, M. I. Jensen-Seaman, A. E. Kwitek, J. Lazar, D. Pasko, P. J. Tonellato, S. Twigger, C. P. P. (Leader, G. a. P. A. Group), J. M. Duarte, S. Rice, L. Goodstadt, S. A. Beatson, R. D. Emes, E. E. Winter, C. Webber, P. Brandt, G. Nyakatura, M. Adetobi, F. Chiaromonte, L. Elnitski, P. Eswara, R. C. Hardison, M. Hou, D. Kolbe, K. Makova, W. Miller, A. Nekrutenko, C. Riemer, S. Schwartz, J. Taylor, S. Yang, Y. Zhang, K. Lindpaintner, T. D. Andrews, M. Caccamo, M. Clamp, L. Clarke, V. Curwen, R. Durbin, E. Eyras, S. M. Searle, G. M. C. (Co-Leader, E. A. Group), S. Batzoglou, M. Brudno, A. Sidow, E. A. Stone, B. A. Payseur, G. Bourque, C. López-Otín, X. S. Puente, K. Chakrabarti, S. Chatterji, C. Dewey, L. Pachter, N. Bray, V. B. Yap, A. Caspi, G. Tesler, P. A. Pevzner, D. H. (Co-Leader, E. A. Group), K. M. Roskin, R. Baertsch, H. Clawson, T. S. Furey, A. S. Hinrichs, D. Karolchik, W. J. Kent, K. R. Rosenbloom, H. Trumbower, M. Weirauch, D. N. Cooper, P. D. Stenson, B. Ma, M. Brent, M. Arumugam, D. Shteynberg, R. R. Copley, M. S. Taylor, H. Riethman, U. Mudunuri, J. Peterson, M. Guyer, A. Felsenfeld, S. Old, S. Mockrin, and F. Collins

2004. Genome sequence of the Brown Norway rat yields insights into mammalian

evolution. Nature, 428(6982):493–521.

Gifford, R., P. Kabat, J. Martin, C. Lynch, and M. Tristem 2005. Evolution and Distribution of Class II-Related Endogenous Retroviruses. *Jour*nal of Virology, 79(10):6478–6486.

- Gifford, W. D., S. L. Pfaff, and T. S. Macfarlan 2013. Transposable elements as genetic regulatory substrates in early development. *Trends in Cell Biology*, 23(5):218–226.
- Gilbert, N., J. M. Bomar, M. Burmeister, and J. V. Moran 2004. Characterization of a mutagenic B1 retrotransposon insertion in the jittery mouse. *Human Mutation*, 24(1):9–13.
- Gilbert, N., S. Lutz-Prigge, and J. V. Moran 2002. Genomic deletions created upon LINE-1 retrotransposition. *Cell*, 110(3):315– 325.
- Ginsburg, M., M. H. Snow, and A. McLaren 1990. Primordial germ cells in the mouse embryo during gastrulation. *Development*, 110(2):521–528.
- Girard, A., R. Sachidanandam, G. J. Hannon, and M. A. Carmell 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*, 442(7099):199–202.
- Goll, M. G. and T. H. Bestor
 - 2005. Eukaryotic cytosine methyltransferases. Annual Review of Biochemistry, 74:481–514.
- Goodier, J. L. and H. H. Kazazian

2008. Retrotransposons Revisited: The Restraint and Rehabilitation of Parasites. Cell, 135(1):23–35.

- Goodier, J. L., E. M. Ostertag, K. Du, and H. H. Kazazian 2001. A novel active L1 retrotransposon subfamily in the mouse. *Genome Research*, 11(10):1677–1685.
- Goodier, J. L., E. M. Ostertag, and H. H. Kazazian 2000. Transduction of 3'-flanking sequences is common in L1 retrotransposition. *Human Molecular Genetics*, 9(4):653–657.

- Gowher, H., K. Liebert, A. Hermann, G. Xu, and A. Jeltsch 2005. Mechanism of Stimulation of Catalytic Activity of Dnmt3a and Dnmt3b DNA-(cytosine-C5)-methyltransferases by Dnmt3l. *Journal of Biological Chemistry*, 280(14):13341–13348.
- Gregory, R. I., K.-p. Yan, G. Amuthan, T. Chendrimada, B. Doratotaj, N. Cooch, and R. Shiekhattar
 - 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature*, 432(7014):235–240.
- Griffiths-Jones, S.
 - 2004. The microRNA Registry. Nucleic Acids Research, 32(suppl 1):D109–D111.
- Grimson, A., K. K.-H. Farh, W. K. Johnston, P. Garrett-Engele, L. P. Lim, and D. P. Bartel
 - 2007. MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. *Molecular Cell*, 27(1):91–105.
- Grishok, A., A. E. Pasquinelli, D. Conte, N. Li, S. Parrish, I. Ha, D. L. Baillie, A. Fire, G. Ruvkun, and C. C. Mello
- 2001. Genes and Mechanisms Related to RNA Interference Regulate Expression of the Small Temporal RNAs that Control C. elegans Developmental Timing. *Cell*, 106(1):23–34.
- Grivna, S. T., E. Beyret, Z. Wang, and H. Lin 2006. A novel class of small RNAs in mouse spermatogenic cells. *Genes & Develop*ment, 20(13):1709–1714.
- Gruenbaum, Y., H. Cedar, and A. Razin 1982. Substrate and sequence specificity of a eukaryotic DNA methylase. *Nature*, 295:620–622.
- Gu, T.-P., F. Guo, H. Yang, H.-P. Wu, G.-F. Xu, W. Liu, Z.-G. Xie, L. Shi, X. He, S.-g. Jin, K. Iqbal, Y. G. Shi, Z. Deng, P. E. Szabó, G. P. Pfeifer, J. Li, and G.-L. Xu
 - 2011. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*, 477(7366):606–610.
- Guelen, L., L. Pagie, E. Brasset, W. Meuleman, M. B. Faza, W. Talhout, B. H. Eussen,

A. de Klein, L. Wessels, W. de Laat, and B. van Steensel

2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, 453(7197):948–951.

- Guenatri, M., D. Bailly, C. Maison, and G. Almouzni 2004. Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *The Journal of Cell Biology*, 166(4):493–505.
- Guibert, S., T. Forné, and M. Weber

2012. Global profiling of DNA methylation erasure in mouse primordial germ cells. Genome Research, 22(4):633–641.

- Guo, H., N. T. Ingolia, J. S. Weissman, and D. P. Bartel 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, 466(7308):835–840.
- Guo, J. U., Y. Su, J. H. Shin, J. Shin, H. Li, B. Xie, C. Zhong, S. Hu, T. Le, G. Fan, H. Zhu, Q. Chang, Y. Gao, G.-l. Ming, and H. Song
 2014. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nature Neuroscience*, 17(2):215–222.
- Guo, J. U., Y. Su, C. Zhong, G.-l. Ming, and H. Song 2011. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell*, 145(3):423–434.
- Ha, M. and V. N. Kim
 - 2014. Regulation of microRNA biogenesis. *Nature Reviews Molecular Cell Biology*, 15(8):509–524.
- Hackett, J. A., R. Sengupta, J. J. Zylicz, K. Murakami, C. Lee, T. A. Down, and M. A. Surani

2013. Germline DNA Demethylation Dynamics and Imprint Erasure Through 5-Hydroxymethylcytosine. *Science*, 339(6118):448–452.

Hackett, J. A. and M. A. Surani

2013. DNA methylation dynamics during the mammalian life cycle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1609).

Hadjimichael, C., C. Nikolaou, J. Papamatheakis, and A. Kretsovali

2016. MicroRNAs for Fine-Tuning of Mouse Embryonic Stem Cell Fate Decision
through Regulation of TGF- Signaling. Stem Cell Reports, 6(3):292–301.

- Hajkova, P., K. Ancelin, T. Waldmann, N. Lacoste, U. C. Lange, F. Cesari, C. Lee, G. Almouzni, R. Schneider, and M. A. Surani
 2008. Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature*, 452(7189):877–881.
- Hajkova, P., S. Erhardt, N. Lane, T. Haaf, O. El-Maarri, W. Reik, J. Walter, and M. A. Surani
- 2002. Epigenetic reprogramming in mouse primordial germ cells. *Mechanisms of Development*, 117(1–2):15–23.
- Hajkova, P., S. J. Jeffries, C. Lee, N. Miller, S. P. Jackson, and M. A. Surani 2010. Genome-Wide Reprogramming in the Mouse Germ Line Entails the Base Excision Repair Pathway. *Science*, 329(5987):78–82.
- Hammond, S. M.

2005. Dicing and slicing. FEBS Letters, 579(26):5822–5829.

- Hammond, S. M., E. Bernstein, D. Beach, and G. J. Hannon 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature*, 404(6775):293–296.
- Han, B. W., W. Wang, P. D. Zamore, and Z. Weng 2015. piPipes: a set of pipelines for piRNA and transposon analysis via small RNAseq, RNA-seq, degradome- and CAGE-seq, ChIP-seq and genomic DNA sequencing. *Bioinformatics*, 31(4):593–595.
- Han, J., Y. Lee, K.-H. Yeom, Y.-K. Kim, H. Jin, and V. N. Kim 2004. The Drosha-DGCR8 complex in primary microRNA processing. *Genes & Development*, 18(24):3016–3027.
- Han, K., J. Lee, T. J. Meyer, P. Remedios, L. Goodwin, and M. A. Batzer 2008. L1 recombination-associated deletions generate human genomic variation. *Proceedings of the National Academy of Sciences of the United States of America*, 105(49):19366–19371.
- Han, K., S. K. Sen, J. Wang, P. A. Callinan, J. Lee, R. Cordaux, P. Liang, and M. A. Batzer

2005. Genomic rearrangements by LINE-1 insertion-mediated deletion in the human

and chimpanzee lineages. Nucleic Acids Research, 33(13):4040–4052.

Havecker, E. R., X. Gao, and D. F. Voytas

2004. The diversity of LTR retrotransposons. *Genome Biology*, 5:225.

Hawkins, R. D., G. C. Hon, L. K. Lee, Q. Ngo, R. Lister, M. Pelizzola, L. E. Edsall, S. Kuan, Y. Luu, S. Klugman, J. Antosiewicz-Bourget, Z. Ye, C. Espinoza, S. Agarwahl, L. Shen, V. Ruotti, W. Wang, R. Stewart, J. A. Thomson, J. R. Ecker, and B. Ren

2010. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell*, 6(5):479–491.

- Hayashi, K., S. M. Chuva de Sousa Lopes, M. Kaneda, F. Tang, P. Hajkova, K. Lao, D. O'Carroll, P. P. Das, A. Tarakhovsky, E. A. Miska, and M. A. Surani 2008. MicroRNA Biogenesis Is Required for Mouse Primordial Germ Cell Development and Spermatogenesis. *PLoS ONE*, 3(3).
- Hayashi, K., H. Ohta, K. Kurimoto, S. Aramaki, and M. Saitou 2011. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell*, 146(4):519–532.
- Hayashi, M., K. Maehara, A. Harada, Y. Semba, K. Kudo, H. Takahashi, S. Oki, C. Meno, K. Ichiyanagi, K. Akashi, and Y. Ohkawa
 2016. Chd5 Regulates MuERV-L/MERVL Expression in Mouse Embryonic Stem Cells Via H3k27me3 Modification and Histone H3.1/H3.2. Journal of Cellular Biochemistry, 117(3):780–792.
- Hayward, B. E., M. Zavanelli, and A. V. Furano 1997. Recombination creates novel L1 (LINE-1) elements in Rattus norvegicus. *Genetics*, 146(2):641–654.
- He, Y.-F., B.-Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C.-X. Song, K. Zhang, C. He, and G.-L. Xu
 2011. Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. Science (New York, N.Y.), 333(6047):1303–1307.

Heintzman, N. D.

2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.*, 39:311–318.

Heitz, E.

1928. Das Heterochromatin der Moose. Jahrb. Wiss. Bot., 69:762–818.

Heitz, E.

1929. Heterochromatin, Chromocentren, Chromomeren. Berichte der Deutschen Botanischen Gesellschaft, 47(4):274–284.

Heitz, E.

1932. Geschlechtschromosomen bei einem Laubmoos. Berichte der Deutschen Botanischen Gesellschaft, 50(4):204–206.

Heras, S. R., S. Macias, M. Plass, N. Fernandez, D. Cano, E. Eyras, J. L. Garcia-Perez, and J. F. Cáceres

2013. The Microprocessor controls the activity of mammalian retrotransposons. Nature Structural & Molecular Biology, 20(10):1173–1181.

Hochedlinger, K., Y. Yamada, C. Beard, and R. Jaenisch 2005. Ectopic Expression of Oct-4 Blocks Progenitor-Cell Differentiation and Causes Dysplasia in Epithelial Tissues. *Cell*, 121(3):465–477.

Holliday, R. and J. E. Pugh

1975. DNA modification mechanisms and gene activity during development. *Science*, 187(4173):226–232.

Holmes, S. E., M. F. Singer, and G. D. Swergold

1992. Studies on p40, the leucine zipper motif-containing protein encoded by the first open reading frame of an active human LINE-1 transposable element. *The Journal of Biological Chemistry*, 267(28):19765–19768.

Hong, L., G. P. Schroth, H. R. Matthews, P. Yau, and E. M. Bradbury 1993. Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA. *The Journal of Biological Chemistry*, 268(1):305–314.

Horwich, M. D., C. Li, C. Matranga, V. Vagin, G. Farley, P. Wang, and P. D. Zamore 2007. The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Current biology: CB*, 17(14):1265–1272.

Hough, S. R., I. Clements, P. J. Welch, and K. A. Wiederholt

2006. Differentiation of Mouse Embryonic Stem Cells after RNA Interference-

Mediated Silencing of OCT4 and Nanog. STEM CELLS, 24(6):1467–1475.

- Howell, C. Y., T. H. Bestor, F. Ding, K. E. Latham, C. Mertineit, J. M. Trasler, and J. R. Chaillet
 - 2001. Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. *Cell*, 104(6):829–838.
- Huang, H., Q. Gao, X. Peng, S.-Y. Choi, K. Sarma, H. Ren, A. J. Morris, and M. A. Frohman
 - 2011. piRNA-associated germline nuage formation and spermatogenesis require MitoPLD profusogenic mitochondrial-surface lipid signaling. *Developmental Cell*, 20(3):376–387.
- Hutnick, L. K., X. Huang, T.-C. Loo, Z. Ma, and G. Fan 2010. Repression of Retrotransposal Elements in Mouse Embryonic Stem Cells Is Primarily Mediated by a DNA Methylation-independent Mechanism. *The Journal* of Biological Chemistry, 285(27):21082–21091.
- Hutvágner, G., J. McLachlan, A. E. Pasquinelli, E. Bálint, T. Tuschl, and P. D. Zamore 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science (New York, N.Y.)*, 293(5531):834–838.

Inoue, A. and Y. Zhang

- 2011. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science (New York, N.Y.)*, 334(6053):194.
- Ip, J., P. Canham, K. H. A. Choo, Y. Inaba, S. A. Jacobs, P. Kalitsis, D. M. Mattiske, J. Ng, R. Saffery, N. C. Wong, L. H. Wong, and J. R. Mann 2012. Normal DNA Methylation Dynamics in DICER1-Deficient Mouse Embryonic Stem Cells. *PLoS Genetics*, 8(9).
- Iqbal, K., S.-G. Jin, G. P. Pfeifer, and P. E. Szabó 2011. Reprogramming of the paternal genome upon fertilization involves genomewide oxidation of 5-methylcytosine. *Proceedings of the National Academy of Sciences*, 108(9):3642–3647.
- Ito, S., L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, and Y. Zhang 2011. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science*, 333(6047):1300–1303.

- Iwasaki, Y. W., M. C. Siomi, and H. Siomi
- 2015. PIWI-Interacting RNA: Its Biogenesis and Functions. Annual Review of Biochemistry, 84:405–433.
- Jacobs, F. M. J., D. Greenberg, N. Nguyen, M. Haeussler, A. D. Ewing, S. Katzman,
 B. Paten, S. R. Salama, and D. Haussler
 2014. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and

SVA/L1 retrotransposons. *Nature*, 516(7530):242-245.

- Jeppesen, P. and B. M. Turner
 - 1993. The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. *Cell*, 74:281–289.
- Jermann, P., L. Hoerner, L. Burger, and D. Schübeler
- 2014. Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation. *Proceedings of the National Academy of Sciences*, 111(33):E3415–E3421.
- Jern, P., G. O. Sperber, and J. Blomberg
 - 2005. Use of Endogenous Retroviral Sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy. *Retrovirology*, 2:50.
- Jia, W., W. Chen, and J. Kang
 - 2013. The Functions of MicroRNAs and Long Non-coding RNAs in Embryonic and
 Induced Pluripotent Stem Cells. *Genomics, Proteomics & Bioinformatics*, 11(5):275–283.

John, R. M.

2017. Imprinted genes and the regulation of placental endocrine function: Pregnancy and beyond. *Placenta*.

Joseph, A., A. R. Mitchell, and O. J. Miller

1989. The organization of the mouse satellite DNA at centromeres. *Experimental Cell Research*, 183(2):494–500.

Jurka, J.

1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proceedings of the National Academy of Sciences of the United States of America*, 94(5):1872–1877. Kaer, K. and M. Speek

2013. Retroelements in human disease. Gene, 518(2):231–241.

Kagiwada, S., K. Kurimoto, T. Hirota, M. Yamaji, and M. Saitou 2013. Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *The EMBO Journal*, 32(3):340–353.

Kajikawa, M. and N. Okada

- 2002. LINEs mobilize SINEs in the eel through a shared 3' sequence. *Cell*, 111(3):433–444.
- Kaneda, M., M. Okano, K. Hata, T. Sado, N. Tsujimoto, E. Li, and H. Sasaki 2004. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature*, 429(6994):900–903.
- Kanellopoulou, C., S. A. Muljo, A. L. Kung, S. Ganesan, R. Drapkin, T. Jenuwein, D. M. Livingston, and K. Rajewsky
 2005. Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes & Development*, 19(4):489–501.

Kapitonov, V. V. and J. Jurka

- 2005. RAG1 Core and V(D)J Recombination Signal Sequences Were Derived from Transib Transposons. *PLOS Biol*, 3(6):e181.
- Kapitonov, V. V., S. Tempel, and J. Jurka

2009. Simple and fast classification of non-LTR retrotransposons based on phylogeny of their RT domain protein sequences. *Gene*, 448(2):207–213.

- Karimi, M. M., P. Goyal, I. A. Maksakova, M. Bilenky, D. Leung, J. X. Tang, Y. Shinkai, D. L. Mager, S. Jones, M. Hirst, and M. C. Lorincz
 2011. DNA methylation and SETDB1/H3k9me3 regulate predominantly distinct sets of genes, retroelements and chimaeric transcripts in mouse ES cells. *Cell stem cell*, 8(6).
- Kato, Y., M. Kaneda, K. Hata, K. Kumaki, M. Hisano, Y. Kohara, M. Okano, E. Li, M. Nozaki, and H. Sasaki

2007. Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Human Molecular Genetics*, 16(19):2272–2280.

- Kawamura, Y., K. Saito, T. Kin, Y. Ono, K. Asai, T. Sunohara, T. N. Okada, M. C. Siomi, and H. Siomi 2008. Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature*, 453(7196):793–797.
- Kazazian, H. H. and J. V. Moran 1998. The impact of L1 retrotransposons on the human genome. *Nature Genetics*,

19(1):19-24.

Kazazian, H. H., C. Wong, H. Youssoufian, A. F. Scott, D. G. Phillips, and S. E. Antonarakis

1988. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature*, 332(6160):164–166.

Kelley, D. R. and J. L. Rinn

2012. Transposable elements reveal a stem cell specific class of long noncoding RNAs. Genome Biology, 13(11):R107.

- Khan, H., A. Smit, and S. Boissinot 2006. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Research*, 16(1):78–87.
- Kidwell, M. G., J. F. Kidwell, and J. A. Sved

1977. Hybrid Dysgenesis in DROSOPHILA MELANOGASTER: A Syndrome of Aberrant Traits Including Mutation, Sterility and Male Recombination. *Genetics*, 86(4):813–833.

- Kigami, D., N. Minami, H. Takayama, and H. Imai 2003. MuERV-L Is One of the Earliest Transcribed Genes in Mouse One-Cell Embryos. *Biology of Reproduction*, 68(2):651–654.
- Kim, J., A. Kollhoff, A. Bergmann, and L. Stubbs

2003. Methylation-sensitive binding of transcription factor YY1 to an insulator sequence within the paternally expressed imprinted gene, Peg3. *Human Molecular Genetics*, 12(3):233–245.

- Kim, S., U. Günesdogan, J. Zylicz, J. Hackett, D. Cougot, S. Bao, C. Lee, S. Dietmann, G. Allen, R. Sengupta, and M. Surani
 - 2014. PRMT5 Protects Genomic Integrity during Global DNA Demethylation in

Primordial Germ Cells and Preimplantation Embryos. *Molecular Cell*, 56(4):564–579.

Klose, R. J.

2006. The transcriptional repressor JHDM3a demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature*, 442:312–316.

Klose, R. J. and A. P. Bird

2006. Genomic DNA methylation: the mark and its mediators. *Trends in Biochemical Sciences*, 31(2):89–97.

Kobayashi, H., T. Sakurai, M. Imai, N. Takahashi, A. Fukuda, O. Yayoi, S. Sato,
K. Nakabayashi, K. Hata, Y. Sotomaru, Y. Suzuki, and T. Kono
2012. Contribution of Intragenic DNA Methylation in Mouse Gametic DNA Methylomes to Establish Oocyte-Specific Heritable Marks. *PLoS Genet*, 8(1):e1002440.

Kobayashi, H., T. Sakurai, F. Miura, M. Imai, K. Mochiduki, E. Yanagisawa,
A. Sakashita, T. Wakai, Y. Suzuki, T. Ito, Y. Matsui, and T. Kono
2013. High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Research*, 23(4):616–627.

Kohli, R. M. and Y. Zhang

2013. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*, 502(7472):472–479.

Kordis, D. and F. Gubensek

1999. Horizontal transfer of non-LTR retrotransposons in vertebrates. *Genetica*, 107(1-3):121–128.

Kornberg, R. D.

1974. Chromatin structure: a repeating unit of histones and DNA. *Science*, 184:868– 871.

Kossel, A.

1884. Ueber einen peptonartigen Bestandtheil des Zellkerns. Zeitschrift für physiologische Chemie, 8(6):511–515.

Kouzarides, T.

2007. Chromatin Modifications and Their Function. Cell, 128(4):693–705.

Kramerov, D. A. and N. S. Vassetzky

2001. Structure and origin of a novel dimeric retroposon B1-diD. Journal of Molecular Evolution, 52(2):137–143.

Kramerov, D. A. and N. S. Vassetzky

2005. Short retroposons in eukaryotic genomes. *International Review of Cytology*, 247:165–221.

Kreahling, J. and B. R. Graveley

2004. The origins and implications of Aluternative splicing. *Trends in genetics: TIG*, 20(1):1–4.

- Krueger, F. and S. R. Andrews 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, 27(11):1571–1572.
- Kuff, E. L. and K. K. Lueders 1988. The intracisternal A-particle gene family: structure and functional aspects. Advances in cancer research, 51:183–276.
- Kulpa, D. A. and J. V. Moran

2006. Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nature Structural & Molecular Biology*, 13(7):655–660.

- Kumar, R., L. DiMenna, N. Schrode, T.-C. Liu, P. Franck, S. Muñoz-Descalzo, A.-K. Hadjantonakis, A. A. Zarrin, J. Chaudhuri, O. Elemento, and T. Evans 2013. AID stabilizes stem-cell phenotype by removing epigenetic memory of pluripotency genes. *Nature*, 500(7460):89–92.
- Kunarso, G., N.-Y. Chia, J. Jeyakani, C. Hwang, X. Lu, Y.-S. Chan, H.-H. Ng, and G. Bourque

2010. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature Genetics*, 42(7):631–634.

- Kunath, T., M. K. Saba-El-Leil, M. Almousailleakh, J. Wray, S. Meloche, and A. Smith 2007. FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development*, 134(16):2895–2902.
- Kuo, M. H.

1996. Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific

lysines. Nature, 383:269–272.

- Kuramochi-Miyagawa, S., T. Watanabe, K. Gotoh, Y. Totoki, A. Toyoda, M. Ikawa, N. Asada, K. Kojima, Y. Yamaguchi, T. W. Ijiri, K. Hata, E. Li, Y. Matsuda, T. Kimura, M. Okabe, Y. Sakaki, H. Sasaki, and T. Nakano
 2008. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes & Development*, 22(7):908–917.
- Labialle, S., V. Marty, M.-L. Bortolin-Cavaillé, M. Hoareau-Osman, J.-P. Pradère,
 P. Valet, P. G. Martin, and J. Cavaillé
 2014. The miR 270/miR 410 eluster at the imprinted Dlk1 Die2 domain controls

2014. The miR-379/miR-410 cluster at the imprinted Dlk1-Dio3 domain controls neonatal metabolic adaptation. *The EMBO Journal*, 33(19):2216–2230.

- Lachner, M., D. O'Carroll, S. Rea, K. Mechtler, and T. Jenuwein 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature*, 410:116–120.
- Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.-F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Navlor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock,

Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. d. l. Bastide, N. Dedhia, H. Blöcker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H.-C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. R. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. A. Smit, E. Stupka, J. Szustakowki, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S.-P. Yang, R.-F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, A. Patrinos, M. J. Morgan, E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Navlor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. StangeThomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.-F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.-F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave,

P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, B. A. Roe, F. Chen, H. Pan, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. d. l. Bastide, N. Dedhia, W. R. McCombie, M. d. l. Bastide, N. Dedhia, H. Blöcker, K. Hornischer, G. Nordsiek, H. Blöcker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H.-C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. R. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. A. Smit, E. Stupka, J. Szustakowki, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S.-P. Yang, R.-F. Yeh, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H.-C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. R. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. A. Smit, E. Stupka, J. Szustakowki, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S.-P. Yang, R.-F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand,

F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, N. Shimizu, K. Kawasaki, S. Minoshima, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, G. A. Evans, M. Athanasiou, R. Schultz, A. Patrinos, A. Patrinos, M. J. Morgan, and M. J. Morgan

2001. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921.

Landthaler, M., A. Yalcin, and T. Tuschl

2004. The Human DiGeorge Syndrome Critical Region Gene 8 and Its D. melanogaster Homolog Are Required for miRNA Biogenesis. *Current Biology*, 14(23):2162–2167.

Langmead, B. and S. L. Salzberg

2012. Fast gapped-read alignment with Bowtie 2. Nature Methods, 9(4):357–359.

Lau, N. C., A. G. Seto, J. Kim, S. Kuramochi-Miyagawa, T. Nakano, D. P. Bartel, and R. E. Kingston

2006. Characterization of the piRNA Complex from Rat Testes. *Science*, 313(5785):363–367.

- Laurent, A. M., J. Puechberty, C. Prades, S. Gimenez, and G. Roizès 1997. Site-specific retrotransposition of L1 elements within human alphoid satellite sequences. *Genomics*, 46(1):127–132.
- Lee, D.-S., J.-Y. Shin, P. D. Tonge, M. C. Puri, S. Lee, H. Park, W.-C. Lee, S. M. I. Hussein, T. Bleazard, J.-Y. Yun, J. Kim, M. Li, N. Cloonan, D. Wood, J. L. Clancy, R. Mosbergen, J.-H. Yi, K.-S. Yang, H. Kim, H. Rhee, C. A. Wells, T. Preiss, S. M. Grimmond, I. M. Rogers, A. Nagy, and J.-S. Seo
 2014a. An epigenomic roadmap to induced pluripotency reveals DNA methylation

as a reprogramming modulator. Nature Communications, 5:5619.

Lee, D. Y., J. J. Hayes, D. Pruss, and A. P. Wolffe

1993a. A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell*, 72(1):73–84.

- Lee, E., R. Iskow, L. Yang, O. Gokcumen, P. Haseley, L. J. Luquette, J. G. Lohr, C. C. Harris, L. Ding, R. K. Wilson, D. A. Wheeler, R. A. Gibbs, R. Kucherlapati, C. Lee, P. V. Kharchenko, P. J. Park, and Cancer Genome Atlas Research Network 2012. Landscape of somatic retrotransposition in human cancers. *Science (New York, N.Y.)*, 337(6097):967–971.
- Lee, H. J., T. A. Hore, and W. Reik
- 2014b. Reprogramming the Methylome: Erasing Memory and Creating Diversity. Cell Stem Cell, 14(6):710–719.
- Lee, J., K. Han, T. J. Meyer, H.-S. Kim, and M. A. Batzer 2008. Chromosomal Inversions between Human and Chimpanzee Lineages Caused by Retrotransposons. *PLOS ONE*, 3(12):e4047.
- Lee, R. C., R. L. Feinbaum, and V. Ambros 1993b. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, 75(5):843–854.
- Lee, Y., C. Ahn, J. Han, H. Choi, J. Kim, J. Yim, J. Lee, P. Provost, O. Rådmark, S. Kim, and V. N. Kim
 - 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature*, 425(6956):415–419.
- Lee, Y., K. Jeon, J.-T. Lee, S. Kim, and V. N. Kim 2002. MicroRNA maturation: stepwise processing and subcellular localization. *The EMBO journal*, 21(17):4663–4670.
- Lee, Y. S., K. Nakahara, J. W. Pham, K. Kim, Z. He, E. J. Sontheimer, and R. W. Carthew

2004. Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell*, 117(1):69–81.

- Leeb, M., D. Pasini, M. Novatchkova, M. Jaritz, K. Helin, and A. Wutz 2010. Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes & Development*, 24(3):265–276.
- 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 Methyla-

tion to Major Satellite Repeats at Pericentric Heterochromatin. *Current Biology*, 13(14):1192–1200.

- Lei, H., S. P. Oh, M. Okano, R. Juttermann, K. A. Goss, R. Jaenisch, and E. Li 1996. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development*, 122(10):3195–3205.
- Leung, D., T. Du, U. Wagner, W. Xie, A. Y. Lee, P. Goyal, Y. Li, K. E. Szulwach,
 P. Jin, M. C. Lorincz, and B. Ren
 2014. Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1. *Proceedings of the National Academy of Sciences*, P. 201322273.
- Lewis, B. P., C. B. Burge, and D. P. Bartel 2005. Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell*, 120(1):15–20.
- Li, E., T. H. Bestor, and R. Jaenisch 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, 69(6):915–926.
- Li, J., M. Kannan, A. L. Trivett, H. Liao, X. Wu, K. Akagi, and D. E. Symer 2014. An antisense promoter in mouse L1 retrotransposon open reading frame-1 initiates expression of diverse fusion transcripts and limits retrotransposition. *Nucleic Acids Research*, 42(7):4546–4562.
- Li, P. W.-L., J. Li, S. L. Timmerman, L. A. Krushel, and S. L. Martin 2006. The dicistronic RNA from the mouse LINE-1 retrotransposon contains an internal ribosome entry site upstream of each ORF: implications for retrotransposition. *Nucleic Acids Research*, 34(3):853–864.
- Li, X., M. Ito, F. Zhou, N. Youngson, X. Zuo, P. Leder, and A. C. Ferguson-Smith 2008. A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. *Developmental Cell*, 15(4):547–557.
- Li, X., C. Roy, X. Dong, E. Bolcun-Filas, J. Wang, B. Han, J. Xu, M. Moore, J. Schimenti, Z. Weng, and P. Zamore

2013a. An Ancient Transcription Factor Initiates the Burst of piRNA Production during Early Meiosis in Mouse Testes. *Molecular Cell*, 50(1):67–81.

- Li, Y., J. Lu, Y. Han, X. Fan, and S.-W. Ding 2013b. RNA Interference Functions as an Antiviral Immunity Mechanism in Mammals. *Science*, 342(6155):231–234.
- Liang, G., J. C. Y. Lin, V. Wei, C. Yoo, J. C. Cheng, C. T. Nguyen, D. J. Weisenberger, G. Egger, D. Takai, F. A. Gonzales, and P. A. Jones
 2004. Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. *Proceedings of the National Academy of Sciences of the United States of America*, 101(19):7357–7362.
- Lingel, A., B. Simon, E. Izaurralde, and M. Sattler 2003. Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. *Nature*, 426(6965):465–469.
- Lister, R., M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q.-M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, and J. R. Ecker 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271):315–322.
- Liu, J., M. A. Carmell, F. V. Rivas, C. G. Marsden, J. M. Thomson, J.-J. Song, S. M. Hammond, L. Joshua-Tor, and G. J. Hannon 2004. Argonaute2 Is the Catalytic Engine of Mammalian RNAi. *Science*, 305(5689):1437–1441.
- Liu, Q., T. A. Rand, S. Kalidas, F. Du, H.-E. Kim, D. P. Smith, and X. Wang 2003. R2d2, a Bridge Between the Initiation and Effector Steps of the Drosophila RNAi Pathway. *Science*, 301(5641):1921–1925.
- Liu, S., J. Brind'Amour, M. M. Karimi, K. Shirane, A. Bogutz, L. Lefebvre, H. Sasaki,
 Y. Shinkai, and M. C. Lorincz
 2014. Setdb1 is required for germline development and silencing of H3k9me3
 - marked endogenous retroviruses in primordial germ cells. *Genes & Development*, 28(18):2041–2055.
- Liu, X., C. Wang, W. Liu, J. Li, C. Li, X. Kou, J. Chen, Y. Zhao, H. Gao, H. Wang, Y. Zhang, Y. Gao, and S. Gao
 - 2016. Distinct features of H3k4me3 and H3k27me3 chromatin domains in pre-

implantation embryos. *Nature*, advance online publication.

- Llave, C., Z. Xie, K. D. Kasschau, and J. C. Carrington 2002. Cleavage of Scarecrow-like mRNA Targets Directed by a Class of Arabidopsis miRNA. *Science*, 297(5589):2053–2056.
- Loh, Y.-H., Q. Wu, J.-L. Chew, V. B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu, K.-Y. Wong, K. W. Sung, C. W. H. Lee, X.-D. Zhao, K.-P. Chiu, L. Lipovich, V. A. Kuznetsov, P. Robson, L. W. Stanton, C.-L. Wei, Y. Ruan, B. Lim, and H.-H. Ng

2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genetics*, 38(4):431–440.

- Love, M. I., W. Huber, and S. Anders 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550.
- Lu, X., F. Sachs, L. Ramsay, P.- . Jacques, J. Göke, G. Bourque, and H.-H. Ng 2014. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nature Structural & Molecular Biology*, 21(4):423–425.
- Luan, D. D., M. H. Korman, J. L. Jakubczak, and T. H. Eickbush 1993. Reverse transcription of R2bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell*, 72(4):595–605.
- Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent, and T. J. Richmond 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, 389:251–260.
- Lund, A. H., M. Duch, and F. S. Pedersen 2000. Selection of functional tRNA primers and primer binding site sequences from a retroviral combinatorial library: identification of new functional tRNA primers in murine leukemia virus replication. *Nucleic Acids Research*, 28(3):791–799.
- Lund, E., S. Güttinger, A. Calado, J. E. Dahlberg, and U. Kutay 2004. Nuclear export of microRNA precursors. *Science (New York, N.Y.)*, 303(5654):95–98.
- Lynch, M. D., A. J. H. Smith, M. De Gobbi, M. Flenley, J. R. Hughes, D. Vernimmen, H. Ayyub, J. A. Sharpe, J. A. Sloane-Stanley, L. Sutherland, S. Meek, T. Burdon,

R. J. Gibbons, D. Garrick, and D. R. Higgs

- 2012. An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *The EMBO Journal*, 31(2):317–329.
- Macfarlan, T. S., W. D. Gifford, S. Driscoll, K. Lettieri, H. M. Rowe, D. Bonanomi, A. Firth, O. Singer, D. Trono, and S. L. Pfaff 2012. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*, 487(7405):57–63.
- Mager, D. L. and J. D. Freeman 2000. Novel Mouse Type D Endogenous Proviruses and ETn Elements Share Long Terminal Repeat and Internal Sequences. *Journal of Virology*, 74(16):7221–7229.
- Mager, D. L. and N. L. Goodchild
- 1989. Homologous recombination between the LTRs of a human retrovirus-like element causes a 5-kb deletion in two siblings. *American Journal of Human Genetics*, 45(6):848–854.
- Mager, D. L. and J. P. Stoye

2015. Mammalian Endogenous Retroviruses. *Microbiology Spectrum*, 3(1).

- Maherali, N., R. Sridharan, W. Xie, J. Utikal, S. Eminli, K. Arnold, M. Stadtfeld,
 R. Yachechko, J. Tchieu, R. Jaenisch, K. Plath, and K. Hochedlinger
 2007. Directly Reprogrammed Fibroblasts Show Global Epigenetic Remodeling and
 Widespread Tissue Contribution. *Cell Stem Cell*, 1(1):55–70.
- Maillard, P. V., C. Ciaudo, A. Marchais, Y. Li, F. Jay, S. W. Ding, and O. Voinnet 2013. Antiviral RNA Interference in Mammalian Cells. *Science*, 342(6155):235–238.
- Maiti, A. and A. C. Drohat
- 2011. Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine POTENTIAL IMPLICATIONS FOR ACTIVE DEMETHYLA-TION OF CpG SITES. *Journal of Biological Chemistry*, 286(41):35334–35338.

Majumdar, S. and D. C. Rio

2015. P transposable elements in Drosophila and other eukaryotic organisms. *Microbiology spectrum*, 3(2).

Maksakova, I. A., P. Goyal, J. Bullwinkel, J. P. Brown, M. Bilenky, D. L. Mager, P. B.

Singh, and M. C. Lorincz

2011. H3k9me3-binding proteins are dispensable for SETDB1/H3k9me3-dependent retroviral silencing. *Epigenetics & Chromatin*, 4(1):12.

Maksakova, I. A., M. T. Romanish, L. Gagnier, C. A. Dunn, L. N. v. d. Lagemaat, and D. L. Mager

2006. Retroviral Elements and Their Hosts: Insertional Mutagenesis in the Mouse Germ Line. *PLOS Genet*, 2(1):e2.

Maksakova, I. A., P. J. Thompson, P. Goyal, S. J. Jones, P. B. Singh, M. M. Karimi, and M. C. Lorincz

2013. Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. *Epigenetics & Chromatin*, 6(1):15.

Malik, H. S., W. D. Burke, and T. H. Eickbush

1999. The age and evolution of non-LTR retrotransposable elements. *Molecular Biology and Evolution*, 16(6):793–805.

- Malik, H. S., S. Henikoff, and T. H. Eickbush 2000. Poised for Contagion: Evolutionary Origins of the Infectious Abilities of Invertebrate Retroviruses. *Genome Research*, 10(9):1307–1318.
- Malone, C. D., J. Brennecke, M. Dus, A. Stark, W. R. McCombie, R. Sachidanandam, and G. J. Hannon

2009. Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. *Cell*, 137(3):522–535.

Malone, C. D. and G. J. Hannon

2009. Small RNAs as Guardians of the Genome. *Cell*, 136(4):656–668.

Manuelidis, L.

1990. A view of interphase chromosomes. *Science (New York, N.Y.)*, 250(4987):1533–1540.

Margueron, R.

2009. Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature*, 461:762–767.

Margueron, R., G. Li, K. Sarma, A. Blais, J. Zavadil, C. L. Woodcock, B. D. Dynlacht,

and D. Reinberg

2008. Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. Molecular Cell, 32(4):503–518.

Margueron, R. and D. Reinberg 2010. Chromatin structure and the inheritance of epigenetic information. Nature Reviews. Genetics, 11(4):285–296.

Marks, H., T. Kalkan, R. Menafra, S. Denissov, K. Jones, H. Hofemeister, J. Nichols, A. Kranz, A. Francis Stewart, A. Smith, and H. Stunnenberg

2012. The Transcriptional and Epigenomic Foundations of Ground State Pluripotency. *Cell*, 149(3):590–604.

Marmorstein, R.

2001. Structure and function of histone acetyltransferases. *Cellular and molecular life sciences: CMLS*, 58(5-6):693–703.

Martin, G. R.

1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 78(12):7634–7638.

Martin, S. L., D. Branciforte, D. Keller, and D. L. Bain

2003. Trimeric structure for an essential protein in L1 retrotransposition. *Proceedings* of the National Academy of Sciences of the United States of America, 100(24):13815– 13820.

- Mathias, S. L., A. F. Scott, H. H. Kazazian, J. D. Boeke, and A. Gabriel 1991. Reverse transcriptase encoded by a human transposable element. *Science* (New York, N.Y.), 254(5039):1808–1810.
- Matranga, C., Y. Tomari, C. Shin, D. P. Bartel, and P. D. Zamore 2005. Passenger-Strand Cleavage Facilitates Assembly of siRNA into Ago2-Containing RNAi Enzyme Complexes. *Cell*, 123(4):607–620.
- Matsui, T., D. Leung, H. Miyashita, I. A. Maksakova, H. Miyachi, H. Kimura,M. Tachibana, M. C. Lorincz, and Y. Shinkai2010. Proviral silencing in embryonic stem cells requires the histone methyltrans-

ferase ESET. Nature, 464(7290):927-931.

- Maunakea, A. K., R. P. Nagarajan, M. Bilenky, T. J. Ballinger, C. D'Souza, S. D.
 Fouse, B. E. Johnson, C. Hong, C. Nielsen, Y. Zhao, G. Turecki, A. Delaney,
 R. Varhol, N. Thiessen, K. Shchors, V. M. Heine, D. H. Rowitch, X. Xing, C. Fiore,
 M. Schillebeeckx, S. J. M. Jones, D. Haussler, M. A. Marra, M. Hirst, T. Wang, and
 J. F. Costello
 - 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*, 466(7303):253–257.
- Mayer, W., A. Niveleau, J. Walter, R. Fundele, and T. Haaf 2000. Embryogenesis: Demethylation of the zygotic paternal genome. *Nature*, 403(6769):501–502.
- McBryant, S. J., V. H. Adams, and J. C. Hansen 2006. Chromatin architectural proteins. Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology, 14(1):39–51.
- McCarthy, E. M. and J. F. McDonald
 - 2004. Long terminal repeat retrotransposons of Mus musculus. *Genome Biology*, 5(3):R14.

McCLINTOCK, B.

- 1951. Chromosome organization and genic expression. *Cold Spring Harbor symposia* on quantitative biology, 16:13–47.
- Medeiros, L. A., L. M. Dennis, M. E. Gill, H. Houbaviy, S. Markoulaki, D. Fu, A. C. White, O. Kirak, P. A. Sharp, D. C. Page, and R. Jaenisch 2011. Mir-290–295 deficiency in mice results in partially penetrant embryonic lethality and germ cell defects. *Proceedings of the National Academy of Sciences of the United States of America*, 108(34):14163–14168.
- Medstrand, P., L. N. van de Lagemaat, and D. L. Mager 2002. Retroelement Distributions in the Human Genome: Variations Associated With Age and Proximity to Genes. *Genome Research*, 12(10):1483–1495.
- Meehan, R. R., J. D. Lewis, and A. P. Bird
 - 1992. Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Research*, 20(19):5085–5092.

- Meissner, A., A. Gnirke, G. W. Bell, B. Ramsahoye, E. S. Lander, and R. Jaenisch 2005. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Research*, 33(18):5868–5877.
- Meissner, A., T. S. Mikkelsen, H. Gu, M. Wernig, J. Hanna, A. Sivachenko, X. Zhang, B. E. Bernstein, C. Nusbaum, D. B. Jaffe, A. Gnirke, R. Jaenisch, and E. S. Lander 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, 454(7205):766–770.
- Meister, G., M. Landthaler, Y. Dorsett, and T. Tuschl 2004. Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing. RNA, 10(3):544–550.
- Melé, M. and J. L. Rinn 2016. "Cat's Cradling" the 3d Genome by the Act of LncRNA Transcription. *Molec*ular Cell, 62(5):657–664.
- Mendenhall, E. M., R. P. Koche, T. Truong, V. W. Zhou, B. Issac, A. S. Chi, M. Ku, and B. E. Bernstein

2010. GC-Rich Sequence Elements Recruit PRC2 in Mammalian ES Cells. *PLOS Genet*, 6(12):e1001244.

Metzger, E., M. Wissmann, N. Yin, J. M. Müller, R. Schneider, A. H. F. M. Peters, T. Günther, R. Buettner, and R. Schüle 2005. LSD1 demethylates repressive histone marks to promote androgen-receptor-

dependent transcription. *Nature*, 437(7057):436–439.

Mi, H., S. Poudel, A. Muruganujan, J. T. Casagrande, and P. D. Thomas 2016. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Research*, 44(Database issue):D336–D342.

Miescher, F.

1871. Ueber die chemische Zusammensetzung der Eiterzellen IN: Hoppe-Seyler(ed) Medicinsich-chemische Untersuchungen, Heft 4, 441-60., Miescher - BooksManuscripts & Prints - Jeremy Norman & Co., Inc.

Mikkelsen, T. S., J. Hanna, X. Zhang, M. Ku, M. Wernig, P. Schorderet, B. E. Bernstein, R. Jaenisch, E. S. Lander, and A. Meissner

2008. Dissecting direct reprogramming through integrative genomic analysis. Nature,

454(7200):49-55.

- 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.
- Mitchell, R. S., B. F. Beitzel, A. R. W. Schroder, P. Shinn, H. Chen, C. C. Berry, J. R. Ecker, and F. D. Bushman
 - 2004. Retroviral DNA Integration: ASLV, HIV, and MLV Show Distinct Target Site Preferences. *PLOS Biol*, 2(8):e234.
- Mitsui, K., Y. Tokuzawa, H. Itoh, K. Segawa, M. Murakami, K. Takahashi,M. Maruyama, M. Maeda, and S. Yamanaka2003. The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in
 - Mouse Epiblast and ES Cells. *Cell*, 113(5):631–642.
- Mohn, F., M. Weber, M. Rebhan, T. C. Roloff, J. Richter, M. B. Stadler, M. Bibel, and D. Schübeler
 - 2008. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Molecular Cell*, 30(6):755–766.
- Molaro, A., I. Falciatori, E. Hodges, A. A. Aravin, K. Marran, S. Rafii, W. R. McCombie, A. D. Smith, and G. J. Hannon
 - 2014. Two waves of de novo methylation during mouse germ cell development. Genes
 & Development, 28(14):1544–1549.
- Molyneaux, K. A., J. Stallock, K. Schaible, and C. Wylie 2001. Time-lapse analysis of living mouse germ cell migration. *Developmental Biology*, 240(2):488–498.
- Moran, J. V., R. J. DeBerardinis, and H. H. Kazazian 1999. Exon shuffling by L1 retrotransposition. Science (New York, N.Y.), 283(5407):1530–1534.
- Moran, J. V., S. E. Holmes, T. P. Naas, R. J. DeBerardinis, J. D. Boeke, and H. H.

Kazazian

1996. High frequency retrotransposition in cultured mammalian cells. *Cell*, 87(5):917–927.

- Morey, L., G. Pascual, L. Cozzuto, G. Roma, A. Wutz, S. A. Benitah, and L. Di Croce 2012. Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. *Cell Stem Cell*, 10(1):47–62.
- Morrish, T. A., N. Gilbert, J. S. Myers, B. J. Vincent, T. D. Stamato, G. E. Taccioli,
 M. A. Batzer, and J. V. Moran
 2002. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition.
 Nature Genetics, 31(2):159–165.
- Morselli, M., W. A. Pastor, B. Montanini, K. Nee, R. Ferrari, K. Fu, G. Bonora, L. Rubbi, A. T. Clark, S. Ottonello, S. E. Jacobsen, and M. Pellegrini 2015. In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. *eLife*, 4:e06205.
- Muller, S., P. Filippakopoulos, and S. Knapp
 2011. Bromodomains as therapeutic targets. *Expert Reviews in Molecular Medicine*, 13.
- Murchison, E. P., J. F. Partridge, O. H. Tam, S. Cheloufi, and G. J. Hannon 2005. Characterization of Dicer-deficient murine embryonic stem cells. *Proceedings* of the National Academy of Sciences of the United States of America, 102(34):12135– 12140.
- Murchison, E. P., P. Stein, Z. Xuan, H. Pan, M. Q. Zhang, R. M. Schultz, and G. J. Hannon

2007. Critical roles for Dicer in the female germline. Genes & Development, 21(6):682–693.

- Muñoz-López, M. and J. L. García-Pérez
 - 2010. DNA Transposons: Nature and Applications in Genomics. *Current Genomics*, 11(2):115–128.

Nagano, T. and P. Fraser

2011. No-Nonsense Functions for Long Noncoding RNAs. Cell, 145(2):178–181.

Najafabadi, H. S., S. Mnaimneh, F. W. Schmitges, M. Garton, K. N. Lam, A. Yang,

M. Albu, M. T. Weirauch, E. Radovani, P. M. Kim, J. Greenblatt, B. J. Frey, and T. R. Hughes

2015. C2h2 zinc finger proteins greatly expand the human regulatory lexicon. *Nature Biotechnology*, 33(5):555–562.

- Nakamura, T., Y. Arai, H. Umehara, M. Masuhara, T. Kimura, H. Taniguchi, T. Sekimoto, M. Ikawa, Y. Yoneda, M. Okabe, S. Tanaka, K. Shiota, and T. Nakano 2006. PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nature Cell Biology*, 9(1):64–71.
- Nakamura, T., Y.-J. Liu, H. Nakashima, H. Umehara, K. Inoue, S. Matoba, M. Tachibana, A. Ogura, Y. Shinkai, and T. Nakano 2012. PGC7 binds histone H3k9me2 to protect against conversion of 5mc to 5hmc in early embryos. *Nature*, 486(7403):415–419.
- Nekrasov, M., B. Wild, and J. Müller 2005. Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. EMBO Reports, 6(4):348–353.
- Nekrutenko, A. and W. H. Li

2001. Transposable elements are found in a large number of human protein-coding genes. *Trends in genetics: TIG*, 17(11):619–621.

Neri, F., D. Incarnato, A. Krepelova, S. Rapelli, A. Pagnani, R. Zecchina, C. Parlato, and S. Oliviero

2013. Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells. *Genome Biology*, 14(8):R91.

- Ng, J.-H., V. Kumar, M. Muratani, P. Kraus, J.-C. Yeo, L.-P. Yaw, K. Xue, T. Lufkin, S. Prabhakar, and H.-H. Ng
 - 2013. In Vivo Epigenomic Profiling of Germ Cells Reveals Germ Cell Molecular Signatures. *Developmental Cell*, 24(3):324–333.
- Nichols, J., B. Zevnik, K. Anastassiadis, H. Niwa, D. Klewe-Nebenius, I. Chambers, H. Schöler, and A. Smith

1998. Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4. *Cell*, 95(3):379–391.

Nielsen, C. B., N. Shomron, R. Sandberg, E. Hornstein, J. Kitzman, and C. B. Burge

2007. Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA*, 13(11):1894–1910.

- Nielsen, P. R., D. Nietlispach, H. R. Mott, J. Callaghan, A. Bannister, T. Kouzarides,
 A. G. Murzin, N. V. Murzina, and E. D. Laue
 2002. Structure of the HP1 chromodomain bound to histone H3 methylated at lysine
 9. Nature, 416(6876):103–107.
- Nishihara, H., A. F. A. Smit, and N. Okada 2006. Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Research*, 16(7):864–874.
- Niwa, H., K. Ogawa, D. Shimosato, and K. Adachi 2009. A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature*, 460(7251):118–122.
- Noma, K., C. D. Allis, and S. I. Grewal

2001. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science*, 293:1150–1155.

Nora, E. P., B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Blüthgen, J. Dekker, and E. Heard

2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature, 485(7398):381–385.

- Norton, V. G., B. S. Imai, P. Yau, and E. M. Bradbury 1989. Histone acetylation reduces nucleosome core particle linking number change. *Cell*, 57(3):449–457.
- Nusse, R. and H. E. Varmus

1982. Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. Cell, 31(1):99–109.

- O'Carroll, D., S. Erhardt, M. Pagani, S. C. Barton, M. A. Surani, and T. Jenuwein 2001. The Polycomb-Group Gene Ezh2 Is Required for Early Mouse Development. *Molecular and Cellular Biology*, 21(13):4330–4336.
- Ohm, J. E., K. M. McGarvey, X. Yu, L. Cheng, K. E. Schuebel, L. Cope, H. P. Mohammad, W. Chen, V. C. Daniel, W. Yu, D. M. Berman, T. Jenuwein, K. Pruitt,

- S. J. Sharkis, D. N. Watkins, J. G. Herman, and S. B. Baylin
- 2007. A stem cell-like chromatin pattern may predispose tumor suppressor genes to
- DNA hypermethylation and heritable silencing. *Nature Genetics*, 39(2):237–242.
- Ohno, S.
 - 1972. So much "junk" DNA in our genome. *Brookhaven Symposia in Biology*, 23:366–370.
- Okada, C., E. Yamashita, S. J. Lee, S. Shibata, J. Katahira, A. Nakagawa, Y. Yoneda, and T. Tsukihara

2009. A high-resolution structure of the pre-microRNA nuclear export machinery. *Science (New York, N.Y.)*, 326(5957):1275–1279.

- Okamura, K., W.-J. Chung, J. G. Ruby, H. Guo, D. P. Bartel, and E. C. Lai 2008. The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. *Nature*, 453(7196):803–806.
- Okamura, K., A. Ishizuka, H. Siomi, and M. C. Siomi 2004. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes & Development*, 18(14):1655–1666.
- Okamura, K. and K. Nakai
 - 2008. Retrotransposition as a Source of New Promoters. *Molecular Biology and Evolution*, 25(6):1231–1238.
- Okano, M., D. W. Bell, D. A. Haber, and E. Li 1999. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell*, 99(3):247–257.
- Okano, M., S. Xie, and E. Li 1998. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature Genetics*, 19(3):219–220.
- Olins, A. L. and D. E. Olins
 - 1974. Spheroid chromatin units (v bodies). Science (New York, N.Y.), 183(4122):330–332.
- Onder, T. T., N. Kara, A. Cherry, A. U. Sinha, N. Zhu, K. M. Bernt, P. Cahan, O. B. Mancarci, J. Unternaehrer, P. B. Gupta, E. S. Lander, S. A. Armstrong, and G. Q.

Daley

2012. Chromatin modifying enzymes as modulators of reprogramming. *Nature*, 483(7391):598–602.

- Ono, M., T. Yasunaga, T. Miyata, and H. Ushikubo 1986. Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. *Journal of Virology*, 60(2):589–598.
- Ooi, S. K. T., C. Qiu, E. Bernstein, K. Li, D. Jia, Z. Yang, H. Erdjument-Bromage,
 P. Tempst, S.-P. Lin, C. D. Allis, X. Cheng, and T. H. Bestor
 2007. DNMT3l connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*, 448(7154):714–717.
- Orgel, L. E. and F. H. C. Crick

1980. Selfish DNA: the ultimate parasite. Nature, 284(5757):604-607.

- Ovchinnikov, I., A. B. Troxel, and G. D. Swergold 2001. Genomic characterization of recent human LINE-1 insertions: evidence supporting random insertion. *Genome Research*, 11(12):2050–2058.
- Pan, G. J., Z. Y. Chang, H. R. Schöler, and D. Pei 2002. Stem cell pluripotency and transcription factor Oct4. *Cell Research*, 12(5):321– 329.
- Papp, B. and K. Plath

2013. Epigenetics of reprogramming to induced pluripotency. Cell, 152(6):1324–1343.

Park, I.-H., R. Zhao, J. A. West, A. Yabuuchi, H. Huo, T. A. Ince, P. H. Lerou, M. W. Lensch, and G. Q. Daley

2008. Reprogramming of human somatic cells to pluripotency with defined factors. *Nature*, 451(7175):141–146.

Pastor, W. A., H. Stroud, K. Nee, W. Liu, D. Pezic, S. Manakov, S. A. Lee, G. Moissiard, N. Zamudio, D. Bourc'his, A. A. Aravin, A. T. Clark, and S. E. Jacobsen

2014. MORC1 represses transposable elements in the mouse male germline. *Nature Communications*, 5:5795.

Pauler, F. M., M. A. Sloane, R. Huang, K. Regha, M. V. Koerner, I. Tamir, A. Sommer,

A. Aszodi, T. Jenuwein, and D. P. Barlow

2009. H3k27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. *Genome Research*, 19(2):221–233.

Peaston, A. E., A. V. Evsikov, J. H. Graber, W. N. de Vries, A. E. Holbrook, D. Solter, and B. B. Knowles

2004. Retrotransposons Regulate Host Genes in Mouse Oocytes and Preimplantation Embryos. *Developmental Cell*, 7(4):597–606.

Peat, J., W. Dean, S. Clark, F. Krueger, S. Smallwood, G. Ficz, J. Kim, J. Marioni,T. Hore, and W. Reik2014. Genome-wide Bisulfite Sequencing in Zygotes Identifies Demethylation Targets

and Maps the Contribution of TET3 Oxidation. Cell Reports, 9(6):1990-2000.

Peters, A. H. F. M., S. Kubicek, K. Mechtler, R. J. O'Sullivan, A. A. H. A. Derijck, L. Perez-Burgos, A. Kohlmaier, S. Opravil, M. Tachibana, Y. Shinkai, J. H. A. Martens, and T. Jenuwein

2003. Partitioning and Plasticity of Repressive Histone Methylation States in Mammalian Chromatin. *Molecular Cell*, 12(6):1577–1589.

Peters, A. H. F. M., D. O'Carroll, H. Scherthan, K. Mechtler, S. Sauer, C. Schöfer, K. Weipoltshammer, M. Pagani, M. Lachner, A. Kohlmaier, S. Opravil, M. Doyle, M. Sibilia, and T. Jenuwein

2001. Loss of the Suv39h Histone Methyltransferases Impairs Mammalian Heterochromatin and Genome Stability. *Cell*, 107(3):323–337.

- Peters, G. and C. Glover
 - 1980. tRNA's and priming of RNA-directed DNA synthesis in mouse mammary tumor virus. *Journal of Virology*, 35(1):31–40.
- Pickeral, O. K., W. Makałowski, M. S. Boguski, and J. D. Boeke 2000. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Research*, 10(4):411–415.
- Pickersgill, H., B. Kalverda, E. de Wit, W. Talhout, M. Fornerod, and B. van Steensel 2006. Characterization of the Drosophila melanogaster genome at the nuclear lamina. *Nature Genetics*, 38(9):1005–1014.

Pinheiro, I., R. Margueron, N. Shukeir, M. Eisold, C. Fritzsch, F. M. Richter, G. Mittler, C. Genoud, S. Goyama, M. Kurokawa, J. Son, D. Reinberg, M. Lachner, and T. Jenuwein

2012. Prdm3 and Prdm16 are H3k9me1 methyltransferases required for mammalian heterochromatin integrity. *Cell*, 150(5):948–960.

Plasterk, R. H., Z. Izsvák, and Z. Ivics

1999. Resident aliens: the Tc1/mariner superfamily of transposable elements. *Trends* in genetics: TIG, 15(8):326–332.

Plath, K.

2003. Role of histone H3 lysine 27 methylation in X inactivation. *Science*, 300:131–135.

Pokholok, D. K., C. T. Harbison, S. Levine, M. Cole, N. M. Hannett, T. I. Lee, G. W. Bell, K. Walker, P. A. Rolfe, E. Herbolsheimer, J. Zeitlinger, F. Lewitter, D. K. Gifford, and R. A. Young

2005. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell*, 122(4):517–527.

- Polo, J., E. Anderssen, R. Walsh, B. Schwarz, C. Nefzger, S. Lim, M. Borkent, E. Apostolou, S. Alaei, J. Cloutier, O. Bar-Nur, S. Cheloufi, M. Stadtfeld, M. Figueroa, D. Robinton, S. Natesan, A. Melnick, J. Zhu, S. Ramaswamy, and K. Hochedlinger 2012. A Molecular Roadmap of Reprogramming Somatic Cells into iPS Cells. *Cell*, 151(7):1617–1632.
- Popp, C., W. Dean, S. Feng, S. J. Cokus, S. Andrews, M. Pellegrini, S. E. Jacobsen, and W. Reik

2010. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature*, 463(7284):1101–1105.

Pradeepa, M. M., G. R. Grimes, Y. Kumar, G. Olley, G. C. A. Taylor, R. Schneider, and W. A. Bickmore

2016. Histone H3 globular domain acetylation identifies a new class of enhancers. Nature Genetics, 48(6):681–686.

Provost, P., D. Dishart, J. Doucet, D. Frendewey, B. Samuelsson, and O. Rådmark2002. Ribonuclease activity and RNA binding of recombinant human Dicer. The

EMBO journal, 21(21):5864–5874.

- Prud'homme, N., M. Gans, M. Masson, C. Terzian, and A. Bucheton 1995. Flamenco, a gene controlling the gypsy retrovirus of Drosophila melanogaster. *Genetics*, 139(2):697–711.
- Quenneville, S., P. Turelli, K. Bojkowska, C. Raclot, S. Offner, A. Kapopoulou, and D. Trono
 - 2012. The KRAB-ZFP/KAP1 system contributes to the early embryonic establishment of site-specific DNA methylation patterns maintained during development. *Cell Reports*, 2(4):766–773.
- Quenneville, S., G. Verde, A. Corsinotti, A. Kapopoulou, J. Jakobsson, S. Offner,
 I. Baglivo, P. Pedone, G. Grimaldi, A. Riccio, and D. Trono
 2011. In Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions.
 Molecular Cell, 44(3):361–372.
- Rada-Iglesias, A.

2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, 470:279–283.

- Ran, F. A., P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott, and F. Zhang 2013. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, 8(11):2281–2308.
- Rand, T. A., S. Petersen, F. Du, and X. Wang 2005. Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell*, 123(4):621–629.
- Rando, O. J.
 - 2012. Combinatorial complexity in chromatin structure and function: revisiting the histone code. *Current Opinion in Genetics & Development*, 22(2):148–155.
- Rao, S. S. P., M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. L. Aiden 2014. A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7):1665–1680.
- Razin, A. and A. D. Riggs

1980. DNA methylation and gene function. *Science*, 210:604–610.

Rea, S.

2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, 406:593–599.

- Rehwinkel, J., P. Natalin, A. Stark, J. Brennecke, S. M. Cohen, and E. Izaurralde 2006. Genome-Wide Analysis of mRNAs Regulated by Drosha and Argonaute Proteins in Drosophila melanogaster. *Molecular and Cellular Biology*, 26(8):2965–2975.
- Reik, W. and M. A. Surani
- 2015. Germline and Pluripotent Stem Cells. Cold Spring Harbor Perspectives in Biology, 7(11).
- Reik, W. and J. Walter
- 2001. Genomic imprinting: parental influence on the genome. *Nature Reviews Genetics*, 2(1):21–32.
- Reuter, M., P. Berninger, S. Chuma, H. Shah, M. Hosokawa, C. Funaya, C. Antony, R. Sachidanandam, and R. S. Pillai

2011. Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature*, 480(7376):264–267.

Reynolds, N., P. Latos, A. Hynes-Allen, R. Loos, D. Leaford, A. O'Shaughnessy,O. Mosaku, J. Signolet, P. Brennecke, T. Kalkan, I. Costello, P. Humphreys,W. Mansfield, K. Nakagawa, J. Strouboulis, A. Behrens, P. Bertone, and B. Hendrich

2012. NuRD Suppresses Pluripotency Gene Expression to Promote Transcriptional Heterogeneity and Lineage Commitment. *Cell Stem Cell*, 10(5):583–594.

Ribet, D., F. Harper, A. Dupressoir, M. Dewannieux, G. Pierron, and T. Heidmann 2008. An infectious progenitor for the murine IAP retrotransposon: Emergence of an intracellular genetic parasite from an ancient retrovirus. *Genome Research*, 18(4):597–609.

1975. X inactivation, differentiation, and DNA methylation. *Cytogenetics and Cell Genetics*, 14(1):9–25.

Rinn, J. L., M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. A. Brugmann, L. H.

Riggs, A. D.

Goodnough, J. A. Helms, P. J. Farnham, E. Segal, and H. Y. Chang2007. Functional demarcation of active and silent chromatin domains in human HOX

loci by noncoding RNAs. Cell, 129(7):1311–1323.

- Rodda, D. J., J.-L. Chew, L.-H. Lim, Y.-H. Loh, B. Wang, H.-H. Ng, and P. Robson 2005. Transcriptional Regulation of Nanog by OCT4 and SOX2. *Journal of Biological Chemistry*, 280(26):24731–24737.
- Ross, J. P., K. N. Rand, and P. L. Molloy
 2010. Hypomethylation of repeated DNA sequences in cancer. *Epigenomics*, 2(2):245–269.
- Rowe, H. M., M. Friedli, S. Offner, S. Verp, D. Mesnard, J. Marquis, T. Aktas, and D. Trono

2013. De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. *Development (Cambridge, England)*, 140(3):519–529.

- Rowe, H. M., J. Jakobsson, D. Mesnard, J. Rougemont, S. Reynard, T. Aktas, P. V. Maillard, H. Layard-Liesching, S. Verp, J. Marquis, F. Spitz, D. B. Constam, and D. Trono
 - 2010. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature*, 463(7278):237–240.
- Rowe, H. M. and D. Trono
 - 2011. Dynamic control of endogenous retroviruses during development. Virology,411(2):273–287.
- Rubin, G. M., M. G. Kidwell, and P. M. Bingham 1982. The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. *Cell*, 29(3):987–994.
- Ryan, R. F., D. C. Schultz, K. Ayyanathan, P. B. Singh, J. R. Friedman, W. J. Fredericks, and F. J. Rauscher

1999. KAP-1 corepressor protein interacts and colocalizes with heterochromatic and euchromatic HP1 proteins: a potential role for Krüppel-associated box-zinc finger proteins in heterochromatin-mediated gene silencing. *Molecular and Cellular Biology*, 19(6):4366–4378.

Saito, K., Y. Sakaguchi, T. Suzuki, T. Suzuki, H. Siomi, and M. C. Siomi

2007. Pimet, the Drosophila homolog of HEN1, mediates 2?-O-methylation of Piwiinteracting RNAs at their 3? ends. *Genes & Development*, 21(13):1603–1608.

Saitou, M. and M. Yamaji

2010. Germ cell specification in mice: signaling, transcription regulation, and epigenetic consequences. *Reproduction*, 139(6):931–942.

- Saksouk, N., T. Barth, C. Ziegler-Birling, N. Olova, A. Nowak, E. Rey, J. Mateos-Langerak, S. Urbach, W. Reik, M.-E. Torres-Padilla, A. Imhof, and J. Déjardin 2014. Redundant Mechanisms to Form Silent Chromatin at Pericentromeric Regions Rely on BEND3 and DNA Methylation. *Molecular Cell*, 56(4):580–594.
- Santos, F., B. Hendrich, W. Reik, and W. Dean 2002. Dynamic Reprogramming of DNA Methylation in the Early Mouse Embryo. *Developmental Biology*, 241(1):172–182.
- Santos, F., J. Peat, H. Burgess, C. Rada, W. Reik, and W. Dean 2013. Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. *Epigenetics & Chromatin*, 6:39.
- Santos, F., A. H. Peters, A. P. Otte, W. Reik, and W. Dean 2005. Dynamic chromatin modifications characterise the first cell cycle in mouse embryos. *Developmental Biology*, 280(1):225–236.
- Sasaki, M., J. Lange, and S. Keeney
- 2010. Genome destabilization by homologous recombination in the germ line. *Nature Reviews. Molecular Cell Biology*, 11(3):182–195.
- Saunders, A., F. Faiola, and J. Wang

2013. Concise Review: Pursuing Self-Renewal and Pluripotency with the Stem Cell Factor Nanog. *STEM CELLS*, 31(7):1227–1236.

Saxton, J. A. and S. L. Martin

1998. Recombination between subtypes creates a mosaic lineage of LINE-1 that is expressed and actively retrotransposing in the mouse genome. *Journal of Molecular Biology*, 280(4):611–622.

Schmitges, F., A. Prusty, M. Faty, A. Stützer, G. Lingaraju, J. Aiwazian, R. Sack, D. Hess, L. Li, S. Zhou, R. Bunker, U. Wirth, T. Bouwmeester, A. Bauer, N. LyHartig, K. Zhao, H. Chan, J. Gu, H. Gut, W. Fischle, J. Müller, and N. Thomä 2011. Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. *Molecular Cell*, 42(3):330–341.

- Schueler, M. G., A. W. Higgins, M. K. Rudd, K. Gustashaw, and H. F. Willard 2001. Genomic and genetic definition of a functional human centromere. *Science* (New York, N.Y.), 294(5540):109–115.
- Schultz, D. C., K. Ayyanathan, D. Negorev, G. G. Maul, and F. J. Rauscher 2002. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes & Development*, 16(8):919–932.
- Schultz, D. C., J. R. Friedman, and F. J. Rauscher
- 2001. Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. *Genes & Development*, 15(4):428–443.
- Schultz, J.

1936. Variegation in Drosophila and the Inert Chromosome Regions. Proceedings of the National Academy of Sciences of the United States of America, 22(1):27–33.

- Schulz, W. A.
 - 2006. L1 Retrotransposons in Human Cancers. Journal of Biomedicine and Biotechnology, 2006.
- Seisenberger, S., S. Andrews, F. Krueger, J. Arand, J. Walter, F. Santos, C. Popp,B. Thienpont, W. Dean, and W. Reik2012. The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse
- Primordial Germ Cells. *Molecular Cell*, 48(6):849–862.
- Seisenberger, S., J. R. Peat, T. A. Hore, F. Santos, W. Dean, and W. Reik 2013a. Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 368(1609):20110330.
- Seisenberger, S., J. R. Peat, and W. Reik
 - 2013b. Conceptual links between DNA methylation reprogramming in the early embryo and primordial germ cells. *Current opinion in cell biology*.
- Seitz, H., H. Royo, M.-L. Bortolin, S.-P. Lin, A. C. Ferguson-Smith, and J. Cavaillé 2004. A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. *Genome Research*, 14(9):1741–1748.
- Seki, Y., K. Hayashi, K. Itoh, M. Mizugaki, M. Saitou, and Y. Matsui 2005. Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Developmental Biology*, 278(2):440–458.
- Seki, Y., M. Yamaji, Y. Yabuta, M. Sano, M. Shigeta, Y. Matsui, Y. Saga, M. Tachibana, Y. Shinkai, and M. Saitou
 2007. Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. *Development*, 134(14):2627–2638.
- Sen, S. K., K. Han, J. Wang, J. Lee, H. Wang, P. A. Callinan, M. Dyer, R. Cordaux,P. Liang, and M. A. Batzer2006. Human Genomic Deletions Mediated by Recombination between Alu Elements.
- Shao, Z., F. Raible, R. Mollaaghababa, J. R. Guyon, C. T. Wu, W. Bender, and R. E. Kingston

American Journal of Human Genetics, 79(1):41–53.

1999. Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell*, 98(1):37–46.

1969. Mutations caused by the insertion of genetic material into the galactose operon of Escherichia coli. *Journal of Molecular Biology*, 40(1):93–105.

- Sharif, J., T. A. Endo, M. Nakayama, M. M. Karimi, M. Shimada, K. Katsuyama,
 P. Goyal, J. Brind'Amour, M.-A. Sun, Z. Sun, T. Ishikura, Y. Mizutani-Koseki,
 O. Ohara, Y. Shinkai, M. Nakanishi, H. Xie, M. C. Lorincz, and H. Koseki
 2016. Activation of Endogenous Retroviruses in Dnmt1-/- ESCs Involves Disruption of SETDB1-Mediated Repression by NP95 Binding to Hemimethylated DNA.
 Cell Stem Cell, 0(0).
- Sharif, J., M. Muto, S.-i. Takebayashi, I. Suetake, A. Iwamatsu, T. A. Endo, J. Shinga, Y. Mizutani-Koseki, T. Toyoda, K. Okamura, S. Tajima, K. Mitsuya, M. Okano,

Shapiro, J. A.

and H. Koseki

2007. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*, 450(7171):908–912.

Shen, L., H. Wu, D. Diep, S. Yamaguchi, A. D'Alessio, H.-L. Fung, K. Zhang, and Y. Zhang

2013. Genome-wide Analysis Reveals TET- and TDG-Dependent 5-Methylcytosine Oxidation Dynamics. *Cell*, 153(3):692–706.

Shi, Y.

- Shin, C., J.-W. Nam, K. K.-H. Farh, H. R. Chiang, A. Shkumatava, and D. P. Bartel 2010. Expanding the MicroRNA Targeting Code: Functional Sites with Centered Pairing. *Molecular Cell*, 38(6):789–802.
- Shrivastava, S., R. Steele, R. Ray, and R. B. Ray 2015. MicroRNAs: Role in hepatitis C virus pathogenesis. *Genes & Diseases*, 2(1):35–45.
- Sienski, G., D. Dönertas, and J. Brennecke 2012. Transcriptional Silencing of Transposons by Piwi and Maelstrom and Its Impact on Chromatin State and Gene Expression. *Cell*, 151(5):964–980.
- Silva, J., W. Mak, I. Zvetkova, R. Appanah, T. B. Nesterova, Z. Webster, A. H. F. M. Peters, T. Jenuwein, A. P. Otte, and N. Brockdorff 2003. Establishment of Histone H3 Methylation on the Inactive X Chromosome Requires Transient Recruitment of Eed-Enx1 Polycomb Group Complexes. *Devel*opmental Cell, 4(4):481–495.
- Silva, J., J. Nichols, T. W. Theunissen, G. Guo, A. L. v. Oosten, O. Barrandon, J. Wray, S. Yamanaka, I. Chambers, and A. Smith

2009. Nanog Is the Gateway to the Pluripotent Ground State. Cell, 138(4):722–737.

Sinkkonen, L., T. Hugenschmidt, P. Berninger, D. Gaidatzis, F. Mohn, C. G. Artus-Revel, M. Zavolan, P. Svoboda, and W. Filipowicz

2008. MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. *Nature Structural & Molecular*

^{2004.} Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell*, 119:941–953.

Biology, 15(3):259-267.

Sinkkonen, L., T. Hugenschmidt, W. Filipowicz, and P. Svoboda

2010. Dicer Is Associated with Ribosomal DNA Chromatin in Mammalian Cells. *PLoS ONE*, 5(8):e12175.

Skardal, A., D. Mack, A. Atala, and S. Soker

2013. Substrate elasticity controls cell proliferation, surface marker expression and motile phenotype in amniotic fluid-derived stem cells. *Journal of the mechanical behavior of biomedical materials*, 17:307–316.

Slotkin, R. K. and R. Martienssen

2007. Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics*, 8(4):272–285.

Smalheiser, N. R. and V. I. Torvik

2005. Mammalian microRNAs derived from genomic repeats. *Trends in genetics: TIG*, 21(6):322–326.

- Smallwood, A., P.-O. Estève, S. Pradhan, and M. Carey 2007. Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes & Development*, 21(10):1169–1178.
- Smallwood, S. A., S.-i. Tomizawa, F. Krueger, N. Ruf, N. Carli, A. Segonds-Pichon, S. Sato, K. Hata, S. R. Andrews, and G. Kelsey

2011. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nature Genetics*, 43(8):811–814.

Smit, A. F.

1993. Identification of a new, abundant superfamily of mammalian LTR-transposons. Nucleic Acids Research, 21(8):1863–1872.

Smit, A. F.

1996. The origin of interspersed repeats in the human genome. Current Opinion in Genetics & Development, 6(6):743–748.

Smit, A. F.

1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. Current Opinion in Genetics & Development, 9(6):657–663.
Smith, A. G.

- 2001. Embryo-Derived Stem Cells: Of Mice and Men. Annual Review of Cell and Developmental Biology, 17(1):435–462.
- Smith, A. G., J. K. Heath, D. D. Donaldson, G. G. Wong, J. Moreau, M. Stahl, and D. Rogers

1988. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature*, 336(6200):688–690.

Smith, Z. D., M. M. Chan, T. S. Mikkelsen, H. Gu, A. Gnirke, A. Regev, and A. Meissner

2012. A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature*, 484(7394):339–344.

- Soifer, H. S., A. Zaragoza, M. Peyvan, M. A. Behlke, and J. J. Rossi 2005. A potential role for RNA interference in controlling the activity of the human LINE-1 retrotransposon. *Nucleic Acids Research*, 33(3):846–856.
- Song, J., O. Rechkoblit, T. H. Bestor, and D. J. Patel 2011. Structure of DNMT1-DNA Complex Reveals a Role for Autoinhibition in Maintenance DNA Methylation. *Science*, 331(6020):1036–1040.
- Song, J.-J., J. Liu, N. H. Tolia, J. Schneiderman, S. K. Smith, R. A. Martienssen, G. J. Hannon, and L. Joshua-Tor 2003. The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nature Structural & Molecular Biology*, 10(12):1026–1032.
- Song, J.-J., S. K. Smith, G. J. Hannon, and L. Joshua-Tor 2004. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science (New York, N.Y.)*, 305(5689):1434–1437.
- Sookdeo, A., C. M. Hepp, M. A. McClure, and S. Boissinot 2013. Revisiting the evolution of mouse LINE-1 in the genomic era. *Mobile DNA*, 4:3.
- Soper, S. F. C., G. W. van der Heijden, T. C. Hardiman, M. Goodheart, S. L. Martin, P. de Boer, and A. Bortvin

2008. Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. *Developmental Cell*, 15(2):285–297.

Soriano, P., M. Meunier-Rotival, and G. Bernardi

1983. The distribution of interspersed repeats is nonuniform and conserved in the mouse and human genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 80(7):1816–1820.

Speek, M.

2001. Antisense Promoter of Human L1 Retrotransposon Drives Transcription of Adjacent Cellular Genes. *Molecular and Cellular Biology*, 21(6):1973–1985.

Sripathy, S. P., J. Stevens, and D. C. Schultz

- 2006. The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression. *Molecular and Cellular Biology*, 26(22):8623–8638.
- Stadler, M. B., R. Murr, L. Burger, R. Ivanek, F. Lienert, A. Schöler, Erik van Nimwegen, C. Wirbelauer, E. J. Oakeley, D. Gaidatzis, V. K. Tiwari, and D. Schübeler 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, 480(7378):490–495.
- Statham, A. L., M. D. Robinson, J. Z. Song, M. W. Coolen, C. Stirzaker, and S. J. Clark

2012. Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. *Genome Research*, 22(6):1120–1127.

Stein, P., N. V. Rozhkov, F. Li, F. L. Cárdenas, O. Davydenk, L. E. Vandivier, B. D. Gregory, G. J. Hannon, and R. M. Schultz 2015. Essential Role for Endogenous siRNAs during Meiosis in Mouse Oocytes. *PLoS Genetics*, 11(2).

Stewart, M. D., J. Li, and J. Wong

- 2005. Relationship between Histone H3 Lysine 9 Methylation, Transcription Repression, and Heterochromatin Protein 1 Recruitment. *Molecular and Cellular Biology*, 25(7):2525–2538.
- Stock, J. K., S. Giadrossi, M. Casanova, E. Brookes, M. Vidal, H. Koseki, N. Brockdorff,

A. G. Fisher, and A. Pombo

2007. Ring1-mediated ubiquitination of H2a restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nature Cell Biology*, 9(12):1428–1435.

Stocking, C. and C. A. Kozak

2008. Murine endogenous retroviruses. *Cellular and molecular life sciences : CMLS*, 65(21):3383–3398.

Suda, Y., M. Suzuki, Y. Ikawa, and S. Aizawa 1987. Mouse embryonic stem cells exhibit indefinite proliferative potential. *Journal* of Cellular Physiology, 133(1):197–201.

Suetake, I., F. Shinozaki, J. Miyagawa, H. Takeshima, and S. Tajima 2004. DNMT3l Stimulates the DNA Methylation Activity of Dnmt3a and Dnmt3b through a Direct Interaction. *Journal of Biological Chemistry*, 279(26):27816–27823.

Surani, M. A., K. Hayashi, and P. Hajkova 2007. Genetic and Epigenetic Regulators of Pluripotency. *Cell*, 128(4):747–762.

Suzuki, M. M. and A. Bird

2008. DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics*, 9(6):465–476.

1998. Perpetually mobile footprints of ancient infections in human genome. *FEBS* Letters, 428(1–2):1–6.

- Svoboda, P., P. Stein, M. Anger, E. Bernstein, G. J. Hannon, and R. M. Schultz 2004. RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Developmental Biology*, 269(1):276–285.
- Symer, D. E., C. Connelly, S. T. Szak, E. M. Caputo, G. J. Cost, G. Parmigiani, and J. D. Boeke

2002. Human l1 retrotransposition is associated with genetic instability in vivo. Cell, 110(3):327-338.

Szak, S. T., O. K. Pickeral, W. Makalowski, M. S. Boguski, D. Landsman, and J. D. Boeke

2002. Molecular archeology of L1 insertions in the human genome. *Genome Biology*, 3(10):research0052.1–research0052.18.

Sverdlov, E. D.

- Tachibana, M., M. Nozaki, N. Takeda, and Y. Shinkai 2007. Functional dynamics of H3k9 methylation during meiotic prophase progression. *The EMBO Journal*, 26(14):3346–3359.
- Tachibana, M., K. Sugimoto, T. Fukushima, and Y. Shinkai 2001. SET domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. J. Biol. Chem., 276:25309–25317.
- Tachibana, M., K. Sugimoto, M. Nozaki, J. Ueda, T. Ohta, M. Ohki, M. Fukuda, N. Takeda, H. Niida, H. Kato, and Y. Shinkai
 2002. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes & Development*, 16(14):1779–1791.
- Tachibana, M., J. Ueda, M. Fukuda, N. Takeda, T. Ohta, H. Iwanari, T. Sakihama, T. Kodama, T. Hamakubo, and Y. Shinkai 2005. Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes & Development*, 19(7):815–826.
- Taddei, A., C. Maison, D. Roche, and G. Almouzni 2001. Reversible disruption of pericentric heterochromatin and centromere function by inhibiting deacetylases. *Nature Cell Biology*, 3(2):114–120.
- Tahiliani, M., K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, and A. Rao 2009. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science*, 324(5929):930–935.
- Takada, S., E. Berezikov, Y. L. Choi, Y. Yamashita, and H. Mano 2009. Potential role of miR-29b in modulation of Dnmt3a and Dnmt3b expression in primordial germ cells of female mouse embryos. *RNA*, 15(8):1507–1514.
- Takahashi, K., K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, and S. Yamanaka

2007. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*, 131(5):861–872.

Takahashi, K. and S. Yamanaka

2006. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, 126(4):663–676.

Takeshita, K., I. Suetake, E. Yamashita, M. Suga, H. Narita, A. Nakagawa, and S. Tajima

2011. Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). Proceedings of the National Academy of Sciences of the United States of America, 108(22):9055–9059.

- Tam, O. H., A. A. Aravin, P. Stein, A. Girard, E. P. Murchison, S. Cheloufi, E. Hodges, M. Anger, R. Sachidanandam, R. M. Schultz, and G. J. Hannon
 2008. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature*, 453(7194):534–538.
- Tanay, A., A. H. O'Donnell, M. Damelin, and T. H. Bestor 2007. Hyperconserved CpG domains underlie Polycomb-binding sites. *Proceedings* of the National Academy of Sciences of the United States of America, 104(13):5521– 5526.
- Tang, G., B. J. Reinhart, D. P. Bartel, and P. D. Zamore 2003. A biochemical framework for RNA silencing in plants. *Genes & Development*, 17(1):49–63.
- Tardat, M., M. Albert, R. Kunzmann, Z. Liu, L. Kaustov, R. Thierry, S. Duan, U. Brykczynska, C. Arrowsmith, and A. F. M. Peters 2015. Cbx2 Targets PRC1 to Constitutive Heterochromatin in Mouse Zygotes in a Parent-of-Origin-Dependent Manner. *Molecular Cell*, 58(1):157–171.
- Taunton, J., C. A. Hassig, and S. L. Schreiber 1996. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science, 272:408–411.
- Thiagalingam, S., K.-H. Cheng, H. J. Lee, N. Mineva, A. Thiagalingam, and J. F. Ponte

2003. Histone deacetylases: unique players in shaping the epigenetic histone code. Annals of the New York Academy of Sciences, 983:84–100.

Thoma, F., T. Koller, and A. Klug

1979. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *The Journal of Cell Biology*, 83(2 Pt 1):403–427.

- Thomas, J. H. and S. Schneider 2011. Coevolution of retroelements and tandem zinc finger genes. *Genome Research*, 21(11):1800–1812.
- Thomas, P. D., A. Kejariwal, N. Guo, H. Mi, M. J. Campbell, A. Muruganujan, and B. Lazareva-Ulitsky

2006. Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Research*, 34(Web Server issue):W645–650.

Thompson, C. B.

1995. New insights into V(D)J recombination and its role in the evolution of the immune system. *Immunity*, 3(5):531-539.

Thomson, T. and H. Lin

2009. The Biogenesis and Function of PIWI Proteins and piRNAs: Progress and Prospect. Annual Review of Cell and Developmental Biology, 25(1):355–376.

- Thurman, R. E., E. Rynes, R. Humbert, J. Vierstra, M. T. Maurano, E. Haugen, N. C. Sheffield, A. B. Stergachis, H. Wang, B. Vernot, K. Garg, S. John, R. Sandstrom, D. Bates, L. Boatman, T. K. Canfield, M. Diegel, D. Dunn, A. K. Ebersol, T. Frum, E. Giste, A. K. Johnson, E. M. Johnson, T. Kutyavin, B. Lajoie, B.-K. Lee, K. Lee, D. London, D. Lotakis, S. Neph, F. Neri, E. D. Nguyen, H. Qu, A. P. Reynolds, V. Roach, A. Safi, M. E. Sanchez, A. Sanyal, A. Shafer, J. M. Simon, L. Song, S. Vong, M. Weaver, Y. Yan, Z. Zhang, Z. Zhang, B. Lenhard, M. Tewari, M. O. Dorschner, R. S. Hansen, P. A. Navas, G. Stamatoyannopoulos, V. R. Iyer, J. D. Lieb, S. R. Sunyaev, J. M. Akey, P. J. Sabo, R. Kaul, T. S. Furey, J. Dekker, G. E. Crawford, and J. A. Stamatoyannopoulos
 - 2012. The accessible chromatin landscape of the human genome. *Nature*, 489(7414):75–82.

Tomizawa, S.-i., H. Kobayashi, T. Watanabe, S. Andrews, K. Hata, G. Kelsey, and

H. Sasaki

2011. Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development (Cambridge, England)*, 138(5):811–820.

- Toyooka, Y., D. Shimosato, K. Murakami, K. Takahashi, and H. Niwa 2008. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development*, 135(5):909–918.
- Trapnell, C., L. Pachter, and S. L. Salzberg 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25(9):1105–1111.
- Tsumura, A., T. Hayakawa, Y. Kumaki, S.-i. Takebayashi, M. Sakaue, C. Matsuoka, K. Shimotohno, F. Ishikawa, E. Li, H. R. Ueda, J.-i. Nakayama, and M. Okano 2006. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes to Cells*, 11(7):805–814.
- Tubio, J. M. C., Y. Li, Y. S. Ju, I. Martincorena, S. L. Cooke, M. Tojo, G. Gundem,
 C. P. Pipinikas, J. Zamora, K. Raine, A. Menzies, P. Roman-Garcia, A. Fullam,
 M. Gerstung, A. Shlien, P. S. Tarpey, E. Papaemmanuil, S. Knappskog, P. V. Loo,
 M. Ramakrishna, H. R. Davies, J. Marshall, D. C. Wedge, J. W. Teague, A. P.
 Butler, S. Nik-Zainal, L. Alexandrov, S. Behjati, L. R. Yates, N. Bolli, L. Mudie,
 C. Hardy, S. Martin, S. McLaren, S. O'Meara, E. Anderson, M. Maddison, S. Gamble, I. B. C. Group, I. B. C. Group, I. P. C. Group, C. Foster, A. Y. Warren,
 H. Whitaker, D. Brewer, R. Eeles, C. Cooper, D. Neal, A. G. Lynch, T. Visakorpi,
 W. B. Isaacs, L. v. Veer, C. Caldas, C. Desmedt, C. Sotiriou, S. Aparicio, J. A.
 Foekens, J. E. Eyfjörd, S. R. Lakhani, G. Thomas, O. Myklebost, P. N. Span, A.-L.
 Børresen-Dale, A. L. Richardson, M. V. d. Vijver, A. Vincent-Salomon, G. G. V. d.
 Eynden, A. M. Flanagan, P. A. Futreal, S. M. Janes, G. S. Bova, M. R. Stratton,
 U. McDermott, and P. J. Campbell
 - 2014. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science*, 345(6196):1251343.
- Tuschl, T., P. D. Zamore, R. Lehmann, D. P. Bartel, and P. A. Sharp

1999. Targeted mRNA degradation by double-stranded RNA in vitro. Genes & Development, 13(24):3191−3197.

van Steensel, B.

2011. Chromatin: constructing the big picture. *The EMBO Journal*, 30(10):1885–1895.

Vassetzky, N. S. and D. A. Kramerov

2013. SINEBase: a database and tool for SINE analysis. Nucleic Acids Research,41(Database issue):D83–D89.

Vincent, J., Y. Huang, P.-Y. Chen, S. Feng, J. Calvopiña, K. Nee, S. Lee, T. Le, A. Yoon, K. Faull, G. Fan, A. Rao, S. Jacobsen, M. Pellegrini, and A. Clark 2013. Stage-Specific Roles for Tet1 and Tet2 in DNA Demethylation in Primordial Germ Cells. *Cell Stem Cell*.

Vinckenbosch, N., I. Dupanloup, and H. Kaessmann 2006. Evolutionary fate of retroposed gene copies in the human genome. *Proceedings* of the National Academy of Sciences of the United States of America, 103(9):3220– 3225.

Voncken, J. W., B. A. J. Roelen, M. Roefs, S. de Vries, E. Verhoeven, S. Marino, J. Deschamps, and M. van Lohuizen

2003. Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. Proceedings of the National Academy of Sciences of the United States of America, 100(5):2468–2473.

von Meyenn, F., M. Iurlaro, E. Habibi, N. Q. Liu, A. Salehzadeh-Yazdi, F. Santos,E. Petrini, I. Milagre, M. Yu, Z. Xie, L. I. Kroeze, T. B. Nesterova, J. H. Jansen,H. Xie, C. He, W. Reik, and H. G. Stunnenberg

2016. Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells. *Molecular Cell*, 62(6):848–861.

von Meyenn, F. and W. Reik

2015. Forget the Parents: Epigenetic Reprogramming in Human Germ Cells. *Cell*, 161(6):1248–1251.

Waddington, C. H.

1942. Canalization of Development and the Inheritance of Acquired Characters.

Nature, 150:563–565.

Waddington, C. H.

1957. The strategy of the genes: a discussion of some aspects of theoretical biology. Allen & Unwin. Google-Books-ID: PdU9AAAAIAAJ.

Walsh, C. P., J. R. Chaillet, and T. H. Bestor

1998. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genetics*, 20(2):116–117.

- Walter, M., A. Teissandier, R. Pérez-Palacios, and D. Bourc'his 2016. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *eLife*, 5:e11418.
- Wang, H., W. An, R. Cao, L. Xia, H. Erdjument-Bromage, B. Chatton, P. Tempst,R. G. Roeder, and Y. Zhang

2003. mAM facilitates conversion by ESET of dimethyl to trimethyl lysine 9 of histone H3 to cause transcriptional repression. *Molecular Cell*, 12(2):475–487.

Wang, H., L. Wang, H. Erdjument-Bromage, M. Vidal, P. Tempst, R. S. Jones, and Y. Zhang

2004. Role of histone H2a ubiquitination in Polycomb silencing. *Nature*, 431(7010):873–878.

- Wang, J., G. Xie, M. Singh, A. T. Ghanbarian, T. Raskó, A. Szvetnik, H. Cai,D. Besser, A. Prigione, N. V. Fuchs, G. G. Schumann, W. Chen, M. C. Lorincz,Z. Ivics, L. D. Hurst, and Z. Izsvák
 - 2014. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature*, 516(7531):405–409.

Wang, K. and H. Chang

- 2011. Molecular Mechanisms of Long Noncoding RNAs. *Molecular Cell*, 43(6):904–914.
- Wang, R., A. B. Taylor, B. Z. Leal, L. V. Chadwell, U. Ilangovan, A. K. Robinson, V. Schirf, P. J. Hart, E. M. Lafer, B. Demeler, A. P. Hinck, D. G. McEwen, and C. A. Kim
 - 2010. Polycomb group targeting through different binding partners of RING1b C-terminal domain. *Structure (London, England: 1993)*, 18(8):966–975.

- Watanabe, T., S. Chuma, Y. Yamamoto, S. Kuramochi-Miyagawa, Y. Totoki, A. Toyoda, Y. Hoki, A. Fujiyama, T. Shibata, T. Sado, T. Noce, T. Nakano, N. Nakatsuji, H. Lin, and H. Sasaki
- 2011. MITOPLD Is a Mitochondrial Protein Essential for Nuage Formation and piRNA Biogenesis in the Mouse Germline. *Developmental Cell*, 20(3):364–375.
- Watanabe, T., A. Takeda, T. Tsukiyama, K. Mise, T. Okuno, H. Sasaki, N. Minami, and H. Imai
 - 2006. Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes & Development*, 20(13):1732–1743.
- Watanabe, T., Y. Totoki, A. Toyoda, M. Kaneda, S. Kuramochi-Miyagawa, Y. Obata, H. Chiba, Y. Kohara, T. Kono, T. Nakano, M. A. Surani, Y. Sakaki, and H. Sasaki 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature*, 453(7194):539–543.
- Watson, J. D. and F. H. C. Crick
 - 1953. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, 171(4356):737–738.
- Weber, M., I. Hellmann, M. B. Stadler, L. Ramos, S. Pääbo, M. Rebhan, and D. Schübeler
 - 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genetics*, 39(4):457–466.
- Wee, L., C. F. Flores-Jasso, W. Salomon, and P. Zamore 2012. Argonaute Divides Its RNA Guide into Domains with Distinct Functions and RNA-Binding Properties. *Cell*, 151(5):1055–1067.
- Wei, W., N. Gilbert, S. L. Ooi, J. F. Lawler, E. M. Ostertag, H. H. Kazazian, J. D. Boeke, and J. V. Moran
 - 2001. Human L1 Retrotransposition: cis Preference versus trans Complementation. Molecular and Cellular Biology, 21(4):1429–1439.
- 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 Genetics*, 41(2):246–250.

- Wicker, T., F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell,
 P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel, and A. H. Schulman
 2007. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, 8(12):973–982.
- Wightman, B., I. Ha, and G. Ruvkun 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell*, 75(5):855–862.
- Williams, K., J. Christensen, M. T. Pedersen, J. V. Johansen, P. A. C. Cloos, J. Rappsilber, and K. Helin

2011. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature*, 473(7347):343–348.

- Williams, R. L., D. J. Hilton, S. Pease, T. A. Willson, C. L. Stewart, D. P. Gearing,
 E. F. Wagner, D. Metcalf, N. A. Nicola, and N. M. Gough
 1988. Myeloid leukaemia inhibitory factor maintains the developmental potential of
 embryonic stem cells. *Nature*, 336(6200):684–687.
- Wissing, S., M. Muñoz-Lopez, A. Macia, Z. Yang, M. Montano, W. Collins, J. L. Garcia-Perez, J. V. Moran, and W. C. Greene

2012. Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility. *Human Molecular Genetics*, 21(1):208–218.

- Wolf, D. and S. P. Goff
 - 2007. TRIM28 Mediates Primer Binding Site-Targeted Silencing of Murine Leukemia Virus in Embryonic Cells. *Cell*, 131(1):46–57.
- Wolf, D. and S. P. Goff
 - 2009. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature*, 458(7242):1201–1204.
- Wong, A. K. and J. B. Rattner
 - 1988. Sequence organization and cytological localization of the minor satellite of mouse. *Nucleic Acids Research*, 16(24):11645–11661.
- Wossidlo, M., T. Nakamura, K. Lepikhov, C. J. Marques, V. Zakhartchenko, M. Boiani, J. Arand, T. Nakano, W. Reik, and J. Walter
 - 2011. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic

reprogramming. Nature Communications, 2:241.

Wray, J., T. Kalkan, and A. G. Smith

2010. The ground state of pluripotency. *Biochemical Society Transactions*, 38(4):1027–1032.

Wu, H. and Y. Zhang

2014. Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions. *Cell*, 156(1–2):45–68.

Xiong, Y. and T. H. Eickbush 1990. Origin and evolution of retroelements based upon their reverse transcriptase

sequences. The EMBO journal, 9(10):3353-3362.

- Xu, M., Y. You, P. Hunsicker, T. Hori, C. Small, M. D. Griswold, and N. B. Hecht 2008. Mice deficient for a small cluster of Piwi-interacting RNAs implicate Piwiinteracting RNAs in transposon control. *Biology of Reproduction*, 79(1):51–57.
- Yamaguchi, S., K. Hong, R. Liu, A. Inoue, L. Shen, K. Zhang, and Y. Zhang 2013. Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. *Cell Research*, 23(3):329–339.
- Yamaguchi, S., K. Hong, R. Liu, L. Shen, A. Inoue, D. Diep, K. Zhang, and Y. Zhang 2012. Tet1 controls meiosis by regulating meiotic gene expression. *Nature*, 492(7429):443–447.
- Yang, N. and H. H. Kazazian

2006. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nature Structural & Molecular Biology*, 13(9):763– 771.

Yekta, S., I.-h. Shih, and D. P. Bartel

2004. MicroRNA-Directed Cleavage of HOXB8 mRNA. Science, 304(5670):594–596.

Ying, Q.-L., J. Wray, J. Nichols, L. Batlle-Morera, B. Doble, J. Woodgett, P. Cohen, and A. Smith

2008. The ground state of embryonic stem cell self-renewal. *Nature*, 453(7194):519–523.

Yoder, J. A., C. P. Walsh, and T. H. Bestor

1997. Cytosine methylation and the ecology of intragenomic parasites. Trends in

Genetics, 13(8):335–340.

- Yokochi, T., K. Poduch, T. Ryba, J. Lu, I. Hiratani, M. Tachibana, Y. Shinkai, and D. M. Gilbert
 - 2009. G9a selectively represses a class of late-replicating genes at the nuclear periphery. *Proceedings of the National Academy of Sciences of the United States of America*, 106(46):19363–19368.
- Yoshimizu, T., N. Sugiyama, M. De Felice, Y. I. Yeom, K. Ohbo, K. Masuko, M. Obinata, K. Abe, H. R. Schöler, and Y. Matsui

1999. Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. *Development, Growth & Differentiation*, 41(6):675–684.

- Youngson, N. A., S. Kocialkowski, N. Peel, and A. C. Ferguson-Smith 2005. A small family of sushi-class retrotransposon-derived genes in mammals and their relation to genomic imprinting. *Journal of Molecular Evolution*, 61(4):481–490.
- Zanni, V., A. Eymery, M. Coiffet, M. Zytnicki, I. Luyten, H. Quesneville, C. Vaury, and S. Jensen

2013. Distribution, evolution, and diversity of retrotransposons at the flamenco locus reflect the regulatory properties of piRNA clusters. *Proceedings of the National Academy of Sciences*, P. 201313677.

- Zhang, T., S. Cooper, and N. Brockdorff 2015. The interplay of histone modifications – writers that read. *EMBO reports*, 16(11):1467–1481.
- Zhang, Y., I. A. Maksakova, L. Gagnier, L. N. van de Lagemaat, and D. L. Mager 2008. Genome-Wide Assessments Reveal Extremely High Levels of Polymorphism of Two Active Families of Mouse Endogenous Retroviral Elements. *PLoS Genet*, 4(2):e1000007.
- Zhang, Z., P. M. Harrison, Y. Liu, and M. Gerstein 2003. Millions of Years of Evolution Preserved: A Comprehensive Catalog of the Processed Pseudogenes in the Human Genome. *Genome Research*, 13(12):2541–2558.
- Zhao, B., D. Yang, J. Jiang, J. Li, C. Fan, M. Huang, Y. Fan, Y. Jin, and Y. Jin 2014. Genome-wide mapping of miRNAs expressed in embryonic stem cells and pluripotent stem cells generated by different reprogramming strategies. BMC Ge-

nomics, 15:488.

Zhao, X. D., X. Han, J. L. Chew, J. Liu, K. P. Chiu, A. Choo, Y. L. Orlov, W.-K. Sung, A. Shahab, V. A. Kuznetsov, G. Bourque, S. Oh, Y. Ruan, H.-H. Ng, and C.-L. Wei

2007. Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell*, 1(3):286–298.

- Zheng, H., B. Huang, B. Zhang, Y. Xiang, Z. Du, Q. Xu, Y. Li, Q. Wang, J. Ma,X. Peng, F. Xu, and W. Xie2016. Resetting Epigenetic Memory by Reprogramming of Histone Modifications in
- Zheng, K., J. Xiol, M. Reuter, S. Eckardt, N. A. Leu, K. J. McLaughlin, A. Stark, R. Sachidanandam, R. S. Pillai, and P. J. Wang

Mammals. Molecular Cell, 63(6):1066–1079.

2010. Mouse MOV10l1 associates with Piwi proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 107(26):11841–11846.

Zhu, J.-K.

2009. Active DNA Demethylation Mediated by DNA Glycosylases. *Annual review of genetics*, 43:143–166.

Nomenclature

2i	2 inhibitors - FGF and MEK signalling
5caC	DNA carboxycytosin
5fC	DNA formylcytosin
5hmC	DNA hydroxymethylation
$5\mathrm{mC}$	DNA methylation
AGO	ARGONAUTE
bp	base pair
cDKO	conditional double knock out
cDNA	complementary DNA
ChIP-seq	Chromatin immunoprecipitation and sequencing
CpG	CpG-Dinukleotid
CRISPR	CRISPR clustered regularly interspaced short palindromic repeats
DE	differentially expressed
DGCR8	DiGeorge syndrome critical region 8
DKO	double knock out
DMEM	Dulbecco modified Eagle's minimal essential medium
DNA	Deoxyribonucleic acid

BIBLIOGRAPHY

DNMT	DNA methyltransferase
dsRNAs	double stranded RNAs
Ε	embryonic day
EN	endonuclease
endosiRNAs	endogenous short interfering RNAs
ERVs	endogenous retroviruses
ESCs	embryonic stem cells
ETns	early transposons
GO	Gene ontology
gRNA	guide RNA
H3K27me3	Histone H3 trimethyl Lys27
H3K9me2	Histone H3 dimethyl Lys9
H3K9me3	Histone H3 trimethyl Lys9
НАТ	histone acetyltransferase
HDAC	histone deactylase complex
IAPs	intracisternal A particles
IF	Immunofluorescence
iPSC	induced-pluripotent stem cells
IRES	internal ribosomal entry sites
kb	kilobases
KD	knockdown
КО	knockout
LIF	leukaemia inhibitory factor

LINEs	long interspersed elements
LNAs	locked nucleic acids
LTRs	Long terminal repeats
MEF	mouse embryonic fibroblast
MERVL	Mouse ERV with a leucine tRNA primer-binding site elements
miRNAs	microRNAs
miscRNAs	miscellaneous other RNAs
myrs	million years
nt	nucleotides
ON	overnight
ORF	open reading frame
Р	postnatal day
PAM	protospacer adjacent motif
PBS	Phosphate Buffered Saline
PcG	polycomb group proteins
PCR	polymerase chain reaction
PGCLC	PGC like cells
PGCs	primodial germ cells
piRNAs	piwi interacting RNAs
pre-miRNA	precursor micro RNA
PTGS	post transcriptional gene silencing
qRTPCR	quantitative real time PCR
RIN	RNA integrity number

BIBLIOGRAPHY

RNA	ribonucleic acid
RNA-seq	RNA sequencing
RNAi	RNA interference
RNP	ribonucleoprotein complexes
RPKM	reads per kilobase per million mapped reads
RPM	reads per million mapped reads
rRNA	ribosomal RNA
RT	reverse transcriptase
SETDB1	SET domain containing 1 B
SINEs	short interspersed elements
siRNAs	short interfering RNAs
snoRNAs	small nucleolar RNAs
snRNAs	small nuclear RNAs
sRNAs	small RNAs
TET	methylcytosine dioxygenase
TF	Transcription Factor
TGS	transcriptional gene silencing
TSD	target site duplications
TSS	Transcription start site
TTS	Transcription termination site
UHRF1	E3 ubiquitin-protein ligase1
UTR	untranslated region
VECs	variably erased CGIs

WGBS-seq Whole genome bisulphite sequencing

WT wild type

Appendix

Namo	organism	Strain	coll type	Ganatura	librony
RNA Sog E14 d0 1 control	Muo muoouluo			Wildtupo	total DNA con
RNA_Seq_E14_d0.1_control	Mus musculus	E14 E14	ESC	Wildtype	total RNA-seq
RNA_Seq_E14_d0.2_control	Mus musculus	E14	ESC	Wildtype	total RNA seg
RNA_Seq_E14_d1.1_control	Mus musculus		ESC	Wildtype	total RNA sog
RNA_Seq_E14_d1.2_control	Mus musculus	E14 E14	ESC	Wildtype	total RNA-seq
RNA_Seq_E14_d3.1_control	Mus musculus		ESC	Wildtype	total DNA and
RINA_Seq_E14_d61_control	Mus musculus		ESC	Wildtype	total RNA-seq
RINA_Seq_E14_d6.1_control	Mus musculus	E14 E14	ESC	Wildtype	total RNA-seq
RINA_Seq_E14_d0.2_control			ESC	Wildtype	total RNA-seq
RINA_Seq_E14_d9.1_control	Mus musculus		ESC	Wildtype	total RNA-seq
RNA_Seq_E14_d9.2_control	Mus musculus	E14	ESC	Wildtype	total RNA-seq
RINA_Seq_E14_d11.1_control			ESC	Wildt we	total RNA-seq
RNA_Seq_E14_d11.2_control	Mus musculus	E14	ESC	vviidtype	total RNA-seq
RNA_Seq_E14_d17.1_control	Mus musculus	E14	ESC	vviidtype	total RNA-seq
RNA_Seq_E14_d17.2_control	Mus musculus	E14	ESC	vviidtype	total RNA-seq
RNA_Seq_E14_d25.1_control	Mus musculus	E14	ESC	Wildtype	total RNA-seq
RNA_Seq_E14_d25.2_control	Mus musculus	E14	ESC	Wildtype	total RNA-seq
RNA_Seq_E14_d0.1_DNM11KO	Mus musculus	E14	ESC	Dnmt1 KO	total RNA-seq
RNA_Seq_E14_d0.2_DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	total RNA-seq
RNA_Seq_E14_d1.1_DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	total RNA-seq
RNA_Seq_E14_d1.2_DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	total RNA-seq
RNA_Seq_E14_d3.1_DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	total RNA-seq
RNA_Seq_E14_d3.2_DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	total RNA-seq
RNA_Seq_E14_d6.1_DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	total RNA-seq
RNA_Seq_E14_d6.2_DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	total RNA-seq
RNA_Seq_E14_d9.1_DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	total RNA-seq
RNA_Seq_E14_d9.2_DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	total RNA-seq
RNA_Seq_E14_d11.1_DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	total RNA-seq
RNA_Seq_E14_d11.2_DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	total RNA-seq
RNA_Seq_E14_d17.1_DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	total RNA-seq
RNA_Seq_E14_d17.2_DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	total RNA-seq
RNA_Seq_E14_d25.1_DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	total RNA-seq
RNA_Seq_E14_d25.2_DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	total RNA-seq
Dicer KO DNMT1 KO d1	Mus musculus	E14	ESC	Dicer/Dnmt1 dKO	total RNA-seq
Dicer KO DNMT1 KO d11	Mus musculus	E14	ESC	Dicer/Dnmt1 dKO	total RNA-seq
Dicer KO DNMT1 WT d1	Mus musculus	E14	ESC	Dicer KO	total RNA-seq
Dicer KO DNMT1 WT d11	Mus musculus	E14	ESC	Dicer KO	total RNA-seq
Dicer WT DNMT1 KO d1	Mus musculus	E14	ESC	Dnmt1 KO	total RNA-seq
Dicer WT DNMT1 KO d11	Mus musculus	E14	ESC	Dnmt1 KO	total RNA-seq
WT d1	Mus musculus	E14	ESC	Wildtype	total RNA-seq
WT d11	Mus musculus	E14	ESC	Wildtype	total RNA-seq

Name	organism	Strain	cell type	Antibody	Genotyp	library
ChIP-seq_H3K9me3_d4.1_control	Mus musculus	E14	ESC	anti-H3K9me3; Active Motif, MABI 0319	Wildtype	ChIP-Sec
ChIP-seq_H3K9me3_d4.1_DNMT1KO	Mus musculus	E14	ESC	anti-H3K9me3; Active Motif, MABI 0319	DNMT1 KO	ChIP-Sec
ChIP-seq_H3K9me3_d8.1_control	Mus musculus	E14	ESC	anti-H3K9me3; Active Motif, MABI 0319	Wildtype	ChIP-Sec
ChIP-seq_H3K9me3_d8.2_control	Mus musculus	E14	ESC	anti-H3K9me3; Active Motif, MABI 0319	Wildtype	ChIP-Sec
ChIP-seq_H3K9me3_d8.1_DNMT1KO	Mus musculus	E14	ESC	anti-H3K9me3; Active Motif, MABI 0319	DNMT1 KO	ChIP-Sec
ChIP-seq_H3K9me3_d8.2_DNMT1KO	Mus musculus	E14	ESC	anti-H3K9me3; Active Motif, MABI 0319	DNMT1 KO	ChIP-Sec
ChIP-seq_H3K27me3_d4.1_control	Mus musculus	E14	ESC	anti-H3K27me3; Active Motif, 39155	Wildtype	ChIP-Sec
ChIP-seq_H3K27me3_d4.1_DNMT1KO	Mus musculus	E14	ESC	anti-H3K27me3; Active Motif, 39155	DNMT1 KO	ChIP-Sec
ChIP-seq_H3K27me3_d8.1_control	Mus musculus	E14	ESC	anti-H3K27me3; Active Motif, 39155	Wildtype	ChIP-Sec
ChIP-seq_H3K27me3_d8.2_control	Mus musculus	E14	ESC	anti-H3K27me3; Active Motif, 39155	Wildtype	ChIP-Sec
ChIP-seq_H3K27me3_d8.1_DNMT1KO	Mus musculus	E14	ESC	anti-H3K27me3; Active Motif, 39155	DNMT1 KO	ChIP-Sec
ChIP-seq_H3K27me3_d8.2_DNMT1KO	Mus musculus	E14	ESC	anti-H3K27me3; Active Motif, 39155	DNMT1 KO	ChIP-Sec
ChIP-seq_H3K9me2_d4.1_control	Mus musculus	E14	ESC	anti-H3K9me2; Abcam ab1220	Wildtype	ChIP-Sec
ChIP-seq_H3K9me2_d4.1_DNMT1KO	Mus musculus	E14	ESC	anti-H3K9me2; Abcam ab1220	DNMT1 KO	ChIP-Sec
ChIP-seq_H3K9me2_d8.1_control	Mus musculus	E14	ESC	anti-H3K9me2; Abcam ab1220	Wildtype	ChIP-Sec
ChIP-seq_H3K9me2_d8.2_control	Mus musculus	E14	ESC	anti-H3K9me2; Abcam ab1220	Wildtype	ChIP-Sec
ChIP-seq_H3K9me2_d8.1_DNMT1KO	Mus musculus	E14	ESC	anti-H3K9me2; Abcam ab1220	DNMT1 KO	ChIP-Sec
ChIP-seq_H3K9me2_d8.2_DNMT1KO	Mus musculus	E14	ESC	anti-H3K9me2; Abcam ab1220	DNMT1 KO	ChIP-Sec
Chip-seq_Input1	Mus musculus	E14	ESC	none	input	ChIP-Sec
Dicer KO DNMT1 KO d8 H3K9me3	Mus musculus	E14	ESC	anti-H3K9me3; Active Motif, MABI 0319	Dicer/Dnmt1 dKO	ChIP-Sec
Dicer KO DNMT1 WT d8 H3K9me3	Mus musculus	E14	ESC	anti-H3K9me3; Active Motif, MABI 0319	Dicer KO	ChIP-Sec
Dicer WT DNMT1 KO d8 H3K9me3	Mus musculus	E14	ESC	anti-H3K9me3; Active Motif, MABI 0319	Dnmt1 KO	ChIP-Sec
WT d8 H3K9me3	Mus musculus	E14	ESC	anti-H3K9me3; Active Motif, MABI 0319	Wildtype	ChIP-Sec
Dicer KO DNMT1 KO d8 H3K27me3	Mus musculus	E14	ESC	anti-H3K27me3; Active Motif, 39155	Dicer/Dnmt1 dKO	ChIP-Sec
Dicer KO DNMT1 WT d8 H3K27me3	Mus musculus	E14	ESC	anti-H3K27me3; Active Motif, 39155	Dicer KO	ChIP-Sec
Dicer WT DNMT1 KO d8 H3K27me3	Mus musculus	E14	ESC	anti-H3K27me3; Active Motif, 39155	Dnmt1 KO	ChIP-Sec
WT d8 H3K27me3	Mus musculus	E14	ESC	anti-H3K27me3; Active Motif, 39155	Wildtype	ChIP-Sec
Dicer KO DNMT1 KO d8 H3K9me2	Mus musculus	E14	ESC	anti-H3K9me2; Abcam ab1220	Dicer/Dnmt1 dKO	ChIP-Sec
Dicer KO DNMT1 WT d8 H3K9me2	Mus musculus	E14	ESC	anti-H3K9me2; Abcam ab1220	Dicer KO	ChIP-Sec
Dicer WT DNMT1 KO d8 H3K9me2	Mus musculus	E14	ESC	anti-H3K9me2; Abcam ab1220	Dnmt1 KO	ChIP-Sec
WT d8 H3K9me2	Mus musculus	E14	ESC	anti-H3K9me2; Abcam ab1220	Wildtype	ChIP-Sec
Chip-seq_Input2	Mus musculus	E14	ESC	none	input	ChIP-Sec

Name	
sRNA	3

Name	organism	Strain	cell type	Genotyp	library
sRNA Seg E14 d1.1 DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 d1.2 DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 d1.3 DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 d1.1 control	Mus musculus	E14	ESC	Wildtype	small RNA-seq
sRNA Seg E14 d1.2 control	Mus musculus	E14	ESC	Wildtype	small RNA-seq
sRNA Seg E14 d1.3 control	Mus musculus	E14	ESC	Wildtype	small RNA-seq
sRNA Seg E14 d3.1 DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 d3.2 DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 d3.3 DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 d3.1 control	Mus musculus	E14	ESC	Wildtype	small RNA-seq
sRNA Seg E14 d3.2 control	Mus musculus	E14	ESC	Wildtype	small RNA-seq
sRNA Seg E14 d3.3 control	Mus musculus	E14	ESC	Wildtype	small RNA-seq
sRNA Seg E14 d6.1 DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 d6.2 DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 d6.3 DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 d6.1 control	Mus musculus	E14	ESC	Wildtype	small RNA-seq
sRNA Seg E14 d6.2 control	Mus musculus	E14	ESC	Wildtype	small RNA-seq
sRNA Seg E14 d6.3 control	Mus musculus	E14	ESC	Wildtype	small RNA-seq
sRNA Seg E14 d9.1 DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 d9.2 DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 d9.3 DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 d9.1 control	Mus musculus	E14	ESC	Wildtype	small RNA-seq
sRNA Seg E14 d9.2 control	Mus musculus	E14	ESC	Wildtype	small RNA-seq
sRNA Seg E14 d9.3 control	Mus musculus	E14	ESC	Wildtype	small RNA-seq
sRNA Seg E14 d11.1 DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 d11.2 DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 d11.3 DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 d11.1 control	Mus musculus	E14	ESC	Wildtype	small RNA-seq
sRNA Seg E14 d11.2 control	Mus musculus	E14	ESC	Wildtype	small RNA-seq
sRNA Seg E14 d11.3 control	Mus musculus	E14	ESC	Wildtype	small RNA-seq
sRNA Seg E14 d25.1 control	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 d25.2 control	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 d25.3 control	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 d25.1 DNMT1KO	Mus musculus	E14	ESC	Wildtype	small RNA-seq
sRNA Seg E14 d25.2 DNMT1KO	Mus musculus	E14	ESC	Wildtype	small RNA-seq
sRNA Seg E14 d25.3 DNMT1KO	Mus musculus	E14	ESC	Wildtype	small RNA-seq
sRNA_Seq_E14_Ago2_1_control	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 Ago2 2 control	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 Ago2 1 DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA_Seq_E14_Ago2_2_DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 Ago2 none 1 control	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seg E14 Ago2 none 1 DNMT1KO	Mus musculus	E14	ESC	Dnmt1 KO	small RNA-seq
sRNA Seq PGC E13 5 F	Mus musculus	B6	PGC	Wildtype	small RNA-seq
sRNA Seg PGC E13 5 M	Mus musculus	B6	PGC	Wildtype	small RNA-seq
sRNA Seq PGC E14 5 F	Mus musculus	B6	PGC	Wildtype	small RNA-seq
sRNA_Seq_PGC_E14_5_M	Mus musculus	B6	PGC	Wildtype	small RNA-seq
sRNA_Seq_PGC_E15_5_F	Mus musculus	B6	PGC	Wildtype	small RNA-seq
sRNA_Seq_PGC_E15_5_M	Mus musculus	B6	PGC	Wildtype	small RNA-seq
sRNA_Seq_PGC_E14_PGCLC	Mus musculus	B6	PGCLC	Wildtype	small RNA-seq
sRNA_Seq_PGC_B6_PGCLC	Mus musculus	B6	PGCLC	Wildtype	small RNA-seq

Name	organism	Strain	cell type	Genotyp	library
WGBS_Seq_E14_d11.1_control	Mus musculus	E14	ESC	Wildtype	WGBS-seq
WGBS_Seq_E14_d11.1_DNMT1KO	Mus musculus	E14	ESC	DNMT1 KO	WGBS-seq
WGBS_Seq_E14_d11.2_control	Mus musculus	E14	ESC	Wildtype	WGBS-seq
WGBS_Seq_E14_d11.2_DNMT1KO	Mus musculus	E14	ESC	DNMT1 KO	WGBS-seq
WGBS_Seq_E14_d1.1_control	Mus musculus	E14	ESC	Wildtype	WGBS-seq
WGBS_Seq_E14_d1.1_DNMT1KO	Mus musculus	E14	ESC	DNMT1 KO	WGBS-seq
WGBS_Seq_E14_d1.2_control	Mus musculus	E14	ESC	Wildtype	WGBS-seq
WGBS_Seq_E14_d1.2_DNMT1KO	Mus musculus	E14	ESC	DNMT1 KO	WGBS-seq
WGBS_Seq_E14_d3.1_control	Mus musculus	E14	ESC	Wildtype	WGBS-seq
WGBS_Seq_E14_d3.1_DNMT1KO	Mus musculus	E14	ESC	DNMT1 KO	WGBS-seq
WGBS_Seq_E14_d3.2_control	Mus musculus	E14	ESC	Wildtype	WGBS-seq
WGBS_Seq_E14_d3.2_DNMT1KO	Mus musculus	E14	ESC	DNMT1 KO	WGBS-seq
WGBS_Seq_E14_d6.1_control	Mus musculus	E14	ESC	Wildtype	WGBS-seq
WGBS_Seq_E14_d6.1_DNMT1KO	Mus musculus	E14	ESC	DNMT1 KO	WGBS-seq
WGBS_Seq_E14_d9.1_control	Mus musculus	E14	ESC	Wildtype	WGBS-seq
WGBS_Seq_E14_d9.1_DNMT1KO	Mus musculus	E14	ESC	DNMT1 KO	WGBS-seq
WGBS_Seq_E14_d9.2_control	Mus musculus	E14	ESC	Wildtype	WGBS-seq
WGBS_Seq_E14_d9.2_DNMT1KO	Mus musculus	E14	ESC	DNMT1 KO	WGBS-seq

Chr	Start	End	WT Day 1	KO Day 1	WT Day 3	KO Day 3	WT Day 6	KO Day 6	WT Day 9	KO Day 9	WT Day 11	KO Day 11
6	146416935	146422725	80.89	77.27	74.14	53.03	87.50	52.78	94.12	47.14	69.05	29.93
2	104983697	104987311	84.29	71.05	68.84	44.90	73.08	55.56	77.66	39.74	69.89	32.05
5	64499421	64502595	76.74	83.33	86.36	50.00	79.41	44.83	89.91	45.00	69.35	32.29
19	56922287	56925073	88.60	63.85	77.14	63.11	78.00	48.81	78.83	56.90	65.04	32.99
10	23537467	23541618	78.90	90.12	76.24	49.65	96.67	71.43	70.33	55.56	72.62	33.33
2	166283341	166286146	90.91	79.69	100.00	58.64	88.89	52.47	93.41	54.88	60.00	33.87
4	32550292	32554491	98.48	73.08	80.49	52.56	90.74	50.00	66.67	40.78	70.48	34.08
1	136511009	136517431	84.00	84.38	80.00	55.43	80.00	44.44	89.74	47.73	86.11	34.09
3	137396025	137402318	75.86	63.83	72.70	50.00	74.07	52.63	69.05	67.31	66.67	34.68
13	30003827	30006113	73.71	91.67	90.82	76.92	80.00	50.00	96.91	52.50	73.53	35.09
5	150188861	150193174	88.49	96.34	92.39	51.04	97.62	48.85	82.08	62.00	63.54	35.77
11	75120631	75123065	93.88	88.89	76.47	56.41	81.48	41.03	89.81	43.48	76.39	36.32
10	36708807	36713946	91.11	85.59	80.77	58.12	95.24	48.61	85.00	45.24	89.58	36.42
17	12463803	12464834	88.60	76.67	76.67	64.10	100.00	70.59	89.51	67.57	57.81	36.59
8	8779014	8782988	88.54	82.11	93.75	60.61	81.08	53.85	82.56	37.50	89.32	36.67
18	58832336	58836443	86.05	88.00	92.31	63.24	80.77	69.23	90.74	57.53	83.65	36.67
6	146172324	146177909	87.93	81.20	78.33	57.21	92.11	51.47	92.89	53.03	62.82	36.90
1	53448902	53453018	79 49	61.25	48.26	41.46	55 56	45 45	68.33	41.67	54 90	36.98
2	143935173	143939667	83.67	87.93	83.67	50.00	100.00	59.09	78.33	47.50	66.09	37 10
1	88154242	88158184	78.86	88 54	95.83	67.67	97.62	48.00	87.76	47 92	60.34	37.23
11	52686117	52689371	85 74	75.56	87.88	61.90	100.00	51.32	82 41	68.18	87.96	37.41
3	9215657	9219946	81 77	76.42	84 85	52 27	87.50	45.83	89.94	46.55	77 42	37.50
4	44255014	44260541	79 44	63.93	77 15	54 69	62.22	58.06	51.56	38 43	71.59	37.71
5	130050411	130054221	91.03	70.83	76.96	58 11	78 95	47.06	83.75	67.95	67 54	37.76
2	74043264	74046761	91.00	75.52	91.07	63 70	72.58	58.33	87.23	64.00	63 51	37.01
12	107542461	107544024	06.19	01.00	00 20	56 22	100.00	54.94	06.07	57.54	74.26	20 20
10	107042401	12725002	66 10	76.00	95.42	47.91	75.06	45.24	76.02	50.00	64.02	20.20
10	102607461	102611920	00.19	70.00	00.00	47.01 EC 10	01.20	45.24	00.51	50.00	70.12	20.00
6	144405691	144409076	04.07	04.44	90.00	50.12	100.00	00.41 AE AE	90.01	60.00	70.13	20.30
17	96909704	96941267	90.00	00.07	91.00	00.10 E4.E4	00.00	45.45	97.07	50.00	69.40	20.44
0	00000794	26720442	00.07	02.21	09.13	54.51	90.00	10.00	07.50	47.20	76.00	20.40
8	30/20323	30730413	88.40	85.56	81.48	52.72	93.75	42.08	81.11	47.29	76.00	38.40
5	77500050	77500540	80.05	80.07	92.80	65.00	87.88	50.00	00.44	53.23	85.71	38.40
5	77506653	77509519	95.16	88.89	72.48	48.84	93.75	08.18	80.11	50.00	75.97	38.57
15	/825/80	7829208	89.92	82.29	70.16	54.90	88.10	07.74	73.90	37.84	72.81	38.62
5	67784830	67788235	83.72	69.35	90.48	60.94	90.48	63.64	90.54	50.00	82.27	38.82
9	45072855	45075973	90.00	/8.5/	91.67	59.46	91.67	48.89	92.42	43.33	79.55	38.89
17	23658368	23660279	79.51	81.75	80.00	67.16	94.74	54.76	72.22	36.73	73.21	39.02
17	87703650	87707729	85.94	80.37	83.33	60.94	83.33	47.83	72.22	/2./3	79.63	39.13
3	143847835	143851052	83.78	92.42	76.14	68.48	86.21	48.28	77.91	42.11	68.92	39.13
2	133359309	133362944	66.67	82.05	97.62	66.67	100.00	50.00	86.84	40.40	91.67	39.19
9	88277527	88282824	88.10	72.50	90.48	55.00	90.48	47.62	88.89	35.71	83.67	39.20
11	79667558	79670303	97.00	86.67	65.00	57.56	93.33	63.89	81.82	53.70	70.59	39.25
1	180680935	180686188	90.14	70.19	90.87	54.63	96.19	51.32	82.62	47.56	81.31	39.33
16	93441284	93444036	84.85	79.84	90.48	57.14	97.14	69.12	93.24	76.47	76.92	39.37
11	79647884	79652815	61.54	75.00	78.28	51.61	96.43	52.63	86.21	60.53	66.09	39.39
9	77284960	77288786	95.71	81.08	96.77	57.72	84.62	43.33	62.22	40.80	75.81	39.41
1	41177744	41181099	92.00	92.14	86.32	50.00	86.00	52.78	81.63	41.94	82.43	39.52
13	89714262	89719103	72.00	87.88	89.29	59.76	100.00	48.61	78.57	53.33	64.52	39.52
11	106558713	106561077	89.29	72.12	94.90	71.88	96.15	44.72	84.47	52.67	78.01	39.63

Methylation level of DMRs that resist global demethylation.

Probe	Chr	Start	End	Strand	Day 0 WT	Day 0 KO	Day 1 WT	Day 1 KO	Day 3 WT	Day 3 KO	Day 6 WT	Day 6 KO	Day 9 WT	Day 9 KO	Day 11 WT	Day 11 KO
Nrp2	1	62749859	62865269	+	2.11	2.37	2.22	1.90	2.81	3.06	1.32	3.53	2.41	3.50	2.41	2.49
Gm17672	1	95005394	95007262	-	2.08	2.19	2.54	2.77	2.93	4.55	2.66	6.05	2.38	5.72	2.49	6.15
Leftv2	1	182823239	182829234	+	4.65	4.82	5.38	5.61	5.99	6.89	4.95	6.49	5.08	6.55	6.88	7.45
Acc1	2	31325727	31376102	+	5.61	5.85	5.46	5.63	5 50	6.23	5 58	6.72	5 56	6 56	5.76	5.04
7631	-	01020121	04040000		5.01	5.05	0.40	0.00	5.55	0.20	0.00	5.00	0.00	0.00	5.70	5.04
трі	2	84273012	84316932	-	4.18	4.13	4.11	4.25	4.51	4.87	4.02	5.32	4.00	5.44	4.26	5.38
Ube2l6	2	84638985	84650492	+	-0.75	-1.20	0.29	-0.75	0.41	2.02	-0.98	2.70	-0.42	2.89	0.72	3.52
Smtnl1	2	84651333	84662809	-	-0.54	-0.62	-2.31	-1.98	-2.25	1.18	-1.27	1.66	-2.47	2.80	-2.26	2.38
Gm14149	2	151030931	151057350	+	-1.61	-0.83	-1.44	-1.41	-0.43	3.30	-0.40	4.61	-0.82	4.18	-0.59	4.79
Car2	3	14886426	14900769	+	2.96	2.96	2.69	3.07	2.95	4.28	1.48	3.68	2.05	3.97	2.72	3.78
Ccrn4l	3	51028369	51055566	+	6.02	6.06	6.43	6.62	6.54	8.19	6.59	9.30	6 42	9.18	6.65	9.62
Hormod1	2	05262500	05201502		1.67	1 00	1 20	1 90	0.07	2 70	1.69	2.61	2.16	2 94	1 27	2.60
Olaradi		50000007	50051050	+	1.07	1.00	1.35	1.00	0.97	2.70	1.00	3.01	2.10	5.04	1.27	5.09
Ctnnal1	4	56823807	56878060	-	4.71	4.73	3.88	4.12	4.39	4.92	3.25	4.90	4.78	5.62	4.57	5.33
Wdr31	4	62109687	62131930	-	2.95	2.99	2.41	2.21	2.49	3.68	2.55	4.97	2.73	4.67	2.66	5.25
Tnc	4	63620819	63681760	-	5.17	5.51	4.76	5.06	4.48	5.74	4.52	5.79	4.92	6.46	4.45	4.90
Gm13225	4	145100662	145129091	+	3.47	3.69	2.41	3.24	3.49	3.78	2.33	4.57	3.34	5.28	3.69	4.51
Gm13242	4	145104787	145294344	+	6.67	6.79	6.34	6.37	6.47	6.85	6.02	7.33	6.59	7.78	6.12	6.90
Rev2	4	145881315	145887868	+	4 87	4 82	4 33	4 84	4 72	5 19	4 15	5.72	4 99	6.21	4 51	5.40
76:600		146104005	140400450		5.04	5.04	5.00	5.05	5.02	6.00	5.24	6.72	5.70	7.40	5.40	0.10
210000	4	140124225	140100150	*	5.64	5.94	5.09	5.05	5.63	0.20	5.31	0.77	5.79	7.10	5.40	0.45
Gm13138	4	146395959	146434903	+	5.07	5.34	4.63	4.89	5.01	5.60	4.46	6.06	5.01	6.62	4.85	5.61
Gm13154	4	146927386	146959307	+	3.09	3.78	2.92	2.75	3.35	4.34	2.47	4.88	3.53	5.52	3.44	4.75
Ppp2r2c	5	37259752	37346317	+	1.12	1.56	2.41	2.35	2.73	4.48	1.33	4.36	2.28	4.30	3.44	4.85
Peg10	6	4697306	4710514	+	4.86	4.88	4.67	4.91	5.01	5.62	5.22	6.70	5.11	6.49	4.76	6.00
Gm17462	6	112387465	112389330	+	-1 61	-0.12	-1.31	0.42	-1.67	2 64	0.60	3 39	0.34	3 21	-0.39	4 02
Gm11402	6	12501400	125000000		-1.01	-0.12	-1.01	2.00	-1.07	2.04	0.00	0.00	2.70	5.07	-0.00	9.02
Empi	0	135312949	122222131	+	3.44	2.99	2.02	3.00	3.57	3.95	0.97	3.30	3.79	5.27	3.12	3.35
AU018091	7	3154665	3169204	-	3.35	3.29	2.79	2.98	3.14	4.03	2.88	4.48	3.41	4.44	4.02	4.60
Peg3	7	6656603	6683132	-	7.05	7.21	6.74	6.67	6.18	6.82	6.64	7.65	6.69	8.03	5.32	7.03
Zscan4c	7	11591071	11595896	+	0.98	0.84	-1.52	-1.41	-2.17	2.05	-1.85	3.36	-2.40	3.16	-0.59	2.83
Zscan4f	7	11983264	11988090	+	0.25	-1.20	-1.81	-1.19	-2.25	1.10	-1.85	3.22	-2.40	3.20	-1.18	2.53
Gm2783	7	104423960	104425825		0.75	-0.33	1 4 5	1.80	1.82	4 01	1 10	4.83	1 24	4 80	1 77	5 16
Son 1	- 7	124257061	124262012		2 47	4 29	4.02	2 71	4 20	4.27	2 10	4.61	4.24	4 71	4.62	4 50
Sep i		134357901	134302012	-	3.47	4.20	4.03	3.71	4.59	4.27	3.19	4.01	4.34	4.71	4.02	4.59
H19	7	149761434	149764048	-	4.93	4.94	5.20	5.24	5.63	6.94	4.99	7.08	5.81	7.66	5.69	7.34
Tnfrsf22	7	150820711	150835566	-	3.08	2.10	2.22	2.41	2.23	3.49	2.06	5.06	2.52	4.85	2.48	4.56
Mt2	8	96696518	96697467	+	4.61	4.79	4.35	4.64	4.70	5.24	3.85	5.71	4.47	5.32	4.28	5.07
Mt1	8	96702989	96704225	+	4.11	4.16	4.26	4.23	4.30	4.50	3.33	4.63	3.71	4.63	3.73	4.34
Gm3662	8	103142530	103144395	+	-0.54	-1.41	-0.44	-0.11	-0.59	2.38	-1.19	3.63	0.26	3 56	-1.26	3.41
Robo4	-	27200492	27221609		0.49	0.75	1.50	0.11	0.10	0.04	1 10	2.10	0.02	2.79	0.17	1.05
R0D04	9	37209462	3/221000	*	0.40	0.75	-1.52	-0.11	0.10	0.94	-1.19	2.10	-0.92	2.70	0.17	1.05
2410004A20Rik	9	72949643	72952250	+	2.05	1.98	1.25	1.29	1.26	2.52	1.13	4.05	1.70	4.35	1.20	3.22
SIc25a20	9	108564429	108586972	+	2.90	2.79	2.66	2.89	2.10	3.18	2.30	3.79	2.60	4.28	3.70	4.02
Slfn9	11	82793131	82805332	-	2.27	2.29	1.83	2.26	1.97	2.91	1.71	3.41	2.29	3.70	0.11	3.02
Calcoco2	11	95960640	95973278	-	2.37	2.44	1.56	1.67	1.84	2.63	-0.53	3.09	1.72	3.61	3.17	2.72
Gm4511	12	33282246	33284111	+	2 01	2.01	2.89	2.87	2 21	4 79	2.86	6.09	2 76	5.80	3.02	6 29
Maa2	10	440770244	110000000		2.01	2.01	2.00	2.07	2.2.1	0.57	2.00	4.64	2.10	5.50	0.02	6.40
wegs	12	110779211	110009930	*	2.05	2.00	2.17	2.70	2.14	2.57	2.21	4.04	2.15	5.50	2.02	0.13
Rian	12	110842155	110899919	+	2.92	3.20	2.71	3.09	2.51	3.23	0.87	6.21	2.74	7.40	1.63	7.38
Mirg	12	110973191	110987665	+	-2.19	-1.99	-2.31	-2.06	-2.17	-0.35	-1.69	1.30	-2.32	3.35	-1.76	3.48
Dnahc12	14	27644487	27704889	+	-2.11	-2.57	-1.15	-2.06	-1.59	2.16	-1.77	1.58	-1.03	1.28	-0.23	3.11
Smc1b	15	84895121	84962387	-	0.46	0.38	1.59	0.29	1.52	2.41	2.86	3.27	1.84	3.22	2.33	4.14
Rpl391	16	10170319	10175004	+	-0.45	0.09	-0.91	1.05	0.41	3.02	-0.69	3.42	0.69	3.76	0.94	3.64
Pros1	16	62854133	62020172	+	3 13	3.67	3.65	3 78	4.08	4.85	2 02	4 38	3.64	5 10	3.87	4 98
D1	10	50440740	50400004		0.10	0.07	0.00	0.70	4.00	4.00	4.70	4.00	0.04	3.13	0.07	4.50
Dazi	17	50418719	50432924	-	2.35	2.55	1.84	2.01	2.75	3.16	1.76	4.01	1.63	4.00	3.64	4.55
Impact	18	13130761	13151457	+	4.87	5.13	5.12	4.85	4.96	5.22	4.67	5.71	4.56	5.93	4.21	5.44
Gm16344	18	32181394	32186264	+	1.58	1.67	1.64	1.48	1.57	2.49	1.48	3.61	1.56	3.76	2.06	3.35
Neat1	19	5824708	5845478	-	5.54	5.34	5.84	5.78	6.01	7.06	5.82	6.73	5.94	6.54	6.27	7.01
Ahnak	19	9063750	9151409	+	6.68	6.75	6.70	6.87	7.22	7.94	6.37	7.62	7.10	8.03	7.39	7.49
Tmem20	10	38470470	38480007	+	2.86	2.02	3.16	3.07	3 31	3 34	1.80	3 77	2.66	3.86	3.24	4.00
Effect	~	46700475	46906404		0.54	1.01	4.04	0.01	1.67	0.00	1.00	0.11	1.00	2.40	0.50	2.60
EIIIGZ	^	10/091/5	10090494	-	-0.54	-1.91	-1.01	-0.91	-1.07	0.92	-1.00	2.30	-1.02	3.10	-0.52	3.00
Klhl13	х	22796397	22942208	-	5.29	5.18	4.76	4.68	5.09	6.19	4.70	6.93	4.47	7.16	4.65	6.63
Usp26	х	49107136	49154410	-	1.60	1.40	0.41	1.47	-0.30	2.99	-0.03	3.79	-0.60	3.89	0.20	4.24
Pnma5	х	70279327	70282442	-	0.92	1.54	0.56	0.79	1.08	2.55	-0.90	3.70	-0.37	4.21	0.41	4.05
XIr3a	х	70331632	70342434		-2.04	-2.49	-1.74	-2.06	-2.25	0.65	-1.77	3.22	-2.47	3.56	-2.33	4.26
Xir3b	х	70437546	70448269	+	-1.32	-1.20	-2 24	-1.98	-2.17	-1 72	-1.85	2.77	-2.32	3.31	-1.46	3.78
76/1	v	£1650	100050		3.44	1 70	2.27	3.00	1.67	1 00	1 77	0.05	2.02	1 24	0.10	2.14
∠ly i		01050	100002		-2.11	-1.70	-2.39	-2.06	-1.07	-1.60	-1.77	0.05	-2.32	1.31	-2.33	3.04
UDelyi	Y	155156	180667	+	0.16	-1.33	-1.24	-1.98	-1.67	0.68	-1.85	1.17	-1.90	1.76	-1.68	3.18
Eif2s3y	Y	346986	365290	+	3.03	3.32	3.13	2.98	2.71	3.28	2.57	3.42	1.81	4.28	0.41	4.32
Uty	Y	433304	582202	-	2.48	2.47	1.93	1.48	1.49	1.34	1.10	1.92	-0.59	2.85	-1.18	3.36
Ddx3y	Y	597158	623056	-	1.78	1.70	1.51	1.50	1.57	1.92	1.14	3.82	-0.20	3.97	-0.02	4.44
Gm3395	Y	2550262	2552957	+	-0.75	-2.49	-2.39	-1.91	-1.17	1.15	-1.27	3.00	-1.60	2.94	-0.09	2.34
						-				-						

Expression of differentially expressed genes between WT and *Dnmt1* **KO.** Expression levels as log2 RPKM levels. Upregulated upon *Dnmt1* KO.

Probe	Chr	Start	End	Strand	Day 0 WT	Day 0 KO	Day 1 WT	Day 1 KO	Day 3 WT	Day 3 KO	Day 6 WT	Day 6 KO	Day 9 WT	Day 9 KO	Day 11 WT	Day 11 KO
lgfbp2	1	72871077	72899048	+	7.08	7.21	6.95	6.95	7.12	6.30	7.11	6.04	7.26	6.15	7.75	6.81
Dnmt1	9	20711650	20757317	-	6.89	6.89	6.79	4.07	6.88	4.37	7.10	4.18	6.96	4.43	6.81	4.18
Usp28	9	48793490	48850622	+	5.49	5.32	5.47	5.69	5.57	5.15	5.84	4.98	5.81	4.68	5.26	4.97
Grb10	11	11830511	11938686	-	6.41	6.39	6.38	6.60	6.40	6.30	6.31	4.15	6.35	4.56	5.86	4.31
Otx2	14	49277354	49287319	-	3.22	3.38	3.29	3.20	4.19	2.16	3.16	1.04	3.87	1.72	3.02	1.47

Expression of differentially expressed genes between WT and *Dnmt1* KO. Expression levels as log2 RPKM levels. Downregulated upon *Dnmt1* KO.

Chr	Start	End	Day 1 WT	Day 1 KO	Day 3 WT	Day 3 KO	Day 6 WT	Day 6 KO	Day 9 WT	Day 9 KO	Day 11 WT	Day 11 KO	
12	110779211	110987665	5.73	5.97	5.56	6.12	4.86	8.41	5.81	9.25	5.07	9.64	
Х	64029932	64092300	3.94	4.09	3.89	7.00	3.79	7.58	3.21	7.42	4.20	7.73	
5	37259752	37346317	2.93	2.26	2.73	3.21	2.50	4.16	0.82	4.10	3.23	4.33	
5	92686686	92707196	5.42	6.11	5.13	6.11	4.03	7.35	5.44	7.64	3.80	5.03	
5	147546385	147606582	4.34	6.19	5.65	2.81	4.93	4.08	5.12	4.83	3.84	3.80	
6	83065490	83068892	4.41	7.11	5.99	5.04	5.19	4.72	6.26	6.22	5.14	4.82	
7	68935405	68935467	4.42	5.14	3.64	5.23	3.89	5.54	3.74	5.96	2.59	4.21	
8	96696518	96697467	4.93	5.86	5.77	6.37	2.80	5.98	4.71	5.87	4.61	5.44	
14	27512760	27704889	3.48	2.64	1.67	3.54	2.36	3.49	2.86	3.72	0.88	4.14	
17	32506250	32515390	2.44	1.68	1.43	1.70	1.98	2.22	1.54	3.08	1.79	4.87	
х	22796397	22942208	3.16	3.98	3.82	4.32	3.22	5.44	3.22	5.34	2.59	5.32	
х	100609806	100619919	2.16	2.98	1.04	3.19	2.17	4.26	2.16	4.85	0.82	1.53	

Expression of differentially expressed small RNAs between WT and *Dnmt1* KO. Expression levels as log2 RPM levels.

Probe	Chr	Start	End	Strand	Diff p-value	Dicer KO	WT d1	Probe	Chr	Start	End	Strand	Diff p-value	Dicer KO	WT d1
Zbp1	2	173032113	173044424	-	0.03	3.58	-5.03	Lin28a	4	133559245	133574756	-	0.00	-0.04	8.12
WIs	3	159502701	159598797	+	0.01	6.02	1.80	Lbh	17	73267645	73291282	+	0.04	3.51	-2.55
Vim	2	13495554	13504453	+	0.00	8.94	4.92	Lars2	9	123276058	123371782	+	0.00	8.02	8.07
VgII3	16	65815878	65863302	+	0.03	4.97	1.21	Krt17	11	100117531	100122343	-	0.00	7.85	2.78
Tnc	4	63620819	63681760	-	0.00	9.23	0.85	Kis2	х	50095740	50098137	-	0.01	-5.05	3.86
Tm4sf1	3	57089533	57105841	-	0.01	4.22	-1.55	Itgav	2	83564554	83647073	+	0.04	7.61	4.58
Timp3	10	85763117	85812251	+	0.00	7.58	2.10	Inhba	13	16103684	16119044	+	0.03	3.60	-4.88
Timp2	11	118162383	118217054	-	0.00	6.78	1.12	Gsto1	19	47929460	47939280	+	0.00	9.32	5.22
Thsd7a	6	12261610	12699410	-	0.04	2.92	-3.88	Grem2	1	176763916	176851950	-	0.00	4.70	-4.88
Thbs2	17	14802507	14831269	-	0.05	3.83	-1.17	Grem1	2	113586321	113598803	-	0.00	6.99	-1.71
Thbs1	2	117937612	117952869	+	0.00	9.82	3.08	Gm2115	7	91677485	91726847	+	0.04	3.18	-5.03
Thbd	2	148230207	148233924	-	0.01	4.38	-1.88	Ghr	15	3267760	3533492	-	0.03	5.17	1.06
Tgfbr2	9	115993415	116084383	-	0.02	5.38	0.33	Fgd3	13	49356923	49415680	-	0.05	2.95	-5.03
Tbx20	9	24527648	24578746	-	0.02	5.75	1.69	Fbn1	2	125126330	125333729	-	0.00	6.68	1.74
Тарbp	17	34052844	34066235	+	0.03	6.42	2.46	Fbln2	6	91162449	91222534	+	0.00	7.38	0.83
TagIn	9	45737711	45744141	-	0.03	7.45	4.22	Fat4	3	38785862	38910907	+	0.01	3.79	-1.88
Syn3	10	85517870	85961641	-	0.00	6.39	0.88	F5	1	166081969	166150408	+	0.01	5.38	-0.88
Slc7a3	х	98274549	98281359	-	0.00	2.01	6.72	Ereg	5	91503643	91522675	+	0.02	3.71	-4.88
Sh3pxd2a	19	47334664	47538901	-	0.01	6.88	2.24	Emp1	6	135312949	135333191	+	0.00	7.73	3.11
Sfmbt2	2	10292137	10516880	+	0.01	0.80	5.58	Egfr	11	16652206	16818161	+	0.05	5.34	0.98
Serpine1	5	137537374	137548142	-	0.00	8.31	2.73	Dppa3	6	122576428	122580290	+	0.01	-3.31	4.20
Serpinb9b	13	33119285	33133753	+	0.00	5.54	1.01	Dnmt3l	10	77504692	77526367	+	0.01	2.48	7.32
Serpinb2	1	109408000	109432055	+	0.05	2.95	-5.03	Dkk2	3	131748256	131843257	+	0.03	3.63	-5.03
Sema3a	5	13125593	13602565	+	0.03	3.68	-2.88	Ddr2	1	171902439	172040632	-	0.01	4.26	-2.07
Sdpr	1	51345970	51359803	+	0.01	4.84	-5.03	Cubn	2	13197965	13413551	-	0.03	1.58	5.55
S100a6	3	90416816	90418336	+	0.01	7.67	4.35	Csf1	3	107543966	107563387	-	0.00	5.86	-1.71
S100a4	3	90407692	90409967	+	0.00	6.68	-1.07	Col8a1	16	57624371	57754850	-	0.01	4.49	-2.55
Rnf128	х	136097855	136207684	+	0.02	6.74	3.30	Col6a3	1	92663435	92740529	-	0.02	4.63	-3.29
Rian	12	110842155	110899919	+	0.03	5.38	0.41	Col6a1	10	76171537	76188913	-	0.04	3.50	-3.29
Rab11fip5	6	85284957	85324628	-	0.05	5.24	0.65	Col5a2	1	45431166	45560127	-	0.00	6.78	-0.12
Ptgs2	1	151947254	151955140	+	0.00	6.28	0.80	Col5a1	2	27741945	27895034	+	0.00	5.85	1.08
Ptges	2	30744991	30785383	-	0.04	6.15	2.16	Col2a1	15	97806033	97835155	-	0.02	5.10	0.17
Prss23	7	96656295	96666096	-	0.00	5.91	-0.55	Col1a2	6	4454814	4491543	+	0.00	6.19	-1.07
Prrx1	1	165175247	165243841	-	0.00	5.40	-5.03	Col1a1	11	94797538	94814356	+	0.00	5.27	-1.42
Ppbp	5	91197544	91199086	+	0.03	3.32	-4.88	Col12a1	9	79446798	79566638	-	0.00	6.46	-1.55
Plau	14	21655884	21662610	+	0.05	5.04	-0.42	Cdh11	8	105156895	105309011	-	0.01	5.25	-1.29
Plagi1	10	12810497	12851500	+	0.01	6.25	1.08	Cd44	2	102651298	102741822	-	0.00	7.21	0.03
Pla2g7	17	43705047	43749150	+	0.01	6.65	2.64	Ccnd2	6	127075797	127101066	-	0.01	5.54	0.83
Phidb1	9	44494387	44543281	-	0.04	6.16	2.96	Ccdc80	16	45093515	45128037	+	0.00	5.61	-2.07
Pdzd4	х	71038698	71070308	-	0.01	1.56	6.68	Cav1	6	17256335	17291452	+	0.00	6.53	-0.35
Pcdh7	5	58109260	58520590	+	0.02	4.01	-2.55	Casq2	3	101890338	101950437	+	0.00	6.27	-0.97
Pcdh19	х	130117402	130223532	-	0.01	4.32	-0.07	Camk2d	3	126299220	126547972	+	0.01	6.51	2.32
Parva	7	119571019	119735206	+	0.00	6.37	1.78	Bgn	х	70728941	70741272	+	0.00	6.02	-2.07
Pappa	4	64785208	65018543	+	0.03	4.14	-1.55	Axi	7	26542292	26573724	-	0.01	6.00	0.58
Nrp1	8	130882973	131029362	+	0.00	5.23	-0.97	Atxn1	13	45650262	46060345	-	0.05	4.41	-1.07
Notum	11	120515102	120522489	-	0.02	6.45	2.96	Aspa	11	73118494	73140309	-	0.01	5.64	0.52
Neat1	19	5824708	5845478	-	0.01	8.65	4.38	Арр	16	84954685	85173952	-	0.04	5.11	0.55
Myof	19	37973526	38118067	-	0.00	6.29	1.17	Apob	12	7984454	8023641	+	0.00	6.08	1.62
MsIn	17	25885559	25891272	-	0.00	7.58	-0.12	Anxa3	5	97222404	97274987	+	0.01	6.69	2.17
Mical2	7	119369370	119498708	+	0.03	4.93	-2.07	Anxa1	19	20447918	20465434	-	0.01	7.12	2.99
Meg3	12	110779211	110809936	+	0.00	4.25	-2.88	Ahnak	19	9063750	9151409	+	0.05	10.22	6.88
Mal	2	127458962	127482431	-	0.01	4.20	-0.63	Actg2	6	83462899	83486259	-	0.04	3.35	-5.03
Lyz2	10	116714390	116719377	-	0.05	3.55	-3.29	Acta2	19	34315580	34329826	-	0.00	7.85	1.10
Lox	18	52675723	52689521	-	0.00	5.97	-0.79	9930013L23Rik	7	91081367	91235012	-	0.04	4.41	-1.42
								9030617O03Rik	12	102017267	102135191	+	0.04	4.79	-0.23
								8030451A03Rik	4	63640888	63810958	+	0.00	4.19	-4.88

Expression of differentially expressed genes between WT and *Dicer* KO. Expression levels as log2 RPKM levels.

Probe	Chromosome	Start	End	Strand	Dicer KO DNMT1 KO d1	Dicer KO DNMT1 KO d11	Dicer KO DNMT1 WT d1	Dicer KO DNMT1 WT d11	Dicer WT DNMT1 KO d1	Dicer WT DNMT1 KO d11	WT d1
1700019G06Rik	7	96658533	96659000	+	3.4766803	6.0712614	3.8274236	5.176032	-2.5315824	3.634632	-1.8752128
8030451A03Rik	4	63640888	63810958	+	4.194398	6.1153736	4.160347	5.389524	-5.268548	3.7097728	-4.8752127
9030617O03Rik	12	102017267	102135191	+	4.7881413	3.9706385	4.58552	4.027476	-0.47268873	2.47036	-0.2313567
9930013L23Rik	7	91081367	91235012	-	4.4145236	5.1873236	4.506481	4.9236865	-0.6571134	4.126876	-1.4157813
Acta1	8	126415669	126418651	-	4.3062115	5.772805	4.353383	4.7750425	-2.116545	3.1006064	-3.2902503
Acta2	19	34315580	34329820	-	7.8529754	10.330122	7.7120095	9.521622	0.83/0013	7.84041	1.102067
Abnak	10	0063750	0151400	-	10 223134	10.661/61	3.33903	4.//31//	-3.200340	2.430307	-3.027210
Ananti2	2	33071589	33103237	+	5 01453	7 581051	5 0481257	6 9274645	1 1872358	5 399152	0.16918124
Ankrd1	19	36186455	36194334	2	4.0931625	6.907188	4.385865	5.7372737	-1.6571134	4.55133	-2.875213
Anxa1	19	20447918	20465434		7.1240087	9.221425	7.3509855	8.576394	2.6113753	7.0928383	2,9889731
Anxa3	5	97222404	97274987	+	6.6936307	7.460073	7.063268	6.992634	2.0230064	5.167184	2.169181
Арр	16	84954685	85173952	-	5.105861	7.860513	5.3751187	7.0073857	-0.02908209	5.9432254	0.5510519
Arhgdib	6	136872044	136890420	-	3.7163508	6.3591533	3.9663599	5.7778354	-0.47268873	4.352833	-1.1747731
Aspa	11	73118494	73140309	-	5.637888	5.645762	5.7985587	6.255	-0.116545	3.4604936	0.5171045
Atxn1	13	45650262	46060345	-	4.4145236	5.146191	4.2454143	4.234135	-2.30919	2.6043732	-1.067858
AxI	7	26542292	26573724	-	5.99873	7.726063	5.8382235	6.9366593	1.241007	5.762577	0.5842188
Bgn	x	70728941	70741272	+	6.016773	7.8516827	5.7649946	7.257337	-0.6571134	5.982729	-2.067858
Cald1	6	34548500	34725473	+	7.3192954	9.066075	7.3615/23	8.307491	3.9682634	7.0276847	3.7059877
Camic2d	3	120299220	12004/9/2	÷	0.0130814	6.001000	0.3444080	0.2009/83	2.59/7006	4.901012	2.3244590
Caul	5	17256335	17201452	÷	6.532003	9.610127	6 407210	7 051167	-0.72422733	6 3506535	-0.30032210
Ccdc141	2	76847960	77008693	2	5 6686463	4 481219	5 725352	5 6786985	2 6450064	2 7780576	1 982768
Ccdc80	16	45093515	45128037	+	5.609435	8.436153	5.7577734	7.644929	-1.6571134	5.4857607	-2.067858
Ccl2	11	81849073	81850955	+	3.3352606	5.8467546	3.5501287	5.550629	-5.116545	3.7561433	-4.8752127
Ccnd2	6	127075797	127101066	-	5.538352	7.5667343	5.3615727	6.6124377	1.223305	5.7141757	0.82522684
Cd44	2	102651298	102741822	-	7.2075496	9.295553	7.2735357	8.780623	0.241007	7.531377	0.031677816
Cdh11	8	105156895	105309011	-	5.2539253	7.63865	5.399187	7.0034137	-1.0290821	5.1787663	-1.2902504
Col12a1	9	79446798	79566638	-	6.463585	8.855615	6.430665	8.022185	-1.5315824	6.1680336	-1.5532849
Col1a1	11	94797538	94814356	+	5.2652903	6.996237	5.096359	5.9016604	-0.8686175	4.0301104	-1.4157813
Col1a2	6	4454814	4491543	+	6.1943974	9.009199	6.4684386	8.209865	-1.5315824	6.388167	-1.067858
Col3a1	1	45368383	45406551	+	4.377765	7.6550965	4.975282	7.1352286	-2.116545	4.7335424	-1.4157813
Col4a2	8	11312805	11449287	+	6.6917596	8.387431	6.873012	7.621278	3.8974752	8.077445	3.534178
Col4a5	x	137909928	138123778	+	3.6369162	6.4652276	3.8155496	5.445639	-1.4161053	3.8373532	-1.8/52128
Col5a2	2	27741940	45560127	+	5.85485/4	8.434019	5.8557224	7.5004992	0.50208307	0.123710	1.0789834
Col6a1	10	76171537	76199013	÷	3 5025203	6 9642806	4 2803907	6.542087	-0.35230307	4 280500	3 2002503
Col6a3	10	92663435	92740529	2	4 631071	6 467156	4 6381536	6 15255	-3.3313020	4.266774	-3 2902503
Col8a1	16	57624371	57754850	-	4,4918094	7.353952	4.6403985	6.3754544	-2.5315824	4,7395144	-2.553285
Crim1	17	78599588	78775932	+	5.054382	6.688263	5.0832133	5.8290067	1.1313826	5.0609303	0.73949695
Csf1	3	107543966	107563387	-	5.8648534	7.4444976	5.6856766	6.656105	-0.30919003	5.585233	-1.7052878
Ctgf	10	24315248	24318489	+	7.3141375	8.751471	7.2487297	7.911232	3.869297	6.829198	4.110629
Cyp1b1	17	80106234	80114401	-	3.9533896	6.7279406	3.6867635	5.942895	-0.7946169	4.7335067	-0.7052878
Cyr61	3	145309935	145312949	-	6.466323	7.992225	6.4747443	7.3358836	3.5128117	6.5048246	3.4601774
Ddr2	1	171902439	172040632	-	4.264032	6.6744204	4.3615727	5.9420652	-2.794617	4.356517	-2.067858
Dkk2	3	131748256	131843257	+	3.6252027	5.838833	3.9079137	5.4700236	-5.116545	3.8078814	-5.027216
Dkk3	7	119259533	119302571	-	1.9596218	5.364681	2.1820886	4.7238173	-0.72422755	3.1948729	-0.17477313
Dnm3os	1	164147754	164155681	+	2.7811208	5.553182	2.7793295	4.9683867	-3.116545	3.300551	-2.553285
Dynit3	x	9231393	9240129	5	4.8/642//	7.088484	5.33963	6.476342	2.012738	5.1983366	2.019605
Egir Emp1	11	135312040	135333101	+	5.342432 7 7316804	0.3053703	5.128/08	0.057766	3.0230064	4.0/0801	3 1134717
Erap1	13	74777320	74829983	+	5 8854575	5.068723	5 782383	5.515359	2 205383	3 2304258	1 9951518
Ereq	.0	91503643	91522675	+	3 7052672	5 5373616	3 8469994	5 006989	-5 116545	3 8594427	-4 8752127
Fam198b	3	79688455	79750202	+	5.0659814	7.810507	5.1805468	7.03994	1.8141923	6.1820393	1.8390326
Fat4	3	38785862	38910907	+	3.788141	5.289664	3.2186143	4.459624	-5.116545	3.3916616	-1.8752128
Fbln2	6	91162449	91222534	+	7.3812513	10.126296	7.3928742	9.314802	0.7903456	7.916973	0.82522684
Fbn1	2	125126330	125333729	-	6.6837826	8.362265	6.7778006	7.7095027	1.0732796	6.046275	1.7394971
Fgd3	13	49356923	49415680	-	2.947131	5.4333615	3.1382723	4.0368333	-5.268548	2.76411	-5.027216
Fhl2	1	43179919	43220806	-	3.8797178	6.3084283	4.1319027	5.361247	-0.59298307	4.562293	-0.4157812
Fxyd6	9	45178268	45204241	+	4.4054217	4.5287704	4.758807	4.8740807	1.0732796	1.1662977	0.3727147
Ghr	15	3267760	3533492	÷	5.1650357	5.748699	4.8605456	5.306847	1.1502415	3.9915211	1.0555245
Gm2115	1	91677485	91/26847	+	3.1/84563	4.965191	3.0666115	4.6030188	-5.116545	3.3809943	-5.02/216
Gpnmb Cort76	6	48986517	49020928	+	3.472328	6.032/18/	3.311/24	5.356272	-1.94662	3.6754267	-1.1/4//31
Gpi176	2	112596221	112509903	-	2.0903234	4.7391333	7 1414466	0.050775	-3.200340	2.3329003	-3.027210
Grem2	2	176763916	176851950	2	4 7034116	9.530093 6.9537754	4 7608724	6.0653114	-1.30919	4 8281965	-1.7032878
Gsn	2	35111927	35163420	+	6 861113	8 966327	6 9250526	8 334935	2 7721982	6.9505873	2 3915737
Gsto1	19	47929460	47939280	+	9.323983	9.00312	9.405555	9.705413	5.1815176	7.329521	5.2162223
Hmga2	10	119798331	119913525	-	6.060193	7.5682325	5.834306	7.0958514	2.576942	6.023378	2.455704
ler5	1	156943497	156946766	-	3.6600616	5.769684	3.7793293	4.7824793	0.69080997	3.6363015	0.7686433
lfit1	19	34715361	34724499	+	1.5773741	4.4134984	1.6867634	2.4642549	-2.794617	-0.2321361	-5.027216
lgfbp7	5	77778265	77837070	-	4.609435	6.6954093	4.624607	6.105149	-0.5315824	4.807951	-0.29025042
li1ri1	1	40486415	40522260	+	3.4766803	6.357768	3.7083287	5.504158	-5.116545	4.25192	-4.8752127
Inhba	13	16103684	16119044	+	3.6014867	5.3212776	3.260091	4.646458	-3.5315826	3.719531	-4.8752127
Itgav	2	83564554	83647073	+	7.6133933	7.983045	7.4693866	7.587713	4.6565943	6.2850695	4.5801144
itgbl1	14	124059193	124374840	+	1.9964585	5.0117335	2.5596516	4.358762	-5.116545	2.418028	-3.875213
r\dHK∠ Kat17	9	215/1217	21602990	-	5.97739	5.3688126	5.9982243	5.385867	3.728945	3.5550513	3.019605
1 bh	17	73267645	73201202	+	7.049414 3.506793	5 7101303	2, 12027 / D 3, 4358454	7.0093097 5.376060	-4 116545	3.9091238	-2 553295
Lhfp	3	52845450	53065601	+	3.6562295	6.1145535	4.019083	5.0368333	-0.6571134	3.5078812	0.07898339

Expression of differentially expressed genes upon *Dicer* and *Dnmt1* KO in cluster III. Expression levels as log2 RPKM levels.

Probe	Chromosome	Start	End	Strand	Dicer KO DNMT1 KO d1	Dicer KO DNMT1 KO d11	Dicer KO DNMT1 WT d1	Dicer KO DNMT1 WT d11	Dicer WT DNMT1 KO d1	Dicer WT DNMT1 KO d11	WT d1
Lox	18	52675723	52689521	-	5.9720054	8.321633	6.0387936	7.1851788	-0.72422755	5.6620603	-0.78775
Loxi2	14	70008978	70095640	+	5.857363	8.709461	5.9282727	7.8534865	3.7691514	6.8843403	3.652264
Loxl3	6	82984167	83002556	+	4.7436924	7.484713	5.0531907	6.6930737	2.2497773	5.2099266	2.2016027
Lrrc15	16	30269388	30283342	-	2.6562295	5.5089397	2.7126033	4.5386024	-1.94662	1.7577839	-3.2902503
Ltbp1	17	75404908	75791852	+	5.160985	7.009089	5.2947173	6.5647154	1.703634	5.1488	1.21225
Lyz2	10	116714390	116719377	-	3.5487242	6.743405	4.0225306	5.7731767	-3.5315826	3.365466	-3.2902503
Mal	2	127458962	127482431	-	4.1996727	5.1669035	4.5113964	5.141749	-5.116545	1.4155917	-0.6272853
Masp1	16	23449490	23520676	-	2.7382653	5.2166567	2.7629352	4.714128	-5.268548	3.2458158	-4.8752127
Mcpt8	14	56701003	56704053	-	1.387043	4.7728057	1.0930837	4.313279	-5.268548	2.01087	-5.027216
Meg3	12	110779211	110809936	+	4.2488456	6.141369	3.987681	5.405265	-5.116545	5.1315117	-2.875213
Map	6	136820957	136824326	-	3,199673	6.138952	3.5015485	6.27359	-3.116545	4,108268	-5.027216
Mical2	7	119369370	119498708	+	4 9313636	6 7852244	4 767051	5 922846	-0 258564	5 0805316	-2 067858
Mmp10	9	7502352	7510238	+	2 7382653	4 9189763	3 0464337	4 937076	-5 268548	2 330051	-5 027216
Mmp3	9	7445822	7455972	+	3 067425	6 2845693	3 4150116	6 630064	-3 5315826	3 1265702	-5 027216
Msin	17	25885559	25891272		7 5821786	7 1806793	7 9181294	7 246623	1 3263985	4 673057	-0 12032533
Merb3	10	120218152	120336157		4 756277	6 8010563	4 762035	6 108111	0.012738076	4.520124	0.06785700
MylQ		156601156	156607394		5 6183257	7 700325	5 746352	6 0084735	1 0058220	5 404562	2 3244596
Myof	19	37973526	38118067		6 2914653	7 81076	6 268826	7 0274754	1 5699557	6 0874305	1 1691812
Ncam1		40310257	49607027		5.0234814	7.058101	5.0740364	6 262332	1.5128117	5 223442	1 2333115
Noot1	10	#3510257 E924709	43007027 E94E479	-	0.0204014	9 990799	0.0740004	9.074040	E 02066	7 0055115	4 2945205
Nedu	19	3024700	100500400	-	0.040333	0.000700	0.000000	0.0/4049	0.00000	7.2000110	4.3043303
Notum	11	120515102	120522469	-	0.4404917	4.941344	0.0020800	5.9127154	2.3912497	4.304/213	2.903991
Nrpi		130662973	131029362	+	5.2309237	7.427431	5.2240137	0.000204	-1.94002	5.3044075	-0.90832210
Pappa	4	04/85208	65018543	+	4.1432990	7.10888	3.9190543	6.00222	-3.5315820	4.540093	-1.5532849
Parva	. /	119571019	119735206	+	6.36/255/	6.862956	6.2283497	6.3949924	2.1122/3/	4.784484	1.7829987
Pcdn19	× _	130117402	130223532	-	4.3208094	5.9769683	4.0898013	5.442122	-1.65/1134	4.250357	-0.06785799
Pcdh7	5	58109260	58520590	+	4.0115337	5.6480303	3.807579	5.452648	-2.30919	3.9287648	-2.553285
Phidb1	9	44494387	44543281	-	6.1562448	6.8653955	6.1020722	6.4438815	3.14555	5.358906	2.963991
Phidb2	16	45746346	45953711	-	5.9758534	8.0889015	5.8557224	7.2896624	2.820093	6.532238	3.0376766
Pla2g7	17	43705047	43749150	+	6.6490173	6.2151275	6.6454387	6.6661906	2.4303493	5.410386	2.640487
PlagI1	10	12810497	12851500	+	6.250753	4.5213656	5.775248	5.4711742	1.6648147	4.191023	1.0789834
Plau	14	21655884	21662610	+	5.0368056	7.189661	5.248362	6.7903385	-0.16234867	5.341163	-0.4157812
Postn	3	54165031	54194961	+	3.467963	7.116193	3.7915041	6.2610016	-5.116545	4.622882	-1.7052878
Ppbp	5	91197544	91199086	+	3.3208091	5.092237	3.163473	4.276885	-5.268548	2.7301662	-4.8752127
Pmp	2	131735664	131764165	+	5.1744432	7.6306434	5.467807	6.846408	2.527311	5.7487397	2.5510519
Prrx1	1	165175247	165243841	-	5.4008484	7.558618	5.192838	7.0184503	-4.116545	5.3840413	-5.027216
Prss23	7	96656295	96666096	-	5.913818	8.608829	6.3764668	7.594858	-0.116545	6.1167765	-0.5532848
Ptges	2	30744991	30785383	-	6.149446	6.4742045	6.252772	6.7257466	2.05338	5.277693	2.15821
Ptgs2	1	151947254	151955140	+	6.2840347	8.418656	6.449996	7.499653	0.16885725	6.2070208	0.7972125
Rab11fip5	6	85284957	85324628	-	5.2437477	5.4989204	5.1414466	5.4374185	1.4380438	4.4655747	0.64834917
Rian	12	110842155	110899919	+	5.3824115	7.903869	5.651574	7.1904297	0.20538312	7.288515	0.41018927
Rnf128	х	136097855	136207684	+	6.7414336	6.248399	6.9401693	6.4934363	3.1826632	5.2767277	3.304696
Rnf213	11	119254414	119348732	+	7.0839214	6.640645	6.768079	6.6433983	4.353097	4.6702523	3.5552397
S100a4	3	90407692	90409967	+	6.6847234	9.090049	7.168538	8.566197	-1.4161053	6.57953	-1.067858
S100a6	3	90416816	90418336	+	7.667457	9.701496	8.048337	9.15658	4.133754	7.7480307	4.3487887
Sdpr	1	51345970	51359803	+	4.8431077	7.007323	5.147774	6.2676406	-2.794617	5,105732	-5.027216
Sema3a	5	13125593	13602565	+	3.6790693	5.6882634	3,7040412	5.369912	-4,116545	4,167431	-2.875213
Sema3c	5	17080099	17236086	+	4.470147	6.050833	4,4358454	5.948691	0.27577242	4,11563	-0.5532848
Serpinb2	1	109408000	109432055	+	2,9533896	3.5819411	2,9483485	3,150396	-5.268548	2.9169712	-5.027216
Serpinb9b	13	33119285	33133753	+	5 5393925	6 307711	5 6471143	6 010951	0 40701693	4 439284	1 0074302
Serpine1	.0	137537374	137548142		8 311551	10 997989	8 37495	9 9058695	2 2843344	8 576738	2 7321174
Serpinf1	11	75223271	75236203		3 621277	6 078848	3 6824114	5 6525583	-0 7946169	3 9369593	-2 553285
Sh3nyd2a	10	47334664	47538901		6 8800405	6 6821885	6 7002707	6 524258	3.0583808	4 5146027	2 2437282
Smv2	v	130442065	130466085	-	2 020504	4 331103	2 409756	3 0337/01	-5 268548	1.8454323	4 8752127
Sugdi	^	E0227666	E0461505		4.01620304	9.006420	2.403130	3.3337401	1.9403303	1.0434323	1 2540702
Susu i	4	05547030	05004044	-	4.2 1000	3.000439	4.300409	3.6/2339/	1.0492393	1.3972497	1.2340702
Syna	10	45727711	45744141	-	0.3928113	0.15/9/8	0.373095	7.420093	1.0929084	2.9084080	0.8790747
Tophp	9	45/5//11	40744141		6 4001007	9.909207	6 2045284	9.040421	3.7536195	1.394/208	4.217344
тарор	17	34032044	34000233	+	0.4201037	5.0407277	0.3943304	0.230073	2.0400004	4.41020	2.404037
TDX20	9	24527648	245/8/40	-	5.749099	4.55076	5.509112	5.3/85253	1.52/3112	3.500554	1.0940428
T gibs	12	8/39/094	8/419991	-	3.0028017	0.3202444	3.0003447	5.1942073	-2.110040	3.800057	-2.2902505
T gfbr2	9	115993415	116084383	-	5.3835707	6.640075	5.4841514	6.1242967	-0.41610524	4.6812825	0.33424056
Inbd	2	148230207	148233924	-	4.3/54363	6.46844	4.4538307	6.0/51853	-3.5315826	5.3/52146	-1.8/52128
Thbs1	2	117937612	117952869	+	9.820815	11.879068	9.783018	11.04782	2.9333034	9.534122	3.0847893
Thbs2	17	14802507	14831269	-	3.83296	6.792933	3.6246068	6.1830735	-2.30919	4.4479613	-1.1747731
Thsd7a	6	12261610	12699410	-	2.9218197	4.252877	2.9337757	3.8190978	-5.116545	2.3949075	-3.875213
Thy1	9	43851467	43856662	+	4.16773	6.9152145	3.8929245	6.679197	0.7163451	5.0628386	0.5842188
Timp1	х	20447292	20451861	+	6.3590283	8.824489	6.726938	8.245278	3.205383	6.677984	3.5468519
Timp2	11	118162383	118217054	-	6.784635	8.958123	6.594233	8.437566	1.2584944	6.9886966	1.1247871
Timp3	10	85763117	85812251	+	7.576614	9.346551	7.5239162	8.615044	1.8607349	7.1368923	2.102067
Tinagl1	4	129841698	129852366	-	4.793384	6.0516906	4.9841495	5.1886816	0.52731127	4.7473836	0.16918124
Tm4sf1	3	57089533	57105841	-	4.220582	6.748171	4.6448793	6.3292346	-5.116545	4.7236896	-1.5532849
Tnc	4	63620819	63681760	-	9.226489	11.387303	9.3748665	10.638444	0.30971974	8.893325	0.8527076
Tnfrsf11b	15	54082174	54110039	-	1.47668	4.5724187	1.7546674	3.7935627	-5.268548	1.5643276	-5.027216
Tpm1	9	66870397	66897213	-	8.718879	11.072027	8.975171	10.121779	6.4805765	8.797049	6.7135015
Tpm2	4	43527583	43536637	-	5.705267	8.388109	6.0548754	7.422324	3.6746178	5.843232	3.551052
Ubd	17	37330837	37333040	+	5.288993	5.3026805	5.4803963	5.8827558	1.6383425	2.5790944	1.7394971
Vgll3	16	65815878	65863302	+	4.973546	5.808734	4.781366	5.474622	1.0127381	3.9001508	1.21225
Vim	2	13495554	13504453	+	8.937888	11.035347	9.043891	10.518946	4.453311	9.290548	4.9175773
Wisp1	15	66722882	66754763	+	3.6752877	6.822356	3.506481	5.593538	-0.59298307	4.011144	-0.2313567
Wisp2	2	163646597	163658882	+	2.032378	5.945961	2.481649	4.828109	-5.268548	2.6474762	-5.027216
Wls	3	159502701	159598797	+	6.016773	7.7962604	6.0963593	7.018057	1.69081	5.191412	1.7972125
Wnt5a	14	29317936	29340633	+	3.286516	5.7465835	3.4202485	5.082734	-3.116545	3.5335407	-0.4157812
Zbp1	2	173032113	173044424	-	3.577374	3.2940156	3.3173485	3.7787652	-5.268548	0.39156204	-5.027216

Expression of differentially expressed genes upon *Dicer* and *Dnmt1* KO in cluster III. Expression levels as log2 RPKM levels.