

**TRANSCRIPTIONAL SILENCING OF ENDOGENOUS RETROVIRUSES
BY THE NOVEL LYSINE METHYLTRANSFERASE CO-REPRESSOR HNRNP K**

by

PETER JEFFREY THOMPSON

B.Sc., University of Alberta, 2007

M.Sc. University of Alberta, 2010

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Medical Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

December 2015

© Peter Jeffrey Thompson 2015

Abstract

Histone lysine methylation is essential for mammalian development and maintenance of somatic cell identity, as evidenced by a group of Mendelian diseases and cancers linked with mutations in lysine methyltransferases (KMTs). The transcriptional silencing of a class of retrotransposons known as endogenous retroviruses (ERVs) in murine embryonic stem cells (mESCs) provides a unique model system in which to investigate epigenetic regulation by the H3K9 family of KMTs and characterize novel molecular mechanisms of relevance to human biology and disease. In mESCs, class I and II ERVs are silenced by the SETDB1/KAP1 complex, which deposits histone H3K9 trimethylation (H3K9me3). In contrast, class III MERVL ERVs are silenced by the G9a/ GLP complex, which deposits H3K9me2. The molecular mechanisms governing the recruitment of these KMTs to their genomic ERV targets remain poorly understood. The goal of this work was to identify and characterize novel factors that regulate the functions of these KMTs in ERV silencing.

In the first part of my thesis work, I identified the RNA-binding protein and transcription factor hnRNP K as a novel co-repressor for the SETDB1/KAP1 complex. HnRNP K coordinates recruitment of the KMT SETDB1 by KAP1 to its ERV targets. This function of hnRNP K involves a previously uncharacterized influence on the levels of chromatin protein SUMOylation. In the second part of my thesis work, I demonstrated that MERVL elements are also repressed by hnRNP K and can remain inactive in the absence of H3K9me2, likely due to the lack of transcriptional activators. HnRNP K forms a novel RNA-dependent

complex with G9a/GLP, is required for global H3K9me2 and provides a repressive barrier to MERVL expression in the presence and absence of H3K9me2.

Taken together my work has provided significant insights into the epigenetic repression of ERV transcription by KMTs and demonstrates that hnRNP K is a novel co-repressor for two different KMT complexes. As recent studies have linked mutations in *HNRNPK* to the novel Mendelian disorder Au-Kline syndrome and cancer, these insights should also guide future studies on the role of hnRNP K in regulation of KMT-mediated signaling pathways in human disease.

Preface

In all of the research presented in this dissertation, I was the lead investigator responsible for the experimental design, data acquisition and analysis of the data. Where highly specialized equipment, methodology and/or expertise were necessary to complete the work, I collaborated with others who provided the relevant data and/or analyses. Miss Carol Chen performed the ChIP-seq of H3K9me2 in the mESC lines described in Chapter 4, generated track files and performed the heatmap analysis for the results in Figures 13 and 14. Dr. Karimi conducted the bioinformatic analysis for the RNA-seq data described Chapters 3 and 4 and the H3K9me2 ChIP-seq data described in Chapter 4. This contribution formed part of the results described in Figures 6, 13, 14 and 16. Dr. Foster and Miss Moon performed the mass spectrometry and related analyses described in Chapter 3, which contributed to the results shown in Figure 1. Dr. Macfarlan contributed the mouse embryonic stem cell line that was analysed in Chapter 4. The results presented in Figure 15 are based on this contribution.

Collaborators:

Dr. Mohammad M. Karimi

Bioinformatics Unit

The Biomedical Research Centre

University of British Columbia

Dr. Leonard J. Foster and Ms. Jenny Moon

Foster Lab

Proteomics, The Centre for High-Throughput Biology

University of British Columbia

Dr. Todd S. Macfarlan

Macfarlan Lab

Program in Genomics of Differentiation

Eunice Kennedy Shriver National Institute of Child Health and Human Development

National Institutes of Health

A version of Chapter 3 has been published. [Peter J. Thompson](#), Vered Dulberg, Kyung-Mee Moon, Leonard J. Foster, Carol Chen, Mohammad M. Karimi and Matthew C. Lorincz (2015) **hnRNP K coordinates transcriptional silencing by SETDB1 in embryonic stem cells**. *PLoS Genetics* 11(1):e1004933. I performed all of the major experiments described in this study including the biochemistry, RNA interference experiments, chromatin immunoprecipitation assays and expression analyses and wrote the paper with input from Dr. Lorincz. Under my supervision and design, a visiting scientist Dr. Vered Dulberg performed the supplementary experiments on MCAF1. The mass spectrometry analyses were carried out by Dr. Leonard Foster and his technician Kyung-Mee Moon and the bioinformatic analyses of the RNA-seq data was performed by Dr. Karimi.

Check the first page of the chapter to see similar notes.

Table of Contents

ABSTRACT	ii
PREFACE	iv
TABLE OF CONTENTS	vi
LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF ILLUSTRATIONS	xii
LIST OF ABBREVIATIONS	xiii
ACKNOWLEDGEMENTS	xxii
DEDICATION	xxiii
1. INTRODUCTION	1
<i>1.1 EPIGENETIC CONTROL OF THE GENOME</i>	1
<i>1.2 CHROMATIN MODIFICATION BY LYSINE METHYLATION</i>	1
<i>1.3 BIOCHEMISTRY OF MAMMALIAN KMTs</i>	4
<i>1.4 THE ROLES OF KMTs IN MAMMALS</i>	7
<i>1.5 MUTATION OF KMTs IN HUMAN DISEASE</i>	14
<i>1.6 REGULATION OF KMTs BY LONG NON-CODING RNAs</i>	17
<i>1.7 THE HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN (hnRNP) FAMILY OF RNA-BINDING PROTEINS</i>	18
<i>1.8 SUMOYLATION AND CHROMATIN REGULATION</i>	18
<i>1.9 TRANSCRIPTIONAL SILENCING OF RETROTRANSPOSONS AS A MODEL SYSTEM FOR STUDYING H3K9-SPECIFIC KMT FUNCTION</i>	19
<i>1.10 PHYLOGENY AND STRUCTURE OF RETROTRANSPOSONS IN THE MAMMALIAN GENOME</i>	20
<i>1.11 EVOLUTIONARY IMPACTS OF RETROTRANSPOSONS ON THE GENOME</i>	23
<i>1.12 MUTAGENIC MECHANISMS OF RETROTRANSPOSONS AND THEIR ROLES IN HUMAN DISEASE</i>	24
<i>1.13 DNA METHYLATION AND RETROTRANSPOSON SILENCING</i>	25
<i>1.14 TRANSCRIPTIONAL SILENCING OF RETROTRANSPOSONS BY H3K9-SPECIFIC KMTs</i>	30
<i>1.15 THESIS OBJECTIVES</i>	34
2. MATERIALS AND METHODS	38
<i>2.1 CELL LINES AND CELL CULTURE</i>	38
<i>2.2 RNA INTERFERENCE</i>	39
<i>2.3 SITE-DIRECTED MUTAGENESIS AND PLASMID TRANSFECTION</i>	40
<i>2.4 IMMUNOFLUORESCENCE AND FLOW CYTOMETRY</i>	41
<i>2.5 NATIVE AND CROSSLINKED CHIP</i>	42
<i>2.6 IMMUNOPRECIPITATION</i>	45
<i>2.7 SILVER STAINING OF SDS-PAGE GELS</i>	46
<i>2.8 WESTERN BLOT ANALYSIS</i>	46
<i>2.9 SUCROSE GRADIENT SEDIMENTATION</i>	47

2.10 MASS SPECTROMETRY OF SETDB1 COMPLEXES.....	48
2.11 RECOMBINANT PROTEINS.....	49
2.12 IN VITRO SUMOYLATION AND GST PULLDOWN ASSAYS	50
2.13 IN VITRO DE-SUMOYLATION ASSAYS.....	51
2.14 NATIVE RNA IMMUNOPRECIPITATION	51
2.15 NUCLEAR RNA PULLDOWN ASSAYS.....	52
2.16 EMSA	53
2.17 CELLULAR FRACTIONATION RNA EXTRACTION AND RT-PCR.....	54
2.18 QUANTITATIVE RT-PCR, qPCR, RNA-SEQ AND CHIP-SEQ BIOINFORMATICS	54
2.19 STATISTICS AND DATA ANALYSIS.....	55
3. HNRNP K IS A NOVEL CO-REPRESSOR FOR THE SETDB1/KAP1 COMPLEX IN PROVIRAL SILENCING	56
3.1 BACKGROUND AND SUMMARY.....	57
3.2 PROTEOMIC ANALYSIS OF SETDB1 COMPLEXES IN MESCOs	60
3.3 HNRNP K IS ASSOCIATED WITH THE SETDB1/KAP1 COMPLEX IN MESCOs	64
3.4 HNRNP K DIRECTLY INTERACTS WITH KAP1	68
3.5 DEPLETION OF HNRNP K COMPROMISES MESC SELF-RENEWAL	72
3.6 HNRNP K IS REQUIRED FOR MAINTENANCE OF SETDB1-DEPENDENT PROVIRAL REPORTER AND ERV SILENCING.....	75
3.7 HNRNP K REPRESSES A COHORT OF SETDB1-TARGETED MALE GERMLINE-SPECIFIC GENES	79
3.8 HNRNP K IS BOUND AT ERVs AND ITS DEPLETION LEADS TO REDUCED LEVELS OF H3K9ME3 AT PROVIRAL CHROMATIN	83
3.9 HNRNP K IS REQUIRED FOR SETDB1 BUT NOT KAP1 RECRUITMENT TO ERVs	86
3.10 DEPLETION OF HNRNP K PHENOCOPIES INHIBITION OF SUMO CONJUGATION, WHICH INTERFERES WITH SETDB1 RECRUITMENT AND PROVIRAL SILENCING	89
3.11 DEPLETION OF HNRNP K DOES NOT AFFECT BULK KAP1 MONO-SUMOYLATION AND HNRNP K DOES NOT DIRECTLY REGULATE KAP1 SUMOYLATION IN VITRO	92
3.12 MCAF1/MAM IS REQUIRED FOR SETDB1-DEPENDENT PROVIRAL SILENCING.....	95
3.13 DISCUSSION	98
4. ROLE OF HNRNP K, G9A/GLP AND H3K9ME2 IN SILENCING OF MERV L ELEMENTS IN MESCOs	106
4.1 BACKGROUND AND SUMMARY.....	106
4.2 MERV L ELEMENTS ARE MARKED BY H3K9ME2 AS A CONSEQUENCE OF GENOMIC LOCATION AND TRANSCRIPTIONAL INACTIVITY.....	110
4.3 HNRNP K PROMOTES GLOBAL H3K9ME2 AND REPRESSES MERV Ls IN H3K9ME2-RICH DOMAINS.....	114
4.4 HNRNP K IS REQUIRED FOR REPRESSION OF A MERV L LTR REPORTER THAT LACKS H3K9ME2	117
4.5 HNRNP K FORMS AN RNA-DEPENDENT COMPLEX WITH G9A/GLP	121
4.6 HNRNP K AND G9A/GLP ARE ASSOCIATED WITH NUCLEAR MERV L TRANSCRIPTS IN MESCOs	

<i>4.7 G400R POINT MUTATION OF THE NUCLEIC ACID BINDING KH3 DOMAIN OF HNRNP K PERTURBS MERVL SILENCING.....</i>	131
<i>4.8 DISCUSSION</i>	134
5. CONCLUDING REMARKS AND FUTURE DIRECTIONS.....	143
<i>5.1 KMT-DEPENDENT SILENCING OF ERVs INVOLVES THE CO-REPRESSIVE ACTIVITIES OF HNRNP K.....</i>	143
<i>5.2 HNRNP K MUTATIONS IN MENDELIAN DISEASE LINKS WITH G9A/GLP.....</i>	146
<i>5.3 ROLE OF HNRNP K IN KMT-DEPENDENT GENE SILENCING IN CANCER</i>	151
BIBLIOGRAPHY	154

List of Tables

Table 1 List of PCR primers used in this study.....	43
Table 2 RNA-seq fold-changes of a subset of the SETDB1-targeted tissue-specific genes upregulated in common in <i>Setdb1</i> KO and <i>Hnrnpk</i> KD cells.....	81
Table 3 A comparison of major clinical features in Au-Kline syndrome and Kleefstra syndrome caused by mutations in <i>EHMT1</i>	151

List of Figures

Figure 1 Proteomic analysis of SETDB1-associated proteins in mESCs	63
Figure 2 hnRNP K interacts with the SETDB1/KAP1 complex in mESCs	67
Figure 3 hnRNP K directly interacts with KAP1	71
Figure 4 Depletion of hnRNP K abrogates mESC self-renewal	74
Figure 5 hnRNP K is required for SETDB1-dependent silencing of proviral reporters and ERVs	78
Figure 6 hnRNP K co-represses genes with SETDB1 and KAP1	82
Figure 7 hnRNP K is required for H3K9me3 at proviral chromatin	85
Figure 8 hnRNP K is required for SETDB1 recruitment to ERVs and is recruited in a KAP1-dependent manner	88
Figure 9 Depletion of hnRNP K leads to loss of SUMOylation on ERV chromatin, which is necessary for SETDB1 recruitment and proviral silencing	91
Figure 10 hnRNP K depletion does not result in loss of bulk KAP1 mono- or di-SUMOylation and hnRNP K does not directly regulate KAP1 SUMOylation <i>in vitro</i>	94
Figure 11 MCAF1 is required for SETDB1-dependent proviral silencing.....	97
Figure 12 Revised model for SETDB1/H3K9me3-dependent proviral silencing incorporating the functions of hnRNP K and MCAF1	101

Figure 13 MERVL elements are marked by H3K9me2 as a consequence of genomic location and transcriptional inactivity.....	113
Figure 14 hnRNP K represses MERVL via H3K9me2-dependent and independent mechanisms.....	116
Figure 15 hnRNP K is required for repression of a newly integrated MERVL LTR reporter lacking H3K9me2	120
Figure 16 hnRNP K forms an RNA-dependent complex with G9a/GLP	123
Figure 17 <i>G9a</i> KO cells do not show changes in hnRNP K localization or protein levels but have increased SETDB1 binding at ERVs.....	125
Figure 18 hnRNP K and G9a/GLP are associated with nuclear MERVL transcripts.....	128
Figure 19 Comparison of the nucleic acid binding activities of wt and Δ -C mutant hnRNP K proteins.....	130
Figure 20 G400R point mutation of the nucleic acid binding KH3 domain of hnRNP K perturbs G9a-dependent MERVL silencing	133
Figure 21 Revised model for MERVL silencing incorporating the repressive effects of hnRNP K and G9a/GLP	137

List of Illustrations

Illustration 1 Chromatin regulation by lysine methylation in mammals.....	3
Illustration 2 Repetitive elements in the human genome and structure of retrotransposons	21
Illustration 3 DNA methylation, its oxidation and reprogramming during embryonic development.....	27
Illustration 4 Current model for transcriptional silencing of class I/II ERVs by the SETDB1/KAP1 complex	60
Illustration 5 Current model for transcriptional silencing of class III MERVL elements by G9a/GLP	107
Illustration 6 Proposed role of hnRNP K in human development with the G9a/GLP and SETDB1/MCAF1 KMT complexes	150

List of Abbreviations

2i	GSK3 β and MEK1/2 inhibitors
4-OHT	4-hydroxy tamoxifen
5-aza-dC	5-aza-2'-deoxycytidine
5caC	5-carboxyl cytosine
5fC	5-formyl cytosine
5hmC	5-hydroxymethyl cytosine
5mC	5-methyl cytosine
A	Adenine
aCGH	Array comparative genomic hybridization
Actb	Beta actin
AEBP2	Adipocyte enhancer-binding protein 2
AGO2	Argonaute 2
ASH1L	Absent small and homeotic disks protein 1 homolog
ASH2L	Absent small and homeotic disks protein 2 homolog
Atf7ip	ATF7-interacting protein
ATRX	α -thalassemia, X-linked mental retardation protein
BAF155	Brahma associated factor 155 kDa
Blm	Bloom syndrome protein
BTB/POZ	BR-C, TTK BAB/Pox virus and zinc finger
C	Cytosine
CaCl ₂	Calcium chloride
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation - massively parallel sequencing
Clr4	Cryptic loci regulator 4
CML	Chronic myeloid leukemia

Cml2	Camello-like 2
Co-IP	Co-immunoprecipitation
COMPASS	Complex of Proteins Associated with Set1
CoREST	REST co-repressor 1
CpG	Cytosine-phosphate-guanine
CSF1R	Colony stimulating factor 1 receptor
Dazl	Deleted in azoospermia-like
dbSNP	NCBI database of single nucleotide polymorphisms
DDX21	DEAD (Asp-Glu-Ala-Asp) box helicase 21
DMEM	Dulbecco's minimal essential media
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNase I	Deoxyribonuclease 1
DNMT1	DNA methyltransferase 1
DNMT3A	DNA methyltransferase 3A
DNMT3B	DNA methyltransferase 3B
DNMT TKO	DNA methyltransferase (1, 3A, 3B) triple knockout
Dot1	Disruptor of telomeric silencing
DOT1L	Disruptor of telomeric silencing like
DTT	Dithiothreitol
EED	Embryonic ectoderm development
Egr1	Early growth response 1
EHMT1	Euchromatic histone methyltransferase 1 (GLP)
eIF4E	Eukaryotic translation initiation factor 4E
ERV	Endogenous retrovirus
ESET	ERG-associated protein with SET domain
EWSAT1	Ewing sarcoma associated transcript 1

EZH1	Enhancer of zeste, homologue 1
EZH2	Enhancer of zeste, homologue 2
ETn	Early transposon
Fkbp6	FK506 binding protein 6
G	Guanine
gag	Group-specific antigen
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
Gata3	GATA binding protein 3
Gata6	GATA binding protein 6
G9a	Euchromatic histone methyltransferase 2
GLP	G9a like protein 1
GSK3 β	Glycogen synthase kinase 3 beta
H2AK119	Histone H2A Lysine 119
H3	Histone H3
H3K4me1	Histone H3 Lysine 4 monomethylation
H3K4me2	Histone H3 Lysine 4 dimethylation
H3K4me3	Histone H3 Lysine 4 trimethylation
H3K9me1	Histone H3 Lysine 9 monomethylation
H3K9me2	Histone H3 Lysine 9 dimethylation
H3K9me3	Histone H3 Lysine 9 trimethylation
H3K27me1	Histone H3 Lysine 27 monomethylation
H3K27me2	Histone H3 Lysine 27 dimethylation
H3K27me3	Histone H3 Lysine 27 trimethylation
H3K36me2	Histone H3 Lysine 36 dimethylation
H3K36me3	Histone H3 Lysine 36 trimethylation
H3S10	Histone H3 Serine 10
H4K20me2	Histone H4 Lysine 20 dimethylation

H4K20me3	Histone H4 Lysine 20 trimethylation
HCl	Hydrochloric acid
HDAC1	Histone deacetylase 1
HDAC2	Histone deacetylase 2
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hnRNP K	Heterogeneous nuclear ribonucleoprotein K
hnRNP A/B	Heterogeneous nuclear ribonucleoprotein A/B
HOTAIR	HOX transcript antisense RNA
HOTTIP	HOXA transcript at the distal tip
HOXA	Homeobox A
HOXC	Homeobox C
HOXD	Homeobox D
HP1	Heterochromatin protein 1
HP1 α	Heterochromatin protein 1 alpha
HP1 β	Heterochromatin protein 1 beta
IAP	Intracisternal A-type particle
Igf2	Insulin-like growth factor 2
Igf2r	Insulin-like growth factor 2 receptor
ILF2	Interleukin enhancer-binding factor 2
IP	Immunoprecipitation
IP-MS	Immunoprecipitation and mass spectrometry
JARID2	Jumonji/ARID domain-containing protein 2
KAP1	Krüppel-associated box associated protein 1
KCl	Potassium chloride
Kcnq10t1	Kcnq1 overlapping transcript 1
KD	Knockdown
kDa	Kilo Daltons

KDM	Lysine demethylase
KDM1A	Lysine (K)-specific demethylase 1A
KH	heterogeneous nuclear ribonucleoprotein K homology
KI	heterogeneous nuclear ribonucleoprotein K interactive
Kid1	Kidney, ischemia and developmentally regulated gene 1
KMT	Lysine methyltransferase
KO	Knockout
KRAB-ZFP	Krüppel-associated box zinc finger protein
lncRNA-p21	Long noncoding RNA associated with p21
LSD1	Lysine specific demethylase 1
LIF	Leukemia inhibitory factor
LINE1	Long interspersed nuclear element 1
LINE2	Long interspersed nuclear element 2
LTR	Long terminal repeat
Leu	Leucine
Mael	Maelstrom spermatogenic transposon silencer
mAM	Murine ATF-associated modulator
MBD	Methyl-CpG-binding domain
MBD1	Methyl-CpG-binding domain 1
MBD5	Methyl-CpG-binding domain 5
MCAF1	MBD1-containing Chromatin Associated Factor 1
MCM2-7	Minichromosome maintenance complex component 2 through 7
MCM5	Minichromosome maintenance complex component 5
MERVL	Mouse endogenous retrovirus with a leucine tRNA primer binding site
MEK1/2	Mitogen-activated protein kinase kinases 1 and 2
MIR	Mammalian-wide interspersed repeat
MLV	Moloney murine leukemia virus

MLL1	Mixed lineage leukemia 1
MLL2	Mixed lineage leukemia 2
MLL3	Mixed lineage leukemia 3
MLL4	Mixed lineage leukemia 4
MMERVK10C	<i>Mus musculus</i> endogenous retrovirus K subfamily 10C
MNase	Micrococcal nuclease
mRNA	Messenger RNA
MSCV	Murine stem cell virus
MTA	Mouse transposon A
MusD	Mouse type D retrotransposon
N	Any base in DNA (A, C, T or G)
NEM	N-ethylmaleimide
NfκB	Nuclear factor kappa B
Nkx2-9	NK2 homeobox 9
NMD	Nonsense mediated decay
NOLC1	Nucleolar and coiled-body phosphoprotein 1
NP-40	Nonidet P-40 (octyl phenoxy polyethoxy ethanol)
NR3L1	Nuclear receptor subfamily 1, group 1, member 3
NSD1	Nuclear receptor binding SET domain protein 1
NSD2	Nuclear receptor binding SET domain protein 2
NSD3	Nuclear receptor binding SET domain protein 3
Nup155	Nuclear pore complex protein 155 kDa
ORF	Open reading frame
p53	Tumour protein p53
PAPBC1	Poly A binding protein cytoplasmic 1
PBS	Primer binding site
PCR	Polymerase chain reaction

PCL1	Polycomb like 1
PCL2	Polycomb like 2
PCL3	Polycomb like 3
PHD	Plant homeodomain
pol	Retroviral reverse transcriptase
Pol II	RNA polymerase II
Pol III	RNA polymerase III
Pou5f1	POU class 5 homeobox 1
PPi	Phosphatase inhibitor cocktail
PRC1	Polycomb repressive complex 1
PRC2	Polycomb repressive complex 2
PRDM3	Positive regulatory domain zinc finger protein 3
PRDM16	Positive regulatory domain zinc finger protein 16
PRMT1	Protein arginine methyltransferase 1
PRMT5	Protein arginine methyltransferase 5
Pro	Proline
PxVxL	Proline-x-Valine-x-Leucine
qPCR	Quantitative PCR
qRT-PCR	Quantitative reverse transcriptase PCR
RBBP5	Retinoblastoma binding protein 5
RBCC	Really Interesting New Gene, B-box, Coiled-coil
RGG	Arginine-Glycine-Glycine
rRNA	Ribosomal RNA
RNA	Ribonucleic acid
RNAi	RNA interference
RNAse	Ribonuclease
S	Svedberg units

SAE1	SUMO1 activating enzyme 1
SAE2	SUMO1 activating enzyme 2
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEN1	Sentrin (SUMO)-specific protease 1
SEN7	Sentrin (SUMO)-specific protease 7
Ser	Serine
SET	Suppressor of variegation 3-9, Enhancer of zeste, Trithorax
SETD1A	SET domain-containing 1A
SETD1B	SET domain-containing 1B
SETD2	SET domain-containing 2
SETD5	SET domain-containing 5
SETD8	SET domain-containing 8
SETDB1	SET domain, bifurcated 1
SETDB2	SET domain, bifurcated 2
SETMAR	SET domain and mariner transposase fusion
siRNA	Small interfering RNA
Slc22a2	Solute carrier family 22 member 2
Slc22a3	Solute carrier family 22 member 3
SMARCB1	Swi-snf related matrix associated actin dependent regulator of chromatin subfamily b, member 1
SSEA1	Stage specific embryonic antigen 1
SUMO	Small ubiquitin-like modifier
SUMO1	Small ubiquitin-like modifier 1
SUMO2	Small ubiquitin-like modifier 2
SUV39H1	Suppressor of variegation 3-9, homolog 1
SUV39H2	Suppressor of variegation 3-9, homolog 2
SUV420H1	Suppressor of variegation 4-20, homolog 1
SUV420H2	Suppressor of variegation 4-20, homolog 2

T	Thymine
TAF7L	TATA-binding protein associated factor 7-like
TET1	Ten-eleven translocation 1
TET2	Ten-eleven translocation 2
TET3	Ten-eleven translocation 3
TF	Transcription factor
TOP2A	Topoisomerase II alpha
TRIM28	Tripartite Motif-containing 28
TRIP12	Thyroid hormone receptor-interacting protein 12
Tris	Tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
Trr	Trithorax related (<i>Drosophila</i>)
Tunar	TCL1 upstream neural differentiation-associated RNA
U	Uracil
Ubc9	Ubiquitin carrier protein 9
VNTR	Variable number of tandem repeats
WDR5	WD repeat-containing protein 5
Wiz	Widely interspaced zinc fingers
Xist	X-inactive specific transcript
ZFP161	Zinc finger protein 161
ZFP809	Zinc finger protein 809 (mouse)
ZFP819	Zinc finger protein 819 (mouse)
ZNF274	Zinc finger protein 274 (human)
ZNF644	Zinc finger protein 644 (human)
Zik1	Zinc finger protein interacting with K-protein
Zscan4	Zinc finger and SCAN domain protein

Acknowledgements

Many people provided guidance and support during my thesis work and I am ever in their debt. First, I thank my supervisor and mentor Matt Lorincz for his guidance along the course of my program. He constantly encouraged and challenged me with insightful ideas and always provided a stimulating intellectual environment, giving me the tools I needed to become an independent researcher. I thank my supervisory committee members Leonard Foster, Louis Lefebvre and Marco Marra for their input and helpful feedback on my work and for collaborative assistance. I also thank my lab members Preeti Goyal, Sheng Liu, Julien Albert and especially Julie Brind'Amour, Irina Maksakova and Carol Chen for technical assistance with mESC proviral reporter lines and next-generation sequencing data and analysis. I thank Jacob Hodgson for technical assistance with the biochemistry in my project. I also thank my colleagues in the Molecular Epigenetics Group who have helped to shape my ideas, given me the opportunity to present my work and have provided technical assistance and support. Lastly, I thank my beautiful wife and mother of my children, Adele, who has patiently journeyed with me over these years, providing me with encouragement and steady support.

Dedication

Soli Deo gloria,

Christi crux est mea lux!

For my dear sons, Callan James and Isaac Peter

1. Introduction

1.1 Epigenetic control of the genome

Information in the form of chemical modifications of DNA and histones, the positioning and composition of nucleosomes and the occupancy of regulatory proteins and RNA species on the chromatin fibre is crucial for organism development, cell identity and function. This information is epigenetic (literally ‘above the genetic’ reviewed by Bird¹) and dictates the expression, replication and repair of the genome in a manner that does not alter the underlying DNA sequence, thus providing a dynamic medium through which organisms can sense, remember and adapt to changes in their environment (reviewed by Berger *et al.*²). In addition to being stable through mitotic division and thus serving as the basis for memory of cellular state and identity, a growing body of evidence suggests that epigenetic information may be transmitted across generations in mammals³ (reviewed by Heard⁴). Moreover, the dysregulation of epigenetic systems is a feature of particular Mendelian diseases, including congenital malformation syndromes, imprinting disorders and some forms of malignancy. Therefore, a better understanding of epigenetic regulation of the genome is critical for the prevention and treatment of human disease.

1.2 Chromatin modification by lysine methylation

In eukaryotes, nuclear DNA is wrapped around an octamer of core histones H2A, H2B, H3 and H4 to form the nucleosome particle,^{5,6} which is stabilized by the linker histone H1 to form chromatin (reviewed by Cutter and Hayes⁷). Histone proteins can be post-translationally modified in many different ways including lysine or arginine methylation, lysine acetylation, serine or threonine phosphorylation, lysine ubiquitination, proline isomerization, arginine

citrullination, glutamine methylation and lysine SUMOylation (reviewed by Tessarz and Kouzarides⁸). A growing body of evidence from model organisms and humans has consistently demonstrated that histone modifications impart a complex information system to nuclear signalling pathways, which translates into a variety of distinct functional states in the genome.⁸

Histone lysine methylation, which is conserved from yeast to humans, is one of the most extensively studied epigenetic marks. This chromatin modification is catalyzed by a diverse group of KMTs that can specifically target different lysines. Conversely, removal of these marks is catalyzed by KDMs (**Illustration 1A**). The vast majority of histone lysine methylation in mammalian cells occurs on the histone H3 and H4 N-terminal tails,⁸ which protrude from the nucleosome. Lysines may be mono-, di- or trimethylated predominantly at positions H3K4, H3K9, H3K27, H3K36 and H4K20 (**Illustration 1B**). Such marks correlate with (if not confer) a variety of transcriptional and structural states depending on the position of the lysine, location in the genome, number of methyl groups, presence of methyl-lysine binding proteins and the co-occurrence of other chromatin modifications. For instance, early genomewide mapping studies using ChIP-seq in mESCs, mouse embryonic fibroblasts (MEFs) and neural progenitors showed for the first time that H3K4me3 and H3K36me3 are associated with transcriptionally active genes, while H3K9me2/3, H4K20me3 and H3K27me3 are associated with transcriptional repression at distinct loci.^{9,10} Notably, the dynamics of these modifications change during cell differentiation.^{9,10} Unlike lysine acetylation, methylation does not neutralize the positive charge on the epsilon amino group

(**Illustration 1A**) and thus predominantly exerts its effects through crosstalk or inhibition of other histone modifications and recruitment or exclusion of methyl-lysine-binding proteins.

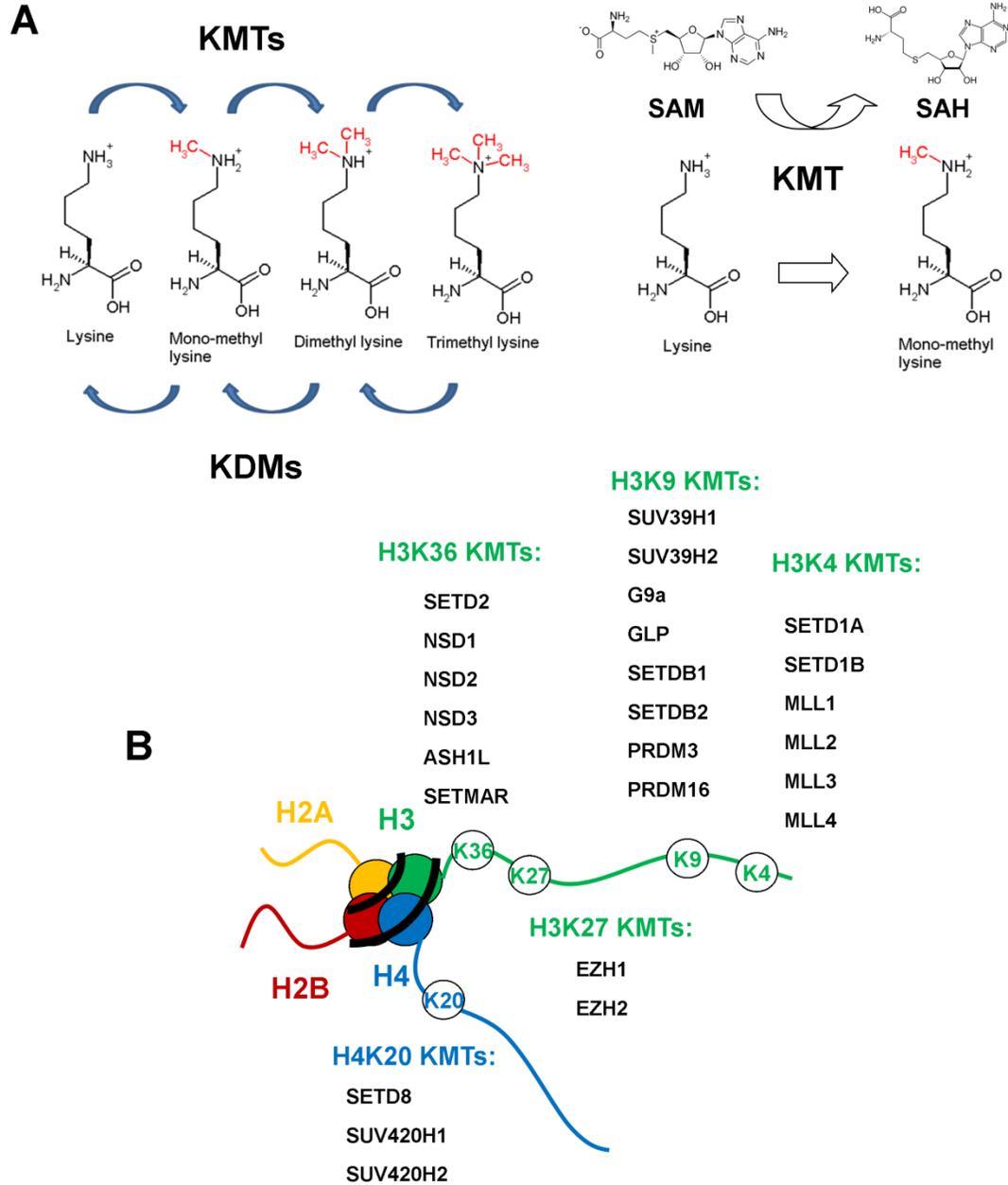


Illustration 1. Chromatin regulation by lysine methylation in mammals. (A) Left panel: Iterative catalysis of lysine methylation by KMTs and demethylation by lysine demethylases (KDMs). Right panel: The basic catalytic mechanism of KMTs utilizes SAM as a co-substrate to transfer the methyl group to the epsilon amino group on the lysine, producing SAH as a by-product. (B) Schematic of a core nucleosome particle, showing the H2A, H2B, H3 and H4 histone subunits and their N-terminal tails. The N-terminal tails of H3 at K4, K9 K27 and K36 and H4K20 are the major sites subjected to mono-, di- and trimethylation by the SET domain-containing KMTs indicated.

1.3 Biochemistry of mammalian KMTs

All known KMTs fall into one of two classes: those with a highly conserved SET catalytic domain named after its founding members identified in *Drosophila* (*Su(var)3-9*, *Enhancer of zeste*, *Trithorax*) and KMTs without a SET domain (reviewed by Herz *et al.*¹¹). Among the SET domain family, a distinct class of SET domain KMTs termed SET and Myeloid-Nearvy-DEAF1 domain (SMYD) possess a MYND domain within the SET domain, which will not be discussed further here (reviewed by Spellmon *et al.*¹²). The second class of KMT includes only one member, DOT1L in mammals and Dot1 in yeast, which is active towards H3K79.¹³ The focus of my thesis is on KMTs in the SET domain family.

The SET domain is ~130 amino acids in length and consists of regulatory modules called pre-SET and post-SET.¹¹ Recent analysis of the human genome suggests there are ~60 annotated SET domain-containing proteins, most of which have not been characterized biochemically or genetically (reviewed by Binda¹⁴). For their catalytic activity, all KMTs require the co-substrate S-adenosyl L-methionine (SAM), which is generated from metabolism of the amino acid methionine. A methyl group from SAM is transferred to the epsilon amino group on the lysine residue in a 1:1 stoichiometry, leaving the product S-

adenosyl homocysteine (SAH) (**Illustration 1A**). *In vitro* enzymatic assays indicate that the methylated histone and SAH products can competitively inhibit SET domain activity.^{15,16} Site-specificity for SET domains includes several residues surrounding the targeted lysine¹⁵ although some SET domains and recombinant full-length KMTs exhibit promiscuity in their target sites during *in vitro* assays.¹¹ Further evidence indicates KMTs are active towards histone-like epitopes in non-histone substrates,¹⁷ suggesting that there are consensus residues for catalysis. KMT activity towards non-histone substrates greatly increases the complexity of signaling information conferred by these enzymes. Although many non-histone substrates have been identified for specific KMTs, the effects of methylation on most non-histone proteins remains to be determined.¹¹ In several cases the methylation of non-histone proteins can regulate the association of binding partners, exert effects on the catalytic activity, or alter the chromatin/DNA binding of the target protein.¹¹ As would be predicted based on their site specificity, methylation of different lysines on the same histone tail, or the presence of other histone modifications can specifically block the activity of KMTs for their lysine target. For example, methylation of H3K4 or phosphorylation of H3S10 can inhibit H3K9 methylation.^{18,19} Although purified SET domains generally possess activity towards recombinant or bulk histones *in vitro*, their specificity and activity towards nucleosomal targets both *in vitro* and *in vivo* is often potentiated by their accessory domains and protein interaction partners.^{11,14}

Most KMTs are known to function in conserved multi-protein assemblies, which serve to enhance catalysis, promote substrate specificity, enhance chromatin binding and otherwise regulate the stability of the KMT.¹¹ For instance, the H3K4-specific family of KMTs

including MLL1-4 interact in a large multi-protein complex called COMPASS in yeast or COMPASS-like in mammals, including the core subunits ASH2L, WDR5 and RBBP5.¹¹ ASH2L and WDR5 are required for global H3K4 di- and trimethylation in mESCs.^{20,21} Similarly, the H3K27-specific KMTs EZH1/2 interact in the PRC2 complex, containing EED, SUZ12, AEBP2, JARID2, PCL1/2/3 and RBBP4/7 in mammals.¹¹ EED and SUZ12 are required for global H3K27me3²² and EED promotes EZH2 catalytic activity toward H3K27me3.²³ AEBP2, JARID2, and PCL proteins are thought to regulate PRC2 targeting and also enhance catalytic activity (reviewed by Margueron and Reinberg²⁴).

H3K9-specific KMTs G9a and GLP interact in a heteromeric complex²⁵ that includes the DNA-binding zinc finger proteins Wiz and/or ZNF644/Zfp644.^{26,27} Wiz controls G9a protein stability²⁷ and both Wiz and ZNF644 can target G9a/GLP to distinct target genes by direct DNA binding.²⁶ The KMTs SUV39H1/2 and SUV420H1/2 form complexes with HP1,²⁸⁻³⁰ which bind the H3K9me3 mark.^{31,32} SETDB1 forms heteromeric complexes with MCAF1 (also called ATF7IP) which enhances its catalytic activity.³³ These H3K9 KMTs have also been identified in a large multimeric assembly consisting of SETDB1, SUV39H1/2, G9a, GLP and HP1 proteins,^{34,35} however the functional significance of this complex remains unclear. In addition to the core subunits in these KMT complexes, other accessory binding partners have been shown to enhance catalysis, promote genomic recruitment via other histone or DNA modifications and/or serve as adaptors to transcription factors or RNA molecules to enable sequence-specific targeting.

1.4 The roles of KMTs in mammals

The roles of specific KMTs were first identified well before it was known that they could actually methylate histones. Forward mutagenesis screens in fruit flies identified polycomb group (PcG) and Trithorax group (TrxG) genes, which when mutated resulted in homeotic transformations in the embryo as a result of transcriptional de-repression or loss of transcriptional activation of the *Hox* genes, respectively, in the embryonic segments (reviewed by Grimaud *et al.*³⁶). These genes were later identified as either encoding KMTs or KMT binding partners and were shown to promote methylation on H3K4 or H3K27, respectively.³⁶ Strikingly orthologues of the genes were subsequently shown to play conserved roles in mammalian body plan organization, underscoring the crucial role of methylation at H3K4 and H3K27 for metazoan development. In addition, homeotic transformation phenotypes, the *Drosophila* eye provided a unique model system to identify mutations which either enhanced or suppressed position-effect variegation (PEV). In classic PEV, the *white* gene (whose mutant phenotype results in white eye colour) was differentially silenced or expressed in ommatidia comprising the eye, depending upon stochastic expansion or contraction of heterochromatin into this gene, a phenomenon influenced by the levels of chromatin regulatory proteins in the cells.³⁷ Enhancers of PEV were genes that normally promoted an open, transcriptionally active chromatin structure, while Suppressors of PEV were genes necessary for heterochromatin formation and/or spreading, including HP1 and Suv3-9, a KMT specific for H3K9.³⁷

The finding that KMTs targeting H3K4 and H3K36 are found in lower organisms such as yeast suggests a fundamental importance of methylation at these residues in the processes

shared by uni- and multicellular organisms. In contrast, while H3K9 methylation occurs in the fission yeast *Schizosaccharomyces pombe*, the budding yeast *Saccharomyces cerevisiae* lack H3K9 methylation, indicating that this mark is dispensable for some unicellular organisms. In contrast, both of these yeast species lack H3K27 methylation, indicating that this mark was acquired during metazoan evolution. Although most KMTs that methylate histones exert their effects by regulating transcription, they also play diverse roles in DNA repair, recombination and genome stability, among other functions.¹¹

There are six H3K4-specific KMTs in mammals: SETD1A, SETD1B and MLL1, MLL2, MLL3 and MLL4 (**Illustration 1B**), which are orthologous to *Drosophila* Trx and Trr and yeast Set1.¹¹ These KMTs play crucial roles in transcriptional activation via regulation of H3K4 methylation states at promoters and enhancers. While SETD1A, SETD1B, MLL1 and MLL2 control levels of H3K4me2 and H3K4me3 at the promoters of active genes,^{11,38} MLL3 and MLL4 are the major H3K4me1-specific KMTs in mammals and control H3K4me1 at enhancers.³⁹ In mESCs, MLL1 and MLL2 redundantly control H3K4me3 at bivalent gene promoters,³⁸ which encode developmentally-regulated proteins whose silencing is necessary to maintain mESC pluripotency.^{9,10} Although catalytically redundant, both SETD1A and SETD1B are essential for embryogenesis, as *Setd1a*^{-/-} embryos die at E8.5 but *Setd1b*^{-/-} embryos die at E11.5.⁴⁰ SETD1A is responsible for global H3K4me2 and H3K4me3 and some H3K4me1 in early embryos.⁴⁰ In contrast, MLL2 controls global H3K4me2 and H3K4me3 in oocytes.⁴¹ The high degree of redundancy in H3K4 KMTs in mammals as compared with fly and yeast suggests that gene duplication of these KMTs co-occurred during the expansion in tissue and organism complexity during evolution. Transcriptional

activation by H3K4 methylation is likely to be achieved by several mechanisms, which likely includes promotion of histone acetylation,⁴² inhibition of H3K9 KMT activity and countering the expansion of H3K9me3^{18,20} and recruitment of activating chromatin remodelling complexes that bind H3K4me3.⁴³

There are only two major H3K27-specific KMTs in mammals: EZH1 and EZH2

(**Illustration 1B**), which are orthologous to *Drosophila* Ez but absent in yeast. Similar to Ez, EZH1/2 proteins are responsible for global H3K27me1, H3K27me2 and H3K27me3,⁴⁴ and generally mediate transcriptional silencing in the context of PRC2. Although EZH2 is responsible for most of the H3K27me3 deposited during development, EZH1 contributes to H3K27me3 and is responsible for most of the H3K27me1.⁴⁵ A modest contribution of H3K27me1 and H3K27me2 was also shown for H3K9-specific KMTs G9a and GLP, which can interact with PRC2.⁴⁶ Notably, EZH1 and EZH2 show cell-type specificity in their incorporation into PRC2, as EZH1 is preferentially found in both dividing and non-dividing cells while EZH2 is found in dividing cells only^{24,44} and EZH1-PRC2 complexes show reduced KMT activity *in vitro* as compared with EZH2.⁴⁴ Consistent with the essential role of H3K27 methylation for developmental progression, genetic ablation of *Ezh2* is lethal by E7.5 with embryos displaying severe growth defects.⁴⁷ Early reports also showed that PRC2 and H3K27 methylation are involved in random X chromosome inactivation in embryonic cells,^{48,49} however, the precise mechanism by which H3K27me3 and PRC2 spread with *Xist* RNA across the inactive X remains unclear (reviewed by Froberg *et al.*⁵⁰). Genome-wide analyses showed that H3K27me3 targets developmentally regulated bivalent promoters in mESCs^{9,10} and EZH1/2 in the PRC2 complex is essential for maintaining developmental gene

silencing, pluripotency and subsequent lineage commitment.⁵¹ Consistent with a highly conserved functional role, PRC2 complexes also silence *Hox* gene expression in mammals.⁵¹ The molecular basis for transcriptional repression by H3K27me3 remains a matter of debate, but a long-standing model contends that PRC2-dependent repression requires subsequent recruitment of PRC1 complexes, which bind the H3K27me3 mark^{52,53} and promote histone H2AK119 monoubiquitination and chromatin compaction, and these together exclude Pol II and transcriptional activators.²⁴ However, it should be noted that PRC2-dependent transcriptional repression is not generally dependent on PRC1 and in many instances PRC1, which lacks KMT activity, still can be recruited and repress transcription independently of PRC2.⁵⁴⁻⁵⁶ PRC2 complexes also bind nascent RNAs as a surveillance mechanism to regulate its activity^{57,58} suggesting that nascent RNAs are integral in H3K27me3 deposition and PRC2 function. Notably, PRC2 recruitment and H3K27me3 can be induced ectopically by inhibition of Pol II transcription⁵⁹, indicating that transcriptional inactivity at some promoters may be sufficient to promote PRC2 binding and H3K27me3. In addition to transcription, it has been shown that PRC2/EZH2 functions in repair of DNA double strand breaks in human cell lines.⁶⁰ Together these studies illustrate the high degree of complexity associated with PRC2/EZH1/2 function in transcriptional silencing and genome regulation.

There are six H3K36-specific KMTs in mammals: SETD2, NSD1, NSD2, NSD3, ASH1L and SETMAR¹¹ (**Illustration 1B**). SETD2 is conserved from yeast to human, while NSD and ASH1L orthologues are found in *Drosophila* and *C. elegans* but not yeast.¹¹ The extensive conservation of H3K36 KMTs from yeast to human indicates a fundamental role of this modification in many different contexts and many studies indicate predominant role for

H3K36 methylation in transcription elongation.¹¹ Out of the six KMTs, SETD2 performs the majority of H3K36me3 in mammals.⁶¹ In contrast, NSD proteins catalyze the majority of H3K36me1 and H3K36me2.⁶² ASH1L also catalyzes the formation of H3K36me1 and H3K36me2⁶³ and regulates H3K36me2 in a localized fashion over the bodies of active *Hox* genes.⁶⁴ SETMAR regulates DNA repair via H3K36me2⁶⁵ while SETD2 has also been shown to promote DNA repair via H3K36me3.⁶⁶ ChIP-seq in mammalian cells reveals that H3K36me2/3 generally marks active gene bodies¹⁰ and earlier studies in yeast demonstrated that H3K36 methylation can inhibit cryptic transcription initiation via recruitment of HDAC activity to promote Pol II elongation (reviewed by Lee and Shilatifard⁶⁷). H3K36 methyltransferases have also been found to regulate alternative splicing⁶⁸ and to promote *de novo* DNA methylation in gene bodies.⁶⁹ Consistent with a fundamental role of H3K36me3 in transcriptional dynamics by Pol II in mammals, mice deficient for *Setd2* exhibit embryonic lethality by E10.5.⁷⁰ Despite their overlapping catalytic specificity, *Nsd1* null mutants show lethality by E10⁷¹ and *Nsd2* null mutants show lethality by P10, due to skeletal and respiratory system defects⁷² demonstrating unique roles of NSD family of KMTs during development *in vivo*.

There are eight H3K9-specific KMTs in mammals: SUV39H1, SUV39H2, SETDB1, SETDB2, G9a, GLP, PRDM3 and PRDM16¹¹ (**Illustration 1B**). H3K9 KMTs are generally involved in transcriptional repression and genome stability, since H3K9 di- and trimethylation is associated with transcriptional repression both at pericentric and facultative heterochromatin including repressed genes, transposable elements, ribosomal RNA genes and the inactive X chromosome in females.^{73,74} The H3K9me3 mark may promote

transcriptional silencing by a variety of mechanisms including recruitment of H3K9me3-binding proteins,³² exclusion of Pol II binding and/or deposition of active histone modifications⁷⁵ and chromatin compaction.⁷⁶ SUV39H1/2 were the first KMTs to be discovered that could methylate histones and were found to be conserved with orthologues in fission yeast, *Drosophila* and *C. elegans*.¹⁹ SUV39H1/2, which catalyzes the bulk of H3K9me3 at major satellite repeats in pericentric heterochromatin, is required for DNA methylation and maintenance of genome stability.^{77,78} Moreover, SUV39H1/2 proteins also contribute to H3K9me3 deposition at IAP ERVs and are required for silencing of a subset of LINE1 retrotransposons in mESCs.⁷⁹ In contrast, G9a/GLP perform the bulk of H3K9me2 primarily in euchromatic regions^{25,80} and can bind the H3K9me2 mark with their ankyrin repeats leading to propagation of the mark to neighbouring nucleosomes to repress genes during mESC differentiation.⁸¹ H3K9me2 deposition by G9a/GLP spreads over large, transcriptionally inactive domains⁸² and represses late-replicating genes on the nuclear periphery.⁸³ Deletion of *G9a* or *Glp* is embryonic lethal by E9.5,^{25,80} underscoring the crucial and non-redundant role of these KMTs and H3K9me2 in mammalian development. G9a/GLP are also required for efficient DNA methylation of specific LTR retrotransposons⁸⁴ and transcriptional silencing of the class III MERVL retrotransposon in mESCs.^{85,86} SETDB1 and SETDB2 perform H3K9 mono-, di- and trimethylation and whereas SETDB1 performs the bulk of H3K9me3 in euchromatin,³³ SETDB2 is reported to promote H3K9me3 at pericentric regions.⁸⁷ SETDB1 is essential for early embryogenesis, as *Setdb1*^{-/-} embryos die at around the expanded blastocyst stage (E4.5-5.5).⁸⁸ In addition, SETDB1-dependent H3K9me3 is required for transcriptional silencing of class I and II LTR retrotransposons during early embryogenesis^{89,90} and also establishes H3K9me3 on imprinted alleles.⁹¹ PRDM3 and

PRDM16 are differentiated from the other H3K9 KMTs in that they possess an N-terminal PR domain, which is related to the SET domain¹¹ and are largely cytoplasmic.⁹² PRDM3/16 perform H3K9me1 on newly synthesized histone H3 in the cytoplasm to facilitate subsequent H3K9me3 by SUV39H1/2 proteins at pericentric heterochromatin and maintenance of repression at major satellite repeats.⁹² Genome-wide analyses by ChIP-seq in specific mammalian cell types reveals marking of retrotransposons by H3K9 methylation^{10,79,93–96} and the *S. pombe* orthologue of mammalian SUV39H1/2, Clr4, is also essential for H3K9 methylation-dependent silencing of retrotransposons (reviewed by Allshire and Ekwall⁹⁷) demonstrating that this role is conserved. Thus it is clear that one of the central functions of H3K9 KMTs in diverse organisms is the repression of repetitive elements, which is the subject of this work and will be the focus of the discussion in later chapters.

There are three H4K20 KMTs in mammals: SETD8, SUV420H1 and SUV420H2¹¹ (**Illustration 1B**) and *Drosophila* and *C. elegans* have SETD8 and SUV420 orthologues. While SETD8 performs H4K20me1,⁹⁸ SUV420H1/2 perform H4K20me2 and H4K20me3.⁹⁹ As with H3K9me2/3, H4K20me2 and H4K20me3 are generally associated with transcriptional silencing of genes and transposable elements. SETD8 apparently functions both in gene repression and activation,¹¹ since H4K20me1 by SETD8 generally occurs at transcriptionally inactive promoters¹⁰⁰ as well as active gene bodies.⁹⁶ In addition, SETD8 also facilitates mitotic division, possibly due to the effect of H4K20me1 on compaction of chromatin.^{98,101} H4K20me2/3 deposited by SUV420H1/2 is tightly correlated with the presence of H3K9me3 at repetitive loci including rRNA genes, retrotransposons and pericentric heterochromatin.^{74,89,102} H4K20me2/3 deposition depends on H4K20me1

generated by SETD8 and consistent with this, genetic ablation of *Setd8* in mice leads to very early embryonic arrest at the 8-cell stage¹⁰³ underscoring the essential function of this KMT in the early embryo. In contrast, *Suv420h1* and *Suv420h2* double null mice exhibit perinatal lethality, with defects in chromosome segregation and rearrangements.⁹⁹ Notably, while H4K20me3 requires the presence of SETDB1/H3K9me3 in mESCs for its deposition at LTR retroelements, it is generally dispensable for their silencing.⁸⁹

1.5 Mutation of KMTs in human disease

As would be predicted given their essential roles in the mouse, KMTs are also crucial for human development. This is most clearly evidenced from the Mendelian overgrowth and malformation disorders arising from loss of KMT function. A common theme that has emerged from a large group of studies is that Mendelian diseases are caused by heterozygous mutations in KMTs, which results in dominant phenotypes due to haploinsufficiency, indicating that the dosage of these enzymes is critical for normal development. For example, the cognitive development and malformation disorder Kleefstra syndrome (OMIM:607001) results from heterozygous loss-of-function mutations in *EHMT1* encoding GLP.^{104,105} Sotos syndrome 1 (OMIM:117550) is a growth disorder in which patients are heterozygous for loss-of-function mutations in the *NSD1* gene,¹⁰⁶ whereas Wolf-Hirschhorn syndrome (OMIM:194190) is a contiguous gene syndrome encompassing *NSD2*.¹⁰⁷ Similarly, heterozygous loss-of-mutations in the gene encoding *MLL2* have been found in patients with Kabuki syndrome¹⁰⁸ (OMIM:147920) an autosomal dominant congenital malformation disorder with associated severe intellectual disability. Weaver syndrome (OMIM:277590), a malformation and overgrowth disorder exhibiting clinical overlap with Sotos syndrome, is

caused by heterozygous loss-of-function mutations in *EZH2*.¹⁰⁹ Autosomal dominant mental retardation 23 (MRD23, OMIM:615761) and intellectual disability were recently shown to be associated with loss-of-function heterozygous mutations in the *SETD5* gene^{110,111} encoding a putative KMT for which the substrate specificity is currently unknown. Therefore, the existence of Mendelian disorders due to germline mutations of KMT-encoding genes demonstrates that although KMTs often have broad cellular functions, the effects of their dysfunction may be tissue- or cell-type restricted and are not generally compensated for by their paralogues.

Strikingly, mutations leading to a gain of KMT function have been identified in many different forms of cancer. Numerous studies have revealed a key role for specific KMTs in the maintenance of somatic cell identity and proliferation, as somatic gain-of-function mutations in several KMTs have been associated with malignancy and metastatic development (reviewed by McGrath et al.¹¹²). For example, *MLL* genes are frequently translocated in hematological malignancies, leading to chimeric proteins that disrupt normal gene expression patterns by rewiring the transcriptome (reviewed by Ford and Dingwall¹¹³). *SETDB1* is recurrently amplified and acts as an oncogenic factor in several cancers including melanoma,³⁵ lung¹¹⁴ and prostate carcinoma.¹¹⁵ Similarly, heterozygous gain-of-function somatic mutations of *NSD2* are frequently found in pediatric acute lymphoblastic leukemia.¹¹⁶ Moreover, *NSD1*, *NSD2* and *NSD3* are frequently overexpressed or amplified in malignancy in multiple tissues (reviewed by Vougiouklakis et al.¹¹⁷). Heterozygous gain-of-function mutation of *EZH2* at Y641 is a frequent feature of diffuse large B-cell and follicular lymphomas.¹¹⁸ The Y641F or Y641N mutations were shown to enhance the ability of *EZH2*

to catalyze H3K27me2 to me3 but abrogate the activity of the enzyme towards unmodified H3, leading to a shift towards increased H3K27me3 *in vivo*.^{118,119} It should be noted, however, that the propensity towards increased KMT expression and/or activity in malignancy is a unique feature of the cell type. This is clearly demonstrated by the recent finding that a H3K27M mutation occurring in pediatric glioblastoma leads to reduced PRC2 binding and globally lower H3K27me3 levels.¹²⁰

The general effect of overexpression of KMTs with repressive activities is increased binding to chromatin and higher levels of their cognate methylation marks, leading to ectopic transcriptional repression of tumour suppressor genes and promotion of malignant tumour growth.^{35,121} In addition to methylation of histone residues, overexpression of KMTs also exerts oncogenic effects via their ability to directly target tumour suppressor proteins for methylation, such as p53 and NFκB, which may alter their functions.^{70,122} The growing number of studies linking somatic gain-of-function mutations in KMT genes with specific malignancies has led to the development of novel therapeutic strategies based on KMT inhibition. For instance, recent evidence suggests that small-molecule inhibition of EZH2 in cancers that show a gain-of-function is a promising therapeutic strategy.^{121,123} As inhibitors for other KMTs with oncogenic effects are characterized, small-molecule KMT inhibition may become a viable therapeutic option for patients suffering from other malignancies.¹¹²

1.6 Regulation of KMTs by long non-coding RNAs

An emerging area of regulatory control over KMTs is the role for RNA molecules, such as long non-coding RNAs (lncRNA) in mediating KMT activity and recruitment (reviewed by Wang and Chang¹²⁴). LncRNAs are defined as RNA species >200 bases in length that are transcribed by RNA Pol II from genes that lack protein-coding potential¹²⁴ and may serve to recruit KMTs by acting as molecular scaffolds that stabilize their binding on chromatin.¹²⁵ A classic example is the ~17 kb *Xist* transcript, which guides PRC2 and H3K27me3 and other chromatin modifiers to establish silencing in *cis* on the inactive X chromosome^{126,127} but there is a growing list of other shorter lncRNAs with specific functions in mammals (reviewed by Batista and Chang¹²⁸). For example, *HOTAIR* is expressed from the *HOXC* gene cluster that acts to recruit PRC2 and H3K27me3 to silence genes in *trans* in the *HOXD* cluster.¹²⁹ In contrast, the *HOTTIP* lncRNA is transcribed from the *HOXA* cluster, binds to WDR5 in the MLL-containing COMPASS-like complex and governs its recruitment to the other genes in *cis* in the *HOXA* cluster to activate transcription via H3K4me3.¹³⁰ G9a is recruited by the lncRNA *Air* to repress imprinted genes *Slc22a2*, *Slc22a3* and *Igf2r* in *cis* in the placenta.¹³¹ Similarly, G9a and PRC2 are also recruited by the lncRNA *Kcnq1ot1* in the placenta to repress nearby imprinted genes in *cis* via H3K9 and H3K27 methylation.¹³² More recently, the roles of several lncRNAs, including *Hotair* have been characterized by genetic knockout which has shown their important roles in development including defects in H3K27 methylation in the case of *Hotair* null mice.^{133,134} *HOTAIR* is also overexpressed in a variety of malignancies, including breast carcinoma where it promotes metastasis in a PRC2/H3K27me3-dependent manner.¹³⁵ These studies highlight the important role of lncRNAs in regulating KMT recruitment and function in mammals.

1.7 The heterogeneous nuclear ribonucleoprotein (hnRNP) family of RNA-binding proteins

The hnRNP group of proteins was first characterized as a large multi-subunit assembly associated with pre-mRNA in the nucleus of HeLa cells.¹³⁶ There were ~20 proteins in the complex, which were termed hnRNP A/B through hnRNP U, using a different letter for each protein on the basis of demonstrating their distinct biochemical properties, including molecular weights and isoelectric points in 2D PAGE.¹³⁶ HnRNPs are also identified as core subunits of the ~40S spliceosome assembly.¹³⁷ The hnRNPs are not related as paralogues, however, they generally function together and possess different RNA binding domains, including the RNA-recognition motif (RRM) and K-homology (KH) domains.¹³⁸ HnRNP K, which is the focus of my thesis work, is one of the hnRNPs found in the spliceosome and also plays roles in transcription, splicing and translation in diverse contexts (reviewed by Han et al. and Bomsztyk et al.^{138,139}). Recent evidence also points to a role for hnRNP K in lncRNA-mediated gene regulation.^{126,140,141}

1.8 SUMOylation and chromatin regulation

SUMOylation is a post-translational modification in which SUMO protein paralogues, SUMO1, SUMO2/3 (which are ~97% identical) can be covalently added to lysine residues on target proteins via the action of conserved SUMO E1 activating and E2 conjugating enzymes, often with the aid of specific SUMO E3 ligases (reviewed by Cubenas-Potts and Matunis¹⁴²). SUMOylation is reversed by the activity of SUMO isopeptidases known as SENPs, which exhibit specific proteolytic activity for the SUMO-lysine isopeptide bond (reviewed by Drag and Salvesen¹⁴³). Unlike ubiquitination, which tags proteins for degradation by the proteasome, SUMOylation is a highly versatile regulatory mechanism

that can control protein-protein interactions, cellular localization, enzymatic activity and protein stability.¹⁴² A growing body of evidence indicates that SUMOylation of chromatin-bound proteins plays an important role in transcription and may impact KMT-dependent regulation of heterochromatin structure. For example, the E2 conjugating enzyme Ubc9 binds chromatin in *S. pombe* and promotes SUMOylation of the Suv39 orthologue Clr4 to promote its activity and maintain gene silencing at centromeres and mating loci.¹⁴⁴ Similarly, SUMO2/3 is enriched in heterochromatin and MBD1 recruits the SETDB1/MCAF1 complex to heterochromatin in a manner dependent on its SUMOylation in HeLa cells.¹⁴⁵ Mutation of *Ubc9* is lethal shortly after E3.5 and mutants show gross defects in chromosome structure and nuclear organization¹⁴⁶ consistent with a critical role for SUMOylation in growth and proliferation by maintaining genome stability. Together these studies point to an important role for SUMOylation in regulating chromatin structure and KMT function in diverse contexts in yeast and mammals.

1.9 Transcriptional silencing of retrotransposons as a model system for studying H3K9-specific KMT function

The diverse roles played by KMTs in development and their dysfunction in developmental syndromes and cancer as discussed above are highly variable, context-dependent and complex necessitating the use of model systems in which to focus on specific aspects of their functions and elucidate novel mechanisms. As mentioned above, one of the central functions of H3K9-specific KMTs is the transcriptional silencing of transposable elements, which is conserved from yeast to human. Recent advances in the mechanisms governing transcriptional repression of retrotransposons in the preimplantation embryo, pluripotent stem cells and adult somatic cells have laid the foundation for more in-depth mechanistic

investigations into the molecular basis of targeting of H3K9 KMTs to these parasitic genetic elements (reviewed by Leung and Lorincz¹⁴⁷), the focus of my thesis work.

1.10 Phylogeny and structure of retrotransposons in the mammalian genome

Approximately 40-45% of mammalian genomes are comprised of repetitive elements, the majority of which (~30%) are retrotransposons (reviewed by Cordaux and Baxter¹⁴⁸

Illustration 2A). These parasitic elements replicate themselves via a ‘copy-and-paste’ mechanism, which involves transcription and subsequent reverse transcription of the RNA intermediate, followed by integration of the resultant cDNA into the genome. Mammalian retrotransposons are further divided into those that are flanked by LTRs, also called endogenous retroviruses (ERVs), or those lacking LTRs (**Illustration 2B**). LTR retrotransposons are thought to have colonized the genome as a consequence of retroviral infections of the host germline and are subsequently transmitted in a Mendelian fashion, having lost the ability to exit the cell in the manner of a typical envelope-encoding replication-competent retrovirus.¹⁴⁹ Non-LTR retrotransposons are the most numerous and active, comprising ~25-27% of the mouse and human genomes¹⁴⁸ (reviewed by Stocking and Kozak¹⁵⁰). This family consists of the autonomous LINE1 (L1) elements, which are ~6 kb in length and present in ~950,000 copies in the human genome of which ~100 are predicted to remain retrotranspositionally competent (reviewed by Friedli and Trono¹⁵¹). L1s harbour two ORFs for their replication and retrotransposition and are flanked by UTRs, of which the 5’ UTR harbours a Pol II promoter (**Illustration 2B**). Related to these are the non-autonomous SINE elements, the majority of which are the ~300 bp long *Alu* elements in primates, which essentially lack ORFs and therefore can only retrotranspose in *trans* using gene products from L1 elements.¹⁴⁸ *Alu* elements are flanked by left and right monomers (**Illustration 2B**),

of which the left or 5' monomer possesses two canonical Pol III promoters in tandem and are present in ~1,800,000 copies in the human genome of which ~1000 are active.¹⁵¹ The remaining non-LTR retrotransposons with active members in humans are SVA elements, which are ~2 kb long and present in ~5500 copies, of which ~50 are active.¹⁵¹

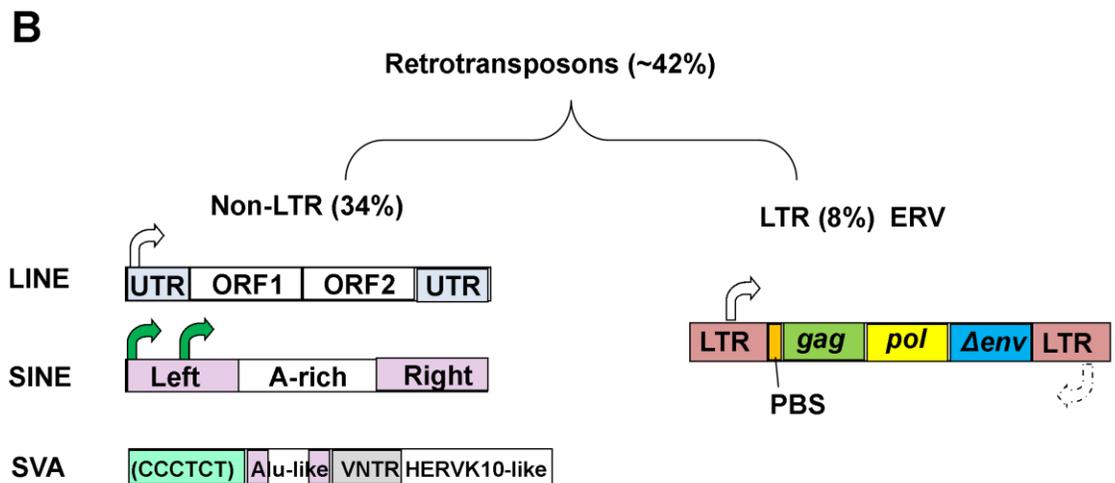
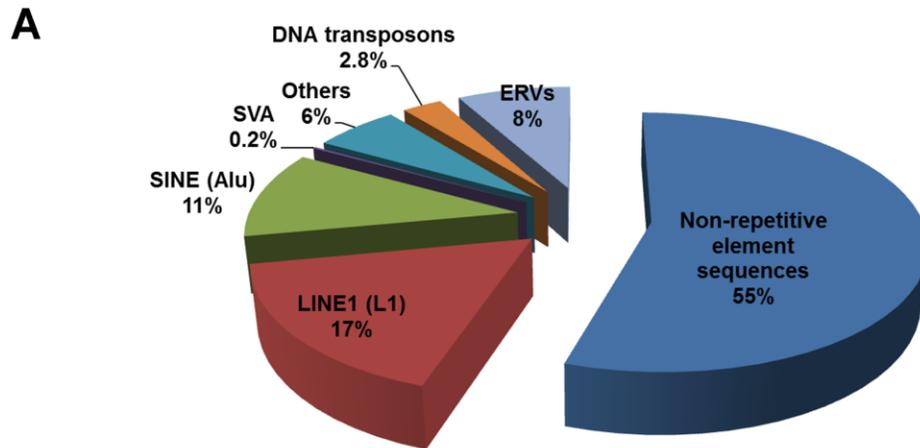


Illustration 2. Repetitive elements in the human genome and structure of retrotransposons. (A) Breakdown of the types of repetitive elements in the human genome based on Cordaux and Baxter.¹⁴⁸ “Others” indicate inactive LINE2 and MIR sequences. (B) Structure of major classes of retrotransposons in the human genome. White arrows indicate Pol II promoters and TSSs. Dashed arrow indicates the presence of a Pol II promoter and TSS in the 3’LTR of some ERVs, which can support antisense transcription. Green arrows indicate Pol III promoter and TSS. The Δenv represents the observation that the vast majority of ERV *env* genes in class I and II ERVs are mutated and thus non-functional. Major sequence units of each element are shown and element size is not to scale.

These elements harbour sequences derived from SINE, VNTR and Alu elements, hence their name, but lack an obvious Pol II promoter region and ORFs involved in retrotransposition and thus are also non-autonomous (**Illustration 2B**).

Approximately 8-10% of the mouse and human genomes are composed of ERVs, which form a superfamily with three distinct classes based on the homology of their reverse transcriptase (*pol*) genes.^{148,150} Their nomenclature in humans are based on their primer-binding site (PBS) tRNA (reviewed by Katoh *et al.*¹⁵²) and they can be classified into 31 different subfamilies representing 50 independent waves of retroviral integration into the ancestral human lineage.¹⁵¹ ERVs from all three classes are generally ~7-11 kb in length and typically possess Pol II promoters in their 5’ and 3’ LTRs. They also harbour a tRNA primer-binding site (PBS) immediately downstream of the 5’ LTR, which is used to prime reverse transcription, along with two other ORFs in addition to their *pol* gene: *gag* encoding a group-specific antigen polyprotein and *env* encoding an envelope protein, although most ERVs actually lack this gene (**Illustration 2B**, reviewed by Jern and Coffin¹⁵³). While the vast majority of ERVs in both mouse and human have accumulated inactivating mutations, some subfamilies

have been shown to be retrotranspositionally competent including the HERVK (HML-2) family in humans¹⁵⁴ and the IAP and MusD ERVs in mice.¹⁵¹

1.11 Evolutionary impacts of retrotransposons on the genome

The differences in retrotranspositional activity of ERVs and non-LTR retroelements in mammals demonstrates evidence for lineage-specific evolutionary forces shaping their activity. For example, in the mouse many class I, II and III ERVs including MLV, IAP and MERVL elements remain intact and transcriptionally active and *de novo* retrotransposition, although rare, has been estimated to account for up to ~10% of spontaneous mutations in the C57BL/6 strain (reviewed by Maksakova *et al.*¹⁵⁵). In contrast, virtually all human ERVs (HERVs) are thought to have acquired mutations and there are no reports of *de novo* HERV retrotransposition.¹⁴⁸ While ~99.9% of L1 elements are also inactive due to 5'UTR deletions and point mutations (reviewed by Beck *et al.*¹⁵⁶), a small subset have continued to retrotranspose throughout primate evolution and remain highly active in humans.¹⁴⁸ Although retrotransposition is generally deleterious to the host genome due to its mutagenic potential, retrotransposons that have escaped negative selection have played important roles in host fitness by increasing genetic diversity,^{148,150} providing the building blocks for gene regulatory networks^{157,158} (reviewed by Cowley and Oakey¹⁵⁹) and tissue-specific promoters,¹⁶⁰ serving as essential promoters or exons,¹⁶¹ seeding the generation of new gene families,^{162,163} and stimulating the evolution of transcriptional repressive mechanisms to counteract the invasion of exogenous retroviruses.^{153,164,165}

1.12 Mutagenic mechanisms of retrotransposons and their roles in human disease

A variety of mutagenic mechanisms of retrotransposons have been observed in mammalian genomes affecting gene expression and genome stability and causing Mendelian disease.^{153,155,159} Insertional mutagenesis occurs as a consequence of retrotransposition into a gene, disrupting the ORF or regulatory sequences and was first shown to occur in humans by an L1 insertion into the gene encoding Factor XIII, resulting in Hemophilia A¹⁶⁶ (OMIM:306700). *De novo* retrotransposition of L1s are an ongoing source of mutagenic activity in humans and has been shown to cause Mendelian disorders in over 60 different cases.¹⁴⁸ *De novo* germline retrotransposition or the presence of retrotransposon-derived sequences such as LTRs can also disrupt gene regulation by introducing novel regulatory sequences such as promoters/enhancers, splice acceptors/donors and polyadenylation signals.^{152,153} For example, the proto-oncogene *CSF1R* gene is aberrantly activated by an LTR from the human ERV THE1B, promoting Hodgkins lymphoma.¹⁶⁷ L1 retrotransposition has also been shown to occur in colon epithelial cells, leading to a familial adenomatous polyposis (OMIM:175100) as a consequence of insertional mutagenesis into the *APC* gene.¹⁶⁸ In addition, retroelements can also induce ectopic epigenetic modifications, such as DNA methylation to silence the expression of neighbouring genes, first demonstrated in mammals in the report of the *Agouti viable yellow* mouse¹⁶⁹ although thus far, this has not be definitively shown as an etiological factor in any Mendelian disease in humans.

In addition to mutagenic effects on single genes, retrotransposons also exert mutagenic effects on a larger structural scale.^{148,156} Retrotransposition in itself leads to double-stranded breaks in the genome, increasing the propensity for sequences to be deleted during DNA

repair and in turn decreasing genome stability. In addition, ERV-derived sequences such as LTRs can mediate recombination events, leading to structural rearrangements, deletions and duplications.^{148,153,156} Retrotransposons can be induced in particular cancers, however, the exact contribution or influence of retrotransposon expression on oncogenesis remains to be determined (reviewed by Kassiotis¹⁷⁰).

1.13 DNA methylation and retrotransposon silencing

The most well-studied transcriptional silencing mechanism for retrotransposons is DNA methylation, which is required for silencing of these parasitic elements in somatic cells and late in male germline development.^{171,172} DNA methylation in mammals is catalyzed by the three DNMT enzymes and occurs at the fifth position on the cytosine base (5mC) in the context of CpG dinucleotides (**Illustration 3A**, reviewed by Schübeler¹⁷³). High-resolution DNA methylation profiling studies have consistently reported that retrotransposons including LINE1 elements and ERVs are marked by high levels of DNA methylation in differentiated somatic cells in mouse and human¹⁷³ (reviewed by Messerschmidt *et al.*¹⁷⁴). DNA methylation in the promoter region or CpG island of genes is thought to mediate transcriptional silencing via direct exclusion of specific TFs and Pol II and recruitment of methyl-CpG-binding domain proteins.¹⁷³

It was previously believed that DNA methylation was stable and generally irreversible with the exception of replication-coupled dilution in the absence of its maintenance. However, over the past six years a large body of work (reviewed by Kohli and Zhang¹⁷⁵) has revealed that 5mC is a substrate for the TET family of Fe²⁺-dependent dioxygenases including TET1, TET2 and TET3¹⁷⁶ that catalyze its iterative oxidation into 5hmC, 5fC and 5caC,¹⁷⁷ which

can be removed by the DNA glycosylase TDG and subsequently replaced with C by the base-excision repair machinery^{178,179} (**Illustration 3A**). Moreover, 5hmC and other oxidized cytosine derivatives are not recognized by the DNMT1 maintenance methylation machinery, leading to passive loss of DNA methylation after mitotic division.¹⁷³ Therefore, DNA methylation can be actively and passively removed from mammalian DNA as a consequence of TET activity, which has important implications for understanding the regulatory mechanisms that maintain 5mC at retrotransposons. In addition to active TET-mediated demethylation, passive demethylation due to the decreased expression of DNMTs also contributes to 5mC reprogramming in the early embryo and germline.¹⁸⁰⁻¹⁸²

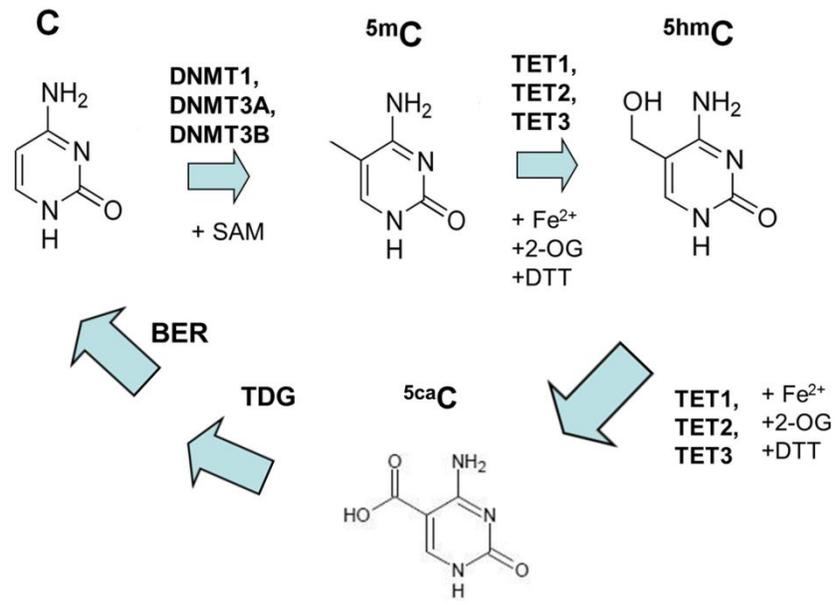
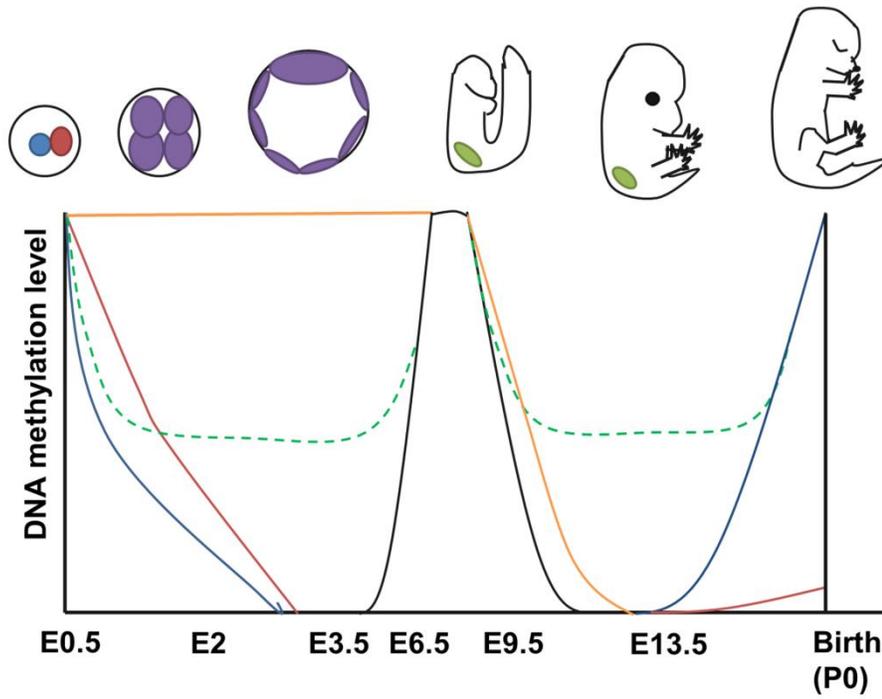
A**B**

Illustration 3. DNA methylation, its oxidation and reprogramming during

embryogenesis in the mouse. (A) Enzymatic pathway for production and oxidation of DNA methylation (5mC) and removal from DNA. 5mC catalyzed from C by DNMTs with the co-substrate SAM, which is oxidized to 5hmC, 5fC (not shown) and 5caC by the TET family of Fe²⁺-dependent dioxygenases, which also require α -ketoglutarate and DTT as co-factors. 5caC can be recognized and its glycosidic bond hydrolyzed by TDG, leaving an abasic site that is repaired by the base excision repair (BER) pathway resulting in C. (B) Reprogramming of DNA methylation during embryogenesis. Mouse embryonic development is depicted above the plot with approximate stages (embryonic days) shown below. Paternal/male PGC DNA is shown in blue while maternal/female PGC DNA is shown in red. Solid black line represents the genome as a whole (maternal and paternal) and includes most CpG islands, introns, exons and intergenic regions. The orange line represents imprinted or differentially methylated regions, whose methylation is only reset in PGCs during their development and these regions are among the last to be demethylated. Green dashed line indicates repetitive elements including retrotransposons, which are generally hypomethylated during both waves of reprogramming. The green circles in the embryo represent the PGCs. After sexual differentiation at E13.5, male prospermatogonia DNA is fully re-methylated by birth while female immature oocyte DNA is only re-methylated after the oocytes fully mature at puberty.

Although DNA methylation patterns are stably inherited in mitotic somatic cells as a consequence of the maintenance methyltransferase DNMT1 and its associated binding partners, 5mC must be reprogrammed initially for the embryo to achieve a totipotent state and subsequently to program lineage commitment decisions and germline development.¹⁷⁴

To this end, DNA methylation is removed from the genome and re-established *de novo* during two windows of embryogenesis (**Illustration 3B**). The first wave of DNA methylation reprogramming occurs immediately following fertilization of the oocyte, where the maternal and paternal genomes undergo programmed loss of DNA methylation during the first cleavage divisions of the zygote. Immunostaining and fine-scale mapping using bisulfite sequencing analyses together indicate that DNA methylation is removed from the maternal and paternal genomes via a combination of passive and active TET3-mediated demethylation mechanisms^{183–186} leading to low levels of 5mC and high 5hmC in the paternal genome as

compared with the maternal genome by the 4-8 cell stages¹⁸⁴ (**Illustration 3B**). By the early blastocyst stage at E3.5, the vast majority of 5mC is lost throughout the genome with the exception of imprinted loci and repetitive elements including retrotransposons, the latter of which are nevertheless hypomethylated relative to differentiated adult somatic cells¹⁸⁷⁻¹⁹⁰ (**Illustration 3B**). Interestingly, during the early erasure of DNA methylation in the paternal genome, the maternal genome is protected from active demethylation activity by PGC7/Stella,¹⁹¹ which binds the H3K9me2 mark on maternally derived chromatin.¹⁹² Upon implantation of the blastocyst at ~E5.5-6.5, DNA methylation is restored *de novo* to facilitate differentiation during gastrulation and germline specification. The second major wave of demethylation occurs in the primordial germ cells (PGCs), which are set aside at E6.5-7.5 and subsequently begin proliferating and migrating to the genital ridge (**Illustration 3B**).¹⁷⁴ PGCs undergo passive demethylation facilitated by 5hmC formation during their proliferation and migration which is essentially complete by E12.5-13.5.^{182,193-197} As a result of this global demethylation process, virtually all 5mC is removed including at promoters, CpG islands, exons, introns, intergenic regions and imprinted regions to allow for the adoption of parental-specific DNA methylation on imprinted alleles. Retrotransposons are among the very few sequences found to be at least partially protected from demethylation in PGCs.⁹⁵ DNA methylation patterns are re-established *de novo* after E13.5, coincident with sexual differentiation. The maternal and paternal imprints are reset in male prospermatogonia prenatally, while immature oocytes in the female remain hypomethylated until after birth, and are *de novo* methylated in the growing oocyte (**Illustration 3B**). While retroelements are generally protected from total DNA demethylation both in the early embryo and in PGCs, their hypomethylated state in conjunction with the expression of TET proteins during these

stages presents a unique window in which DNA methylation turnover is more dynamic, necessitating the requirement for additional epigenetic mechanisms to ensure their repression.¹⁴⁷ A growing body of evidence points to a critical role for H3K9 methylation as the predominant silencing mechanism for retroelements both in the blastocyst and E13.5 PGCs where DNA methylation is erased genome-wide and retrotransposons become hypomethylated.^{89,90,95} Together, these findings have demonstrated specific windows in development that may serve as models for investigating the role of H3K9-specific KMTs in transcriptional silencing of retrotransposons.

1.14 Transcriptional silencing of retrotransposons by H3K9-specific KMTs

Early studies suggesting DNA methylation-independent transcriptional silencing of retrotransposons and related sequences in the embryo used retroviral vectors based on the naturally occurring type C ecotropic retrovirus MLV and investigated production of intracisternal A-type particles derived from IAP ERVs.¹⁹⁸⁻²⁰⁰ Initially, it was observed that DNA methylation accumulated on integrated proviruses and the incubation with the DNA methylation inhibitor 5azaC could induce expression of both newly integrated proviruses and IAP ERVs in MEFs lines.^{198,199} In contrast with these observations, it was subsequently shown that a distinct mechanism operated in pluripotent mouse embryonal carcinoma cells (mECCs), which suppressed transcription from integrated proviral DNA and was unaffected by 5azaC.²⁰⁰ Intriguingly, when mECCs were induced to differentiate they then reverted to a DNA methylation-dependent proviral silencing pathway.²⁰⁰ Notably, this mechanism was generally independent of retroviral integration site and critically relied upon specific

sequences in the PBS and 5'UTR, which when mutated relieved proviral silencing.^{201,202} The PBS was particularly important, since a single A to G point mutation in the PBS^{Pro} of the vector, termed B2, relieved silencing.²⁰¹ Moreover, specifically engineering the MLV-based retroviral vectors by replacing the proline tRNA complementary PBS (PBS^{Pro}) sequence with PBS^{Gln} resulted in modest proviral expression in both mECCs and mESCs²⁰³ indicating that sequences within the provirus directed transcriptional silencing. Later experiments in which *Dnmt3a*^{-/-}; *Dnmt3b*^{-/-} mESCs were infected with MLV-based retroviral vectors confirmed the earlier work and showed that proviral silencing is established by day 3-4 and is unaffected by the absence of *de novo* methylation.²⁰⁴ The main conclusions of this early body of work on retroviral silencing was later extended to ERVs in a definitive study by Hutnick *et al.*²⁰⁵ that showed that undifferentiated mESCs lacking *Dnmt1* and thus globally depleted of DNA methylation nevertheless maintained silencing of IAP ERVs, but upon their differentiation IAP was dramatically induced. Moreover, this DNA methylation-independent silencing pathway in mESCs was shown to be unaffected by deletion of *Dicer1*, demonstrating that it was not dependent on the production of siRNAs.²⁰⁵

The first component of this apparently pluripotent stem cell-specific retroviral silencing pathway to be unequivocally identified was the transcriptional co-repressor KAP1 (also called TRIM28).^{206,207} Previous work had demonstrated that KAP1 is a co-repressor for KRAB-ZFPs, which represent the single largest family of transcriptional repressors in the mouse and human genomes (reviewed by Lupo *et al.*²⁰⁸). KRAB-ZFPs were identified on the basis of the N-terminal KRAB domain,²⁰⁹ which was found to be a strong transcriptional repression module and later shown to exert its effects on transcription through direct

interactions with the RBCC domain of KAP1.²¹⁰⁻²¹³ In HeLa cells, KAP1 was found to repress transcription from KRAB domain-bound promoter constructs via recruitment of different repressive chromatin modifying activities including chromatin remodelling by CHD3, histone deacetylation by HDAC1 and HDAC2²¹⁴ and H3K9 methylation by SETDB1.²¹⁵ In addition, KAP1 was also found to directly interact with HP1 proteins^{216,217} and direct their recruitment to chromatin to enforce a heterochromatic environment and gene silencing.²¹⁸ Despite that the involvement of these additional repressive factors in proviral silencing had not been established, mutation of the HP1 binding site on KAP1 perturbed the silencing of newly integrated MLV-based retroviral vectors in mECCs pointing to the involvement of HP1 proteins.²¹⁹ Subsequently, the finding that the KRAB-ZFP ZFP809 could directly bind the PBS^{Pro} sequence and target KAP1-dependent silencing to MLV-based retroviral vectors²²⁰ was the seminal observation that led to the general model positing that KRAB-ZFPs evolved to bind diverse retrovirus-derived sequences and could recruit KAP1 and repressive chromatin modifications to enforce their silencing (reviewed by Robbez-Masson and Rowe²²¹). However, it remained unclear whether this KRAB-ZFP/KAP1 silencing pathway actually acted on retrotransposons and whether it was indeed independent of DNA methylation.

The last pieces of this puzzle came together with two studies that reported the requirement for KAP1 and SETDB1 in transcriptional silencing of ERVs and MLV-based retroviral vectors specifically in mESCs and preimplantation embryos.^{89,90} ERVs including IAP, MusD and MLV were found to be highly de-repressed upon deletion of *Kap1*, concomitant with loss of H3K9me3 and importantly, KAP1-dependent silencing was shown to be directed by

sequences within the IAP element.⁹⁰ SETDB1 was shown to be the crucial H3K9 KMT for class I and II ERV silencing in mESCs, since deletion of *G9a*, *Glp*, *Suv39h1* and *Suv39h2* had no effect on silencing of these ERVs and SETDB1 catalytic activity recruited by KAP1 was essential for maintenance of silent proviral chromatin.⁸⁹ Furthermore, total loss of DNA methylation in undifferentiated mESCs due to deletion of all three catalytic DNMTs (*Dnmt* TKO) had only a modest effect on IAP silencing and did not de-repress the newly integrated MLV-based vector MSCV-PBS^{Gln} or affect its H3K9me3 levels.⁸⁹ These data were consistent with the results of Rowe *et al.* who reported that KAP1 deletion was synergistic with 5azaC treatment of mESCs.⁹⁰ More recently it has been shown that the SETDB1/KAP1 complex regulates *de novo* DNA methylation of proviral sequences²²² and that SETDB1 regulates DNA methylation turnover at ERV LTRs in mESCs.²²³ SETDB1 is also required for ERV silencing in E13.5 PGCs and directs DNA methylation at H3K9me3-enriched ERV sequences,⁹⁵ supporting the notion that the SETDB1/KAP1 silencing machinery can direct DNA methylation to maintain retroelement silencing. Although it was thought that the SETDB1/KAP1 ERV silencing pathway would only operate in cell types derived from embryonic stages where DNA methylation is being reprogrammed consistent with the earlier work described^{89,147} several studies have now shown that SETDB1/KAP1 are also required for ERV silencing in more differentiated cell types including neural progenitors^{224,225} and B lymphocytes²²⁶ suggesting a higher degree of cell type specificity than previously assumed. Moreover, recent studies have also pointed to the conservation of the KRAB-ZFP/KAP1/SETDB1 pathway in silencing an evolutionarily distinct subset of ERV, L1 and SVA elements in human embryonic stem cells.^{165,227,228} In summary, this comprehensive body of work has demonstrated KRAB-ZFP/KAP1/SETDB1 complexes to be the essential

epigenetic silencing machinery that acts independently of the presence of DNA methylation to silence newly integrated retroviral vectors and ERVs in specific cell types.

In addition to the role of SETDB1 in retrotransposon silencing, other H3K9-specific KMTs have also been shown to promote retroelement silencing in mESCs. For example, while G9a/GLP is not necessary for maintaining ERV silencing, its catalytic activity is required for *de novo* DNA methylation and establishing a silent state at newly integrated MSCV PBS^{Gln} proviruses.²²⁹ In addition, class III MERVL elements, which are not targeted by SETDB1 or H3K9me3 are enriched in H3K9me2 and are de-repressed in *G9a*^{-/-} and *Glp*^{-/-} mESCs.^{85,86} Furthermore, G9a is also necessary and sufficient to silence L1 retrotransposons in spermatogonia.²³⁰ Similarly, SUV39H1/2 are not necessary for ERV silencing yet they contribute to H3K9me3 at class I and II ERVs and are required for maintaining this mark and silencing a distinct subset of L1 elements in mESCs but not more differentiated somatic cell types.⁷⁹

1.15 Thesis Objectives

An emerging theme from prior work is that H3K9-specific KMTs are critical for DNA methylation-independent retrotransposon and retroviral vector silencing in specific cell types, including pluripotent stem cells in mouse and human. These findings pave the way for detailed mechanistic analyses of these silencing pathways with the aim of elucidating novel insights of relevance to Mendelian diseases and cancers linked with mutations in these KMTs. Many questions still remain regarding the mechanisms of retrotransposon silencing by H3K9-specific KMTs in these contexts. The primary objective of this work was to

identify novel factors that interact with the SETDB1/KAP1 and G9a/GLP complexes and characterize their role in ERV silencing pathways, using mESCs as a model system.

In Chapter 3 of this thesis, I isolated SETDB1-associated proteins in mESCs with the intention of finding novel co-repressors. I identified and validated a novel SETDB1 interactor, the RNA binding protein hnRNP K, which was previously shown to play roles in transcriptional regulation in various contexts. Biochemical analysis demonstrated that hnRNP K interacts with SETDB1 indirectly via its direct binding to KAP1 and thus is associated with the SETDB1/KAP1 complex. Genetic analysis showed a critical function for hnRNP K in maintenance of ERV and retroviral vector silencing, via its impact on H3K9me3 deposition and SETDB1 recruitment. HnRNP K was also found to repress a cohort of germline genes directly targeted by SETDB1 and H3K9me3, indicating that its role in SETDB1-mediated transcriptional repression extends to genes. I further showed that SUMO conjugation, which is necessary for SETDB1 to bind to KAP1 *in vitro*, is also important for proviral silencing. Importantly, the depletion of hnRNP K generally phenocopied the depletion of SUMO conjugating enzyme Ubc9 with respect to proviral silencing pointing to a novel role for hnRNP K in promoting SUMOylation of chromatin-associated proteins to ensure efficient SETDB1 recruitment at proviral chromatin. Additional analysis of the role of hnRNP K in protein SUMOylation suggested that although hnRNP K deficiency leads to loss of SUMOylation detected on proviral chromatin by ChIP, it does not generally regulate KAP1 SUMOylation *in vitro* and thus might require additional factors or otherwise control SUMOylation of an unidentified protein(s) or histones themselves at proviral chromatin.

In Chapter 4 of this thesis, I performed a detailed mechanistic investigation of the roles of hnRNP K, G9a/GLP and H3K9me2 in silencing class III MERVL elements. By analyzing H3K9me2 ChIP-seq datasets generated in the Lorincz lab and investigating the silencing of a newly integrated MERVL LTR construct, I found that H3K9me2 is not recruited to MERVL elements in an autonomous fashion as with H3K9me3 at class I and II ERVs, but instead marks MERVLs as a general consequence of their genomic location in large transcriptionally inert domains. Depletion of hnRNP K led to reduced H3K9me2 both at a global level and as detected by ChIP at MERVL elements and other loci, implicating hnRNP K in G9a/GLP function. Moreover, hnRNP K physically interacted with the G9a/GLP complex in a manner dependent on intact RNAs, demonstrating that it forms a novel complex with these KMTs. Interestingly, hnRNP K and G9a/GLP mutually supported the chromatin binding of each other in that loss of either factor leads to reduced chromatin enrichment of the other. In an effort to identify RNAs associated with hnRNP K/G9a/GLP complexes, I verified interactions between hnRNP K and GLP with MERVL RNA in mESCs and showed that hnRNP K can bind to MERVL transcripts in a manner dependent on one of its RNA binding domains. A point mutation which abolishes binding of this domain to DNA or RNA leads to upregulation of MERVL elements in wt but not *G9a*^{-/-} cells, implicating hnRNP K binding to nucleic acids in G9a/GLP-dependent MERVL repression.

Taken together, this work provides novel insights into the mechanisms of retrotransposon silencing by H3K9-specific KMTs in mammals. While class I and II ERV silencing by

SETDB1/KAP1 and H3K9me3 is coordinated in an autonomous, hierarchal fashion and promoted by hnRNP K, class III MERVL elements are likely repressed by combination of H3K9me2, hnRNP K, G9/GLP and the absence of transcriptional activators. Since heterozygous loss-of-function mutations in *HNRNPK* have been recently found to cause a new Mendelian disorder²³¹⁻²³³ and *HNRNPK* is overexpressed and exerts oncogenic activity in several malignancies,²³⁴⁻²³⁶ my work linking this factor with H3K9-specific KMTs during development has important implications for understanding the epigenetic pathways affected in these diseases.

2. Materials and methods

2.1 Cell lines and cell culture

The following previously characterized mESC lines were used in this study: TT2 wt and 33#6 *Setdb1*^{lox/-} (expressing the Cre-ER transgene, hereafter referred to as *Setdb1* CKO) infected with a silent MSCV PBS^{Gln} vector,⁸⁹ *Setdb1* CKO expressing 3XFLAG-*Setdb1*,⁸⁹ HA36 IAP-GFP,²³⁷ J1 wt and *Dnmt1*^{-/-}; *Dnmt3a*^{-/-}; *Dnmt3b*^{-/-} (*Dnmt* TKO),²³⁸ *G9a*^{-/-} clone 22-10 and *Glp*^{-/-} clone cd12^{25,80} and A2lox.²³⁹ MERVL LTR-*Gfp-T2A-Puro* reporter (*2C::Gfp*) mESC line was generated as previously described for the *tdTomato* reporter⁸⁵ and provided by Dr. Todd Macfarlan. The pcDNA3-Hygro vector was modified by exchange of the CMV promoter for the MERVL clone #9 5'LTR-PBS-*gag* region (737 bp) and a *Gfp-T2A-Puro* construct was cloned downstream of the LTR. The MERVL LTR reporter construct was transfected into the A2lox line using lipofectamine 2000 and stable clones were isolated upon selection with hygromycin. The *2C::Gfp* cell line was not viable grown in feeder-free conditions in serum media (data not shown) and therefore was cultured in a modified form of the previously described 2i media with LIF²⁴⁰: 3 μ M MEK1/2 inhibitor PD03259010 (Stemgent), 1 μ M GSK3 β inhibitor CHIR99021 (Stemgent), 0.05% BSA, 48.5% neurobasal media (Gibco), ~10-50 ng/ml recombinant LIF (purified in-house from BL21 *E.coli* expressing a thrombin-cleavable GST-LIF fusion protein and assayed on TT2 mESCs for supporting self-renewal), 48.5% Dulbecco's modified Eagle's Medium (DMEM)/F12 (Gibco), 1% B-27 supplement (Gibco), 0.5% N-2 supplement (Gibco) and 100U/ml penicillin-streptomycin (HyClone). All other mESC lines were cultured in standard feeder-free conditions on 0.2% type A gelatinized tissue culture plates in complete mESC media: DMEM high glucose, 15% fetal bovine serum, 20 mM HEPES, 1 mM L-glutamine

(HyClone), 100 U/ml penicillin-streptomycin (HyClone), 1 mM nonessential amino acids (HyClone), ~10-50 ng/ml of recombinant LIF, 1 mM sodium pyruvate (HyClone) and 0.1 mM β -mercaptoethanol. Human 293T cells were cultured in DMEM high glucose containing 10% FBS and 100 U/ml penicillin-streptomycin. The *Hnrnpk* homozygous genetrapped mutant mESC line and the corresponding parental wt line generated by Horie *et al.*²⁴¹ were cultured on a MEF feeder layer in complete media. MEFs feeders were prepared by culturing HM1 immortalized MEFs to confluence in MEF media (DMEM high glucose containing 10% FBS, 1 mM glutamine, 100 U/ml penicillin-streptomycin) and incubation for 2-3 h in 10 μ g/ml mitomycin C (Sigma-Aldrich). MEF feeder plates were used within 7 days from preparation. All cell lines were grown at 37°C with 5% CO₂. Cells were generally passaged every 2-3 days and experiments were performed between passages 2-16 and frozen after passage 16. For induction of endogenous *Setdb1* deletion in the *Setdb1*^{lox/-} 3XFLAG-*Setdb1* line, cells were cultured in 800 nM 4-OHT for 4 days as described.⁸⁹ For blocking SUMO E1 activity, anacardic acid (Sigma-Aldrich) diluted to 5–100 μ M in DMSO was incubated with mESCs for 18 h prior to harvest. For inhibition of Pol II, mESCs were cultured in 100 μ M Triptolide in DMSO for 8.5 h prior to harvest. For growth curve assays, after the second siRNA transfection, 30,000 cells were plated in triplicate wells of a 24-well plate and wells were cultured for the indicated time-points in complete mESC media which was changed every ~24 h. Viable cells were counted using an automated counter (BioRad) and trypan blue staining.

2.2 RNA interference

RNAi by siRNA transfection of mESCs was performed as described.^{89,237} SMARTpool siRNAs containing 4 independent siRNAs to the target gene were synthesized by Dharmacon

(ThermoFisher). Cells were grown to ~70-80% confluence in the absence of antibiotics and transfected with 100 nM siRNA using Opti-MEM serum-free media (Life Technologies). After passaging, cells were transfected again with 50 or 100 nM siRNA (determined empirically based on mRNA or protein depletion and effects on cell viability). Twenty-four hours later, cells were generally harvested to assay mRNA and/or protein depletion. Downstream phenotypes were often assessed 48-96 h post-transfection, where indicated.

2.3 Site-directed mutagenesis and plasmid transfection

The pcDNA3.3-T7-*HNRNPK* plasmid expressing N-terminal T7 epitope-tagged full-length human hnRNP K isoform 1²⁴² (NCBI accession NP_001288270) was mutagenized using the Quickchange II lightning kit (Stratagene). To generate the G400R, Y458D, and Δ RGG (Δ 240-338 a.a.) mutants, primer sequences were as follows where the asterisk denotes the mutant base: 5' ACT ATT CCC AAA GAT TTG GCT A*GA TCT ATT ATT GGC AAA GGT G 3' (G400R), 5' CAG AAC AGT GTG AAG CAG G*AT GCA GAT GTT AAG GGA T 3' (Y458D) and 5'GAA ACC TAT GAT TAT GGT GGT TTT ACA*----TTC AGT GCT GAT GAA ACT TGG 3' (Δ -RGG). Successful mutagenesis was verified by Sanger sequencing using a CMV promoter primer: 5'CGC AAA TGG GCG GTA GGC GTG 3'. The hnRNP K expression constructs were transiently transfected into TT2 or *G9a*^{-/-} mESCs using lipofectamine 3000 (Life Technologies) according to the product instructions. The pSG5 plasmid harbouring FLAG-tagged mouse *Setdb1* cDNA²⁴³ and the pcDNA3.3-T7-*HNRNPK* plasmid were transiently transfected into 293T cells using lipofectamine 3000 according to the product instructions. Cells were harvested at 48 h post-transfection for IP and western blot analysis. The pET16b-*HNRNPK* plasmid expressing 6X-His-tagged human hnRNP K isoform 1 was expressed in BL21(DE3) *E.coli* and purified as described.²⁴⁴

2.4 Immunofluorescence and Flow Cytometry

Indirect immunofluorescence staining was performed using standard methods. Cells were grown on coverslips or harvested by trypsinization were crosslinked with 4% formaldehyde, permeabilized with 0.25% triton-X-100 and blocked with 1% bovine serum albumin (Sigma-Aldrich). Cells were then incubated with anti-hnRNP K (Abcam ab70492) overnight at 4°C and subsequently incubated with Alexa Fluor 488-labeled secondary antibody (Life Technologies). DNA was counterstained with 10 µg/ml Hoescht 33342 (Sigma-Aldrich). Flow cytometry was performed according to previous methods.^{89,237,245} Cells were trypsinized and incubated in 0.5 µg/ml propidium iodide (Sigma-Aldrich) in FACS buffer (phosphate buffered saline containing 3% FBS) and analyzed on a BD LSR II flow cytometer using BD FACS Diva software. Cells were successively gated on forward and side scatter, then PI- (live cells) and lastly GFP⁺ cells, using the untransfected mESC line (MSCV-GFP, IAP-GFP or 2C::*Gfp*) as a GFP⁻ population to set the gates. 10,000 PI cells were sampled in each replicate. SSEA1 and Annexin V staining were detected on mESCs using 1:400 anti-SSEA1 PE-conjugate (BD Pharmingen) or 1:1000 anti-Annexin V Alexa Fluor 488-conjugate (Life Technologies). Where indicated, cells were gated for the SSEA1⁺ population prior to GFP gating to identify SSEA1⁺; GFP⁺ (double-positive) cells. Cell cycle distributions were determined by PI staining. Cells were trypsinized, washed with phosphate buffered saline and fixed for >2 h in ice-cold 70% ethanol. Cells were subsequently incubated with 10 µg/ml PI stain (Sigma-Aldrich) for 15-30 minutes on ice and analyzed on the BD LSR II flow cytometry with FACS Diva software. Profiles were fitted to the Dean-Jett-Fox cell cycle model using FlowJo software to determine percentages of cells in G1, S and G2/M.

2.5 Native and crosslinked ChIP

For native ChIP of histone lysine methylation, equivalent numbers of cells per line (approximately 1, 2 or 3×10^6) were lysed in NChIP buffer (20 mM HEPES pH 7.9, 50 mM KCl, 1 mM $MgCl_2$, 3 mM $CaCl_2$, 1 mM DTT, 0.5% NP-40, 10% glycerol) containing protease inhibitors (Roche). Cells were digested with 50 U/ml MNase (Worthington Biochemicals) at 37°C for 7-10 minutes to produce mono- and di-nucleosomes, quenched with 5 mM EDTA and 5 mM EGTA and clarified by centrifugation. KCl concentration was adjusted to 150 mM and soluble chromatin was subsequently immunoprecipitated with: 5 μ g of anti-H3K9me3 (Active Motif 39161), 5 μ g anti-H4K20me3 (Active Motif 39180), 5 μ g anti-H3K9me2 (Abcam Ab1220) or 5 μ g of mouse IgG (Sigma-Aldrich I8765) overnight at 4°C. Native ChIP for H3K9me2 in TT2 wt and *Glpr*^{-/-} cells for ChIP-seq was performed according to a previous method²⁴⁶ by Carol Chen in Lorincz lab. Crosslinked ChIP for hnRNP K, KAP1, G9a, SETDB1 or SUMO1 was performed according to a previous method.⁸⁶ Crosslinked chromatin was sonicated to a size range of ~200-600 bp and immunoprecipitated overnight at 4°C with anti-hnRNP K (ab70492), anti-KAP1 (ab22553), anti-G9a (R&D Systems PP-A8620A-00), anti-SETDB1 (sc-66884) or anti-SUMO1 (Santa Cruz Biotechnology FL-101, sc-9060). After reversal of crosslinks samples were RNase A-treated and DNA was purified over silica columns. Quantitative PCR was performed with primers indicated in **Table 1**.

Table 1. List of PCR primers used in this study.

ChIP Target	ChIP primer Name	Sequence (5'-3')
MLV 5' LTR-gag	MLV +2	TGGGCAGGGGTCTCCAAATCT
	MLV -2	ATAAAGCCTCTTGCTGTTTGCATC
IAPEz 5' LTR-gag	IAP-LTR-ChIP-s	GCTCCTGAAGATGTAAGCAATAAAG
	IAP-retro-ChIP-as	CTTCCTTGCGCCAGTCCCGAG
IAPEz pol	4300-s	AAGCAGCAATCACCCACTTTGG CAATCATTAGATG(T/C)
	4400-as	GGCTGCCAAG
MMERVK10c 5' LTR-gag	MMERVK10C_LTR_344- fw	CAAATAGCCCTACCATATGTCAG
	MMERVK10C_int_481-rv	GTATACTTTCTTCTTCAGGTCCAC
Etn 5' LTR-MusD gag	MusD&Etn2	CCCTTCCTTCATAACTGGTGTGCGCA
	MusD3&Etn	TAGCATCTCTCTGCCATTCTTCAGG GCT GCG TGA GAC TGA GTA ACT GC
MERVL 5' LTR-gag	MERVL_LTR-277-fw	GTG GAT CAA CAA GCG AAT TCC A
	MERVL_LTR-524-rv	ATCTCCTGGCACCTGGTATG
MERVL pol	pol-F	AGAAGAAGGCATTTGCCAGA
	pol-R	AACCATCAGATGTTTCCAGGGTG
MSCV 5' LTR-PBS	LTR 302	TTCGGATGCAAACAGCAAGAGGC
	LTR -500	ACTACAACAGCCACAACGTCTA- TATCA
MSCV Gfp internal region	GFPcF	GGCGGATCTTGAAGTTCACC
	GFPcR	AGAACCAAACAGCCCAGAAG
Ifna5 core promoter	Ifna5 TSS-F	TCA GGA AAG CAC AGA GCC TA
	Ifna5 TSS-R	CCTCCCGAGTTCCCAAAGC
Myc core promoter	Myc TSS-F	CCT TCC CCA CCT CTC TCT ATT
	Myc TSS-R	TTTCAAGGGTCTGGAACAGC
Egr1 core promoter	Egr1 TSS-F	GGAGGGCTTCACGTCACTC
	Egr1 prom-R	TAC AAA ATG CCC GCA GAA ATA G
Dazl core promoter	Dazl qPCR F2	CCG GAC TCA ACC TTC TCA ATG
	Dazl qPCR R2	GTGGGATTCCGTTCTCTGAG
Fkbp6 core promoter	Fkbp6 prom-F	CTATCGATCACGTGCCGTTT
	Fkbp6 prrom-R	GTGGGGAGAGCACTGTGATT
Mael core promoter	Mael prom-F	GTT CGG GAA TGT GAA ACA CC
	Mael prom-R	CTCCCTCAGTGGGTGCTTC
Taf7l core promoter.+1.5kb	Taf7l prom-F	AGG CGA CGT GTC TCG TTA GT
	Taf7l prom-R	AAAGAAGCTAATGCCCTTGATTT
2C::Gfp gag-Gfp	gag-Gfp-F	GCTCTCGTCTGCTCTCCATAA
	gag-Gfp-R	TACAGCTTCGTGGTGGACAG
2C::Gfp Gfp-T2A	Gfp-T2A-F	CTTGAAGGCGTGCTGGTACT
	Gfp-T2A-R	GGCTTGATGAAGTCACCAGTAA
Chr1 Intergenic region	Chr1-Int-F	TCAGGCAGCCTCAAGTCT
	Chr1-Int-R	CAAAATCATTAGAA- TAATCGACAGGGC
Chr7 Intergenic region	Chr7-Int-F	TGGATGGTTAGTATCGCATGG
	Chr7-Int-R	

qRT Target	qRT primer Name	Sequence (5'-3')
MLV RLTR4 MM-int	EndPro-	AGAGGTATGGTTGGAATAAGTA
	EndPro+	TAGATGGAGCCTACCAAGCTCTCAA
IAPEz 5' LTR-gag	IAP-LTR-ChIP-s	Same as ChIP
	IAP-retro-ChIP-as	Same as ChIP
IAPEz pol	4300-s	Same as ChIP
	4400-as	Same as ChIP
MMERVK10C 5' LTR-int	MMERVK10C_LTR_344-fw	Same as ChIP
	MMERVK10C_int_481-rv	Same as ChIP
Etn 5' LTR- MusD int	Etn/II-MusD 29 up_fw	TCTTAATAATAGGGAGGAGATGTAGT
	Etn/II-MusD LTR_123_rv	GGTCCAGTAGAAAGGTGCGACG
MERVL 5' LTR-gag set 1	MERVL_LTR-365-fw	CTTCCATTACAGCTGCGACTG
	MERVL_int-519-rv	CTAGAACCCTCCTGGTACCAAC
MERVL 5'LTR-gag set 2	MERVL_LTR-277-fw	Same as ChIP
	MERVL_int-524-rv	Same as ChIP
MERVL pol	pol-F	Same as ChIP
	pol-R	Same as ChIP
MTA	MTA-F	ATGTTTTGGGGAGGACTGTG
	MTA-R	AGCCCCAGTTAACCAGAAC
Setdb1 exons 18-19	Setdb1 ex18-F	CTTCTGGCTCTGACGGTGATG
	Setdb1 ex18-R	GGAAGCCATGTTGGTTGATT
Hnrnpk exons 8-9	Hnrnpk ex8-F	ATCAGAGTCTGGCAGGAGGA
	Hnrnpk ex9-R	TCAGTAGAGTGAGGGCAGCA
Mcaf1 exons 11-12	Mcaf1 ex11-F	GACGATGAAGAGAGCGGAAC
	Mcaf1 ex12-R	CTGATGGCTGAAGAGGTGGT
Ubc9 exons 4-5	Ubc9 ex4-F	GATGGCACAATGAACCTGATGAAC
	Ubc9 ex5-R	TTTGGTGGTGAGGACGGATAGTC
Gata6 exons 3-4	Gata6 ex3-F	CCCTCATCAAGCCACAGAAG
	Gata6 ex4-R	CATATAGAGCCCCGAAGCAT
Gata3 exons 3-4	Gata3 ex3-F	TACCACCTATCCGCCCTATG
	Gata3 ex4-R	ACACACTCCCTGCCTTCTGT
Nkx2-9 exon 1	Nkx2-9 ex1-F	CCTCGCACTTCCATCCTTAG
	Nkx2-9 ex1-R	CTAGGCTTCGCATCCTGTTC
Dazl exon 4	Dazl ex4-F	GTGGCTTCTGCTCCACCTTCG
	Dazl ex4-R	CCTTGACTTGTGGTTGCTGA
Fkbp6 exons 3-4	Fkbp6 ex3-F	TACCTGGAGCACATGGACAA
	Fkbp6 ex4-R	ATAGGCTGGCTTGAACAGGA
Mael exons 11-12	Mael ex11-F	GGTTATTATCCATCTGGGGTGA
	Mael ex12-R	CATCTTTTTGGGGCGTGTA
Taf7l exons 10-11	Taf7l ex10-F	TAGCCTCCATGAAGCAGACC
	Taf7l ex11-R	TTGAGGGTCAAGTTTTCCACTT
Egr1 exons 1-2	Egr1 ex1-F	GAGCGAACAACCCTATGAGC
	Egr1 ex2-R	TTTGGCTGGGATAACTCGTC
Gapdh exons 4-5	Gapdh ex4-F	GTG CTG AGT ATG TCG TGG AG
	Gapdh ex5-R	GAG ATG ATG ACC CTT TTG GCT C
B-actin exons 2-3	b-actin-ex2-fw	GAACCCTAAGGCCAACCGTG
	b-actin-ex3-rv	GGAGTCCATCACAATGCCTG
45S pre-rRNA	45S-F	GGTTGGAATGGTGGAGCC
	45S-R	AAAGAGCGGAGGTTCTGGG
28S rRNA	28S-F	GGTTCAGATCCCCGAATC
	28S-R	AAAGAGAACTCTCCCCGAGG
2C::Gfp gag-Gfp	gag-Gfp-F	Same as ChIP
	gag-Gfp-R	Same as ChIP

RT-PCR primer	Primer name	Sequence (5'-3')
Random 15-mer	15N	NNNNNNNNNNNNNNNN
Oligo dT(18)	18T	TTTTTTTTTTTTTTTTTT
MERV1 LTR-gag-s	MERV1 277-LTR-f	Same as ChIP
MERV1 gag-as	MERV1 524-gag-r	Same as ChIP

2.6 Immunoprecipitation

Nuclear extracts were prepared from mESCs as previously described.⁸⁶ Extracts were prepared with or without 10 or 20 mM NEM and phosphatase inhibitors (Roche), as indicated, diluted to ~170 mM KCl and immunoprecipitated overnight at 4°C with anti-hnRNP K (Abcam ab39975), anti-GLP (R&D Systems PP-B0422-00), anti-SETDB1 (kind gift from Dr. H.H., Ng Singapore Genome Institute), anti-KAP1 (ab22553), rabbit IgG (Sigma-Aldrich I8640) or mouse IgG (Sigma-Aldrich I8765). Where indicated 40 U/ml Ribolock (-RNase) or 40 µg of RNase A and 100 units of RNase T1 as a mix (Fermentas) (+RNase) was added to the extracts. DNase and RNase treatment together were performed by adding 50 U/ml of DNase I (Promega) and 50 µg/ml RNase A (BioShop) to the diluted extract. Complexes were collected with protein A and G dynabeads (Life Technologies), washed three times in IP wash buffer (20 mM HEPES pH 7.9, 170 mM KCl, 1.5 mM MgCl₂, 0.3% NP-40, 10% glycerol) and eluted by boiling in NuPAGE LDS sample buffer (Life Technologies). For IP of FLAG-tagged SETDB1 from the *Setdb1*^{lox/-} cells expressing 3XFLAG-*Setdb1*, the endogenous *Setdb1* was first deleted using 4-OHT and then nuclear extracts were prepared from these cells with uninduced *Setdb1*^{lox/-} lacking the FLAG-*Setdb1* construct, as a negative control. Extract was diluted to ~170 mM KCl, NEM was added to 10 mM and immunoprecipitated overnight at 4°C with anti-DYKDDDDK (FLAG) antibodies. Antibodies were captured on protein G dynabeads, washed in IP wash buffer and eluted with

phosphate buffered saline containing 0.1% Tween-20 and 500 µg/ml 3XFLAG peptide (Sigma-Aldrich).

2.7 Silver staining of SDS-PAGE gels

After SDS-PAGE gels were stained with silver nitrate or colloidal coomassie. For silver staining, gels were first fixed in 40% methanol 10% acetic acid for 1 h or overnight at room temperature. Gels were washed 3 times for 5 minutes each in deionized water then incubated in 0.02% sodium thiosulfate for 1 minute and subsequently rinsed in water for 1 minute. Gels were then incubated for 15-20 minutes in chilled 0.1% silver nitrate solution in the dark. The impregnated silver was reduced with two washes of 2% sodium carbonate, 0.04% formaldehyde until bands developed. Staining was stopped by washes in 10% acetic acid and gels were stored in 1% acetic acid prior to drying in cellophane.

2.8 Western blot analysis

Western blots were performed using standard methods.⁸⁶ Nuclear extracts, IP samples, whole-cell extracts prepared with RIPA buffer (50 mM Tris pH 8.0, 150 mM NaCl, 1% NP-40, 0.25% sodium deoxycholate, 0.1% SDS), total cell lysates prepared in 1X NuPAGE loading buffer containing 100 mM DTT or 0.2 M HCl-extracted histones were separated by SDS-PAGE transferred to nitrocellulose or PVDF membranes, blocked in 5% skim milk in Tris-buffered saline (20 mM Tris-HCl pH 7.5, 150 mM NaCl) or Odyssey blocking reagent (LiCOR Biosciences). Blots were probed with primary antibodies overnight at 4°C including: anti-hnRNP K (Abcam ab39975 or ab70492), anti-LSD1 (Abcam ab37165 or ab17721), anti-G9a (R&D systems PP-A8620A-00), anti-GLP (R&D systems PP-B0422-00), anti-KAP1 (Abcam ab22553), anti-T7 (EMD Millipore 69522), anti-GAPDH (EMD Millipore

ABS16), anti-H3K9me2 (Abcam Ab1220), anti-H3 (Active Motif 39163), anti-SETDB1 (H300 sc-66884), anti-Ubc9 (Santa Cruz sc-5231), anti-SUMO1 (Santa Cruz sc-9060), anti-SENPI (Novus Biologicals NB100–92101), anti-H3K9me3 (Active Motif 39161), anti-pan H3ac (Millipore 06–599), anti-H4 (Millipore 04–858), anti-GST (GenScript A00097–100), anti-DYKDDDDK (FLAG epitope, GenScript). Primary antibodies were detected with IRDYE-conjugated secondary antibodies and scanned on the Odyssey imaging system. Quantification of bands was performed with Odyssey software (or Image J v1.6 for data shown in Figure 7G) and band intensities were normalized to loading controls (typically GAPDH). For experiments involving SUMO-modified KAP1, extracts were typically prepared with 10 or 20 mM NEM to enrich for SUMOylation.

2.9 Sucrose gradient sedimentation

Sedimentation of nuclear proteins over sucrose gradients was performed according to previous methods.^{247,248} Approximately ~2 mg of mESC nuclear extract or ~4 µg of recombinant proteins in 500 µl was layered onto a 5 ml linear 5–50% gradient and centrifuged in parallel with identical gradients containing purified molecular weight standards (blue dextran 52.6S/~2 MDa, thyroglobulin 19.4S/670 kDa, catalase 11.4S/250 kDa, BSA 4.3S/67 kDa all from Sigma-Aldrich) at 27,500 rpm (~91,900 g) in a SW-55Ti rotor (Beckman Coulter) at 4°C for 18.5 h. Fractions of 200 µl were collected from top to bottom including the pellet fraction and 20 µl samples were assayed by western blot. Peaks for migration of the molecular weight standards were determined by absorbance at 280 nm.

2.10 Mass spectrometry of SETDB1 complexes

Nuclear extracts were prepared from TT2 mESCs with or without 10 or 20 mM NEM and clarified by centrifugation. For immunoprecipitation of SETDB1 complexes after anionic column fractionation, approximately 12–15 mg of TT2 mESC nuclear extract (4 ml) was prepared without NEM, diluted with 2 volumes with 56 mM HEPES pH 7.9, 5% glycerol and passed over a 2 ml column of Macro HiQ anionic exchange media (BioRad) in an equilibration buffer (50 mM HEPES pH 7.9, 100 mM KCl, 10% glycerol). Bound proteins were washed with 5 column volumes of equilibration buffer and then eluted stepwise in 2 volumes of buffer containing 250 mM KCl, then 2 volumes of buffer containing 500 mM KCl. The 500 mM KCl fraction containing SETDB1 and depleted of SENP1 (4 ml) was then diluted with 2 volumes IP dilution buffer (20 mM HEPES pH 7.9, 0.5% NP-40, 10% glycerol containing 2 mM PMSF) and divided into two equal aliquots and immunoprecipitated overnight at 4°C with protein G sepharose beads crosslinked with ~100 µg of rabbit IgG (Sigma Aldrich) or rabbit anti-SETDB1 H300 (Santa Cruz Biotechnology) using dimethylpimelimidate as described.²⁴⁹ Beads were washed extensively with a wash buffer (20 mM HEPES pH 7.9, 200 mM KCl, 1% NP-40, 0.1% sodium deoxycholate, 10% glycerol) and eluted by boiling in SDS-PAGE loading buffer. For direct SETDB1 IP from mESC nuclear extract, ~7–8 mg of nuclear extract (1.5 ml) was diluted with 2 volumes of IP dilution buffer as above and incubated with 30 µg rabbit IgG or anti-SETDB1 H300 overnight at 4°C. Immunocomplexes were captured on protein G dynabeads, washed extensively with wash buffer as described above except omitting deoxycholate, eluted with 0.1 M glycine pH 2.5 and neutralized with 1.5 M Tris pH 8.8. Immunoprecipitated samples were analyzed by SDS-PAGE, western blot and silver staining.

For mass spectrometry, IgG and SETDB1 IP samples were resolved by SDS-PAGE and stained with colloidal coomassie as described above. The IgG heavy (~50 kDa) and light (~25 kDa) chain bands were removed first and discarded then the rest of each gel lane was excised and subjected to in-gel tryptic digestion as described.²⁵⁰ Extracted peptides were then analyzed by nano-flow liquid chromatography-tandem mass spectrometry (LC-MS/MS) on a LTQ-Orbitrap Velos Pro mass spectrometer (ThermoFisher).²⁵¹ Tandem mass spectra were searched against the UniProt mouse database using Mascot (v2.4, Matrix Science). Each IP sample was analyzed independently twice. The final refined hit list of proteins was filtered for nuclear proteins with enrichment ratios of SETDB1 IP/IgG IP (medium/light) of >2, >2 unique peptides and >2 independent spectra.

2.11 Recombinant proteins

GST-tagged hnRNP K and GST-tagged KAP1 and KAP1^{PxVxL} (residues 379–524) were purchased from Novus Biologicals. GST-KAP1^{PB} (residues 624–811) was from Cayman Chemical. Purified GST was from Sigma-Aldrich, the C-terminal GST-RanGAP1 fragment (residues 419–587) and GST-SUMO2 were from Enzo Life Sciences and GST-p53 was from Millipore. Purified FLAG-tagged SETDB1 protein was from Active Motif. 6XHis-tagged wt hnRNP K was purified from *E.coli* as described above and the hnRNP K C-terminal deletion mutant (hnRNP K Δ C a.a. 1-276) was from Novus Biologicals. Recombinant full-length human ZIK1 was from Abcam.

2.12 In vitro SUMOylation and GST pulldown assays

In vitro SUMOylation assays were previously described²⁴⁵ and performed according to previous methods²⁵² with minor modifications. For most assays, approximately 500 ng of GST-fused proteins were mixed with 125 ng Aos1/ Uba2 SUMO E1 heterodimer (Enzo Life Sciences BML-UW9330-0025), 500 ng SUMO E2 Ubc9 (Enzo Life Sciences BML-UW9320-0100) and 2 µg 6X-His-tagged SUMO1 (Enzo Life Sciences ALX-201-045-C500) and incubated in 20 µl of 1X SUMOylation buffer (50 mM Tris pH 8.0, 50 mM KCl, 5 mM MgCl₂, 1 mM DTT, 1 mM ATP) for 90 minutes at 30°C. For the *in vitro* SUMOylation of KAP1 under limiting conditions, the amounts of SUMO E1 and SUMO E2 were reduced to 62.5 ng and 250 ng, respectively and reactions were incubated for 15 minutes at 30°C. To determine the influence of various factors on SUMOylation efficiency of KAP1 and p53, 2, 4 or 6 µg of 6XHis-tagged hnRNP K was added (for titration experiments) or 1 µg of ZIK1 was added. Experiments with presence/absence of hnRNP K used 4 µg in each reaction. Negative control reactions were performed by omitting SUMO1. Following this, reactions were stopped either by addition of SDS-PAGE loading buffer for western blotting or prepared for pulldown assays. For pulldown assays, GST-tagged proteins (from SUMOylation reactions or just purified proteins) were immobilized on glutathione-agarose paramagnetic beads (GenScript), washed twice with pulldown buffer (50 mM Tris pH 8.0, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 0.01% Tween-20, 10% glycerol) and incubated with 0.5–5 µg recombinant prey proteins (SETDB1, Ubc9 or hnRNP K) in 150 µl pulldown buffer for 1.5 h at 4°C. For pulldowns with GST-tagged KAP1 mutant baits and 6X-His-tagged hnRNP K prey, BSA was included in the binding reaction at 1 mg/ml. Beads were washed again three times pulldown buffer for SETDB1 and Ubc9 or in pulldown buffer containing 300 mM

NaCl for hnRNP K and subsequently eluted with SDS-PAGE loading buffer for western blotting. Alternatively, glutathione elution buffer was used to elute bound proteins (50 mM Tris pH 8.0, 20 mM reduced L-glutathione, 1 mM DTT).

2.13 In vitro de-SUMOylation assays

GST-tagged KAP1 was SUMOylated *in vitro* as described above and purified from the reaction using glutathione-agarose paramagnetic beads (GenScript). DTT was added to 10 mM with bead-bound SUMOylated KAP1 and then mixed with 0-100 nM recombinant SENP1 catalytic domain (SENP1^{CD}) (Cayman Chemical) in de-SUMOylation buffer (50 mM Tris pH 8.0, 100 mM NaCl, 10 mM DTT, 0.01% Tween-20, 10% glycerol) and incubated at 30°C for 1 h. This determined that ~50-80 nM was the minimal concentration needed to de-SUMOylate KAP1 under these conditions. To test the influence of various factors on SENP1^{CD} de-SUMOylation, 500 ng of the following recombinant proteins were added to the reaction: FLAG-tagged SETDB1, T7-tagged ZIK1, or 6XHis-tagged hnRNP K. To confirm the specificity of de-SUMOylation, NEM was added to the reaction to 10 mM.

2.14 Native RNA immunoprecipitation

Native nuclear RIP assays were performed as described with modifications.²⁵³

Approximately 10⁸ TT2 mESCs were resuspended and lysed hypotonic lysis buffer (10 mM HEPES, pH 7.9, 10 mM KCl, 0.5 mM DTT, 1.5 mM MgCl₂) containing protease inhibitor cocktail (Roche), phosphatase inhibitors (Roche) and 40 U/ml Ribolock RNase inhibitor (Fermentas) with a dounce and tight pestle. After collection of the nuclear pellet by

centrifugation at 5000g for 5 minutes at 4°C the pellet was extracted for 30 minutes with high-salt nuclear extraction buffer (20 mM HEPES pH 7.9, 500 mM KCl, 1.5 mM MgCl₂, 20% glycerol) containing protease and phosphatase inhibitors and 40 U/ml Ribolock. The soluble nuclear extract was clarified 16,000g for 10 minutes and approximately 10% of the RIP volume of nuclear extract was set aside for input RNA isolation. The remaining nuclear extract was diluted with two volumes of IP dilution buffer (20 mM HEPES pH 7.9, 1.5 mM MgCl₂, 0.5% NP-40) and immunoprecipitated overnight at 4°C with 5 µg of anti-hnRNP K (Abcam ab39975), anti-GLP (R&D systems PP-B0422-00) or mouse IgG (Sigma-Aldrich I8765). Immunocomplexes were collected with protein G dynabeads (Life technologies) and washed 3 times in RIP buffer (20 mM HEPES pH 7.9, 170 mM KCl, 1.5 mM MgCl₂, 0.17% NP-40, 8.5% glycerol) containing protease inhibitors, phosphatase inhibitors and 40U/ml Ribolock. Bound protein-RNA complexes were incubated with RIP elution buffer (10 mM Tris pH 8.0, 10 mM DTT, 0.5 mM EDTA, 1% SDS) at 37°C for 15 minutes. For western blot analysis 50% of the bead volume was boiled in 1X NuPAGE loading buffer. RNA was isolated from the eluted volume with Trizol reagent (Sigma-Aldrich or Ambion) and precipitated with isopropanol in the presence of 300 mM sodium acetate and 20 µg glycogen at -80°C overnight. Pellets were washed with 75% ethanol, air-dried and resuspended in RNase-free 10 mM Tris pH 8.5 (Sigma-Aldrich) and stored at -80°C until cDNA synthesis.

2.15 Nuclear RNA pulldown assays

Nuclear RNA was purified from TT2 mESCs by first preparing nuclei as outlined above for native RIP. After cell lysis and centrifugation the nuclear pellet was washed once in PBS containing 40U/ml Ribolock inhibitor and then RNA was extracted with Trizol reagent

(Sigma-Aldrich or Ambion) and precipitated with isopropanol. Contaminating genomic DNA was removed with RNase-free DNase I (Promega) incubation for 1 h at 37°C. For pulldown assays, approximately 10 µg of recombinant 6X-His-tagged wt hnRNP K or C-terminal deletion mutant ($\Delta 277-464$ a.a.) purchased from Novus Biologicals were immobilized on Ni²⁺ HIS-select affinity gel (Sigma-Aldrich). After removal of ~0.6 µg DNase I-digested nuclear RNA as the 10% input, bound hnRNP K proteins were incubated with ~6 µg of the nuclear RNA for 1 h at 4°C rotating in a binding buffer (50 mM Tris, 150 mM NaCl, 5 mM MgCl₂, 10 mM DTT, 0.01% Tween-20, 10% glycerol) containing 40 U/ml Ribolock inhibitor. Beads were washed once in binding buffer and twice in binding buffer except containing 350 mM NaCl and eluted in RIP elution buffer, as outlined in the RIP protocol above. RNA was purified with Trizol, precipitated with isopropanol and stored at -80°C until cDNA synthesis.

2.16 EMSA

For electrophoretic mobility shift assays, approximately 100-200 pmol of 5'IRDYE700-conjugated DICE probe (prepared by IDTDNA) was incubated with 2.5 or 10 µg of purified 6X-His-tagged wt hnRNP K or the hnRNP K ΔC mutant protein ($\Delta 277-464$ a.a.) (Novus Biologicals) in 10 µl of EMSA buffer (10 mM Tris pH 7.5, 50 mM NaCl, 15 mM DTT, 0.5% Tween-20) at 23°C. Purified recombinant 6X-His-tagged wt hnRNP K was produced in BL21(DE3) *E.coli* as previously described.²⁴⁴ Samples were run on 5% (37.5:1 acrylamide:bisacrylamide) non-denaturing gels in 50 mM Tris pH 7.5, 380 mM glycine, 2 mM EDTA at 60-80 V until the free probe had reached the bottom of the gel. Gels were imaged on the Odyssey scanner (LiCOR Biosciences).

2.17 Cellular fractionation RNA extraction and RT-PCR

For RT-PCR assays on cell fractionated RNA, nuclear RNA was isolated as described above for RNA pulldown assays. After cell lysis and centrifugation, the supernatant was kept as the cytoplasmic RNA fraction. RNA was isolated from nuclei and the cytoplasmic supernatant with Trizol reagent (Ambion). Total RNA was extracted from mESCs using the GenElute total RNA Kit (Sigma-Aldrich) according to product instructions. For end-point RT-PCR assays, RNA was treated with 500 U/ml DNase I (Promega), and first strand cDNA was synthesized with SuperScript IV (Life Technologies) and random 15-mer oligos, oligo d(T)₁₈, 5'LTR forward (for antisense RNA) or gag reverse (for sense RNA) primers according to the product instructions. MERVL sequences were amplified by 24-30 cycles, whereas rRNA was amplified by 17-20 cycles.

2.18 Quantitative RT-PCR, qPCR, RNA-seq and ChIP-seq bioinformatics

For expression and enrichment analysis by qRT-PCR, RIP and nuclear RNA pulldown, total RNA, or RNA isolated from 10% input and native RIP and RNA pulldown samples was DNase I-treated and first-strand cDNA synthesis was performed using SuperScript III or SuperScript IV reverse transcriptase (Life Technologies) as described with random 15-mers.^{86,245} Quantitative PCR was performed on cDNA and isolated input or ChIP DNA as described⁸⁶ with primer sequences indicated in **Table 1**. For all RT-PCR-based assays, the specificity of the PCR for detecting cDNA was confirmed using samples lacking reverse transcriptase (-RT). *Gapdh* or *Actb* (β actin) mRNAs were amplified in qRT-PCR experiments as endogenous control genes.

Strand-specific, paired-end mRNA-seq on poly(A) RNA was performed as previously described.^{245,246} Libraries were sequenced on the Illumina HiSeq 2000 at the BC Genome Sciences Centre. Reads per kilobase per million mapped reads (RPKM) was calculated and genes up- or downregulated relative to control siRNA KD cells were determined by applying fold-change threshold of 2 and minimum read count of 25 for genes up (down) regulated in hnRNP K KD (control) cell lines by Dr. M. Karimi. Gene ontology analysis was performed with DAVID bioinformatic resource version 6.7 at <http://david.abcc.ncifcrf.gov/home.jsp>. RNA-seq coverage over RefSeq annotated genes and Rebase annotated retrotransposons in control or hnRNP K KD cells was calculated Dr. M. Karimi. Dr. Karimi also generated a list of genes upregulated >2-fold in *G9a* and *Glp* KO mESCs, from previous RNA-seq analysis of these cells in the lab. ChIP-seq RPKM coverage for H3K9me2 in TT2 wt and *Glp*^{-/-} cells and corresponding track files were generated by Carol Chen of the Lorincz lab. Chase software <http://chase.cs.univie.ac.at/> was used to generate H3K9me2 heatmaps over the flanking 6 kb 5' upstream or 3' downstream of 656 individual MERVL elements in the C56Bl/6 genome. For the H3K9me2 ChIP-seq data, only uniquely aligned reads were used in generating RPKM values, heatmaps and UCSC browser track files.

2.19 Statistics and data analysis

Experiments were generally repeated at least once, with most of the major experiments performed 3 or more times to gauge the level of reproducibility. Data presented are from representative experiments. To determine statistical significance, student's two-tailed T-tests were performed on n = 3 replicates, where indicated.

3. hnRNP K is a novel co-repressor for the SETDB1/KAP1 complex in proviral silencing

The results presented in this chapter have been published: Peter J. Thompson, Vered Dulberg, Kyung-Mee Moon, Leonard J. Foster, Carol Chen, Mohammad M. Karimi and Matthew C. Lorincz (2015) **hnRNP K coordinates transcriptional silencing by SETDB1 in embryonic stem cells**. *PLoS Genetics* 11(1):e100493.

3.1 Background and Summary

As detailed in the introduction, undifferentiated PGCs and pluripotent stem cells derived from the inner cell mass of the blastocyst, such as murine embryonic stem cells (mESCs) and mECCs utilize a DNA methylation-independent pathway to maintain ERV silencing.²⁰⁵ Key effectors in this silencing pathway are the conserved KRAB-ZFPs.²⁰⁸ KRAB-ZFPs are thought to function in retroviral silencing by binding to specific proviral sequences such as the PBS, to direct the recruitment of a large silencing complex that includes the obligate co-repressor KAP1^{212,220} and KMT SETDB1, which deposits H3K9me3 to maintain a repressive chromatin state^{89,215} (**Illustration 4**). Although prototypical KRAB-ZFP candidates for this pathway have been identified, such as ZFP809 and ZFP819,^{220,254} and other work has shown that SETDB1/KAP1 are recruited to specific loci by other KRAB-ZFPs outside the context of ERV silencing, such as ZNF274, ZNF350 and ZFP57^{91,255,256} it remains unclear whether PBS binding is a general property of most KRAB-ZFPs or only a select few. Consistent with observations that PBS sequences alone are insufficient to confer SETDB1/KAP1-mediated silencing,⁹⁰ the transcription factor YY1 was shown to be required for silencing of the newly integrated MLV-based retroviruses in F9 embryonal carcinoma cells and mESCs,²⁵⁴ revealing that additional sequence-specific factors may collaborate with KRAB-ZFPs. In addition, KAP1 is apparently recruited to IAP elements via sequences in the 5'UTR⁹⁰ and in the *gag* coding region.²⁵⁷

In mESCs but not MEFs, both class I and II ERVs and newly integrated MLV-based retroviral vectors are marked with H3K9me3 by a SETDB1/KAP1-containing complex⁸⁹ and similarly ERVs are also marked by H3K9me3 and are silenced in a SETDB1-dependent

manner in E13.5 PGCs.⁹⁵ Conditional knockout of *Setdb1* in undifferentiated mESCs or E13.5 PGCs abolishes H3K9me3 at ERVs and leads to reduced levels of DNA methylation and increased 5-hydroxymethylation²²³, concomitant with pervasive de-repression of distinct class I and II ERV families including MLV, IAP, MMERVK10C and MusD elements.^{89,95,246} A similar phenotype is apparent upon deletion of *Kap1* in mESCs⁹⁰ and KAP1 is required for SETDB1 recruitment, since depletion of KAP1 leads to a loss of SETDB1 binding and H3K9me3 at ERVs and newly integrated MLV-based vectors^{89,90} (**Illustration 4**). Although global DNA methylation at most ERVs is unperturbed in *Setdb1* deletion mESCs,⁸⁹ in the absence of *de novo* methylation the *Setdb1* KO leads to reduced 5mC and increased 5hmC at ERVs²²³ suggesting that a proportion of 5mC is maintained by H3K9me3 and SETDB1 in mESCs (**Illustration 4**).

The Small ubiquitin-like modifier (SUMO) paralogue SUMO1 is conjugated to KAP1 via the autocatalytic SUMO E3 ligase activity of the plant homeodomain (PHD) zinc finger towards the bromodomain at the major lysine acceptor sites K554, K779 and K804 to direct SETDB1 recruitment and H3K9 methylation at heterologous promoters.^{258,259} SUMOylation of KAP1 is also required for SETDB1 binding *in vitro*,²⁵⁸ however, the role of SUMOylation in SETDB1-mediated repression of ERVs and the involvement of additional co-repressors in ERV silencing have not been addressed (**Illustration 4**). For example, although HP1 proteins participate in silencing heterologous promoters in transformed human cell lines²¹⁸ and are bound at ERVs in a SETDB1-dependent manner in mESCs,⁸⁹ Maksakova *et al.*²³⁷ reported that knockout of the genes encoding HP1 α or HP1 β in mESCs, or siRNA KD of the paralogous members in each KO line did not reactivate MLV-based or IAP LTR-driven

proviral reporter constructs in mESCs. In addition, siRNA KD screening of other known H3K9me3-binding proteins indicated that they were dispensable for proviral silencing.²³⁷ This raises the intriguing possibility that H3K9me3-binding proteins are not required for SETDB1/KAP1-mediated ERV silencing in mESCs and that the mark itself might be sufficient to prevent transcriptional activity on proviral elements. Similarly, KD of HDAC1 and MBD1, the latter of which was shown to associate with SETDB1 in HeLa cells,²⁶⁰ does not de-repress newly integrated MSCV proviruses in mESCs.⁸⁹ Furthermore, mESCs deficient for the chromatin remodeler ATRX only show modest de-repression of ERVs and de-repression is potentiated in *Setdb1* KO cells,²⁵⁷ suggesting it acts in a different pathway.

Although SETDB1-containing complexes have been purified from human transformed cell lines (e.g. HeLa and HEK293T),^{33,34} no studies have systematically investigated the binding partners of SETDB1 in a pluripotent stem cell type where this KMT functions in ERV silencing. In addition, both of the previous reports of purifying the SETDB1 complex made no attempts to address the presence of SUMOylation in stabilizing the binding of SETDB1 with KAP1 and other binding partners. To investigate the mechanisms governing the function of the SETDB1/KAP1 complex in proviral silencing in mESCs, I initially used a biochemical approach of isolating SETDB1 binding partners under different conditions with the aim of identifying novel co-repressors. This led to the identification of hnRNP K as a novel SETDB1-interacting protein in mESCs. Subsequent genetic analyses supported the role for hnRNP K as a unique member of the SETDB1/KAP1 assembly by virtue of its influence, chromatin protein SUMOylation, H3K9me3 deposition and SETDB1 recruitment, independently of an effect on KAP1 binding at proviral chromatin.²⁴⁵

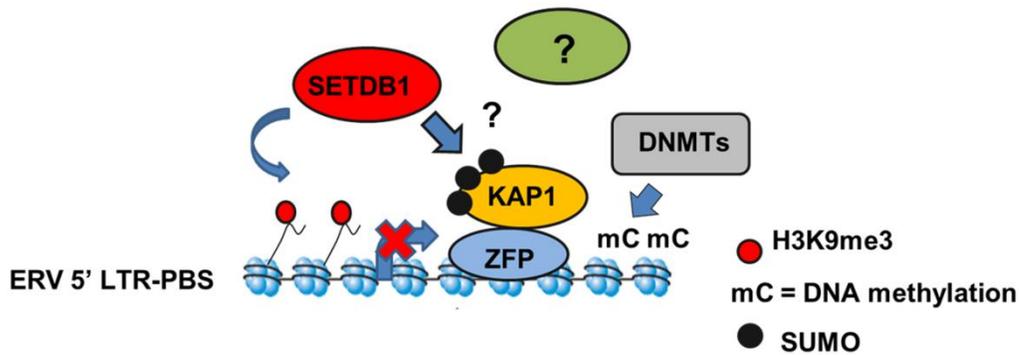


Illustration 4. Current model for transcriptional silencing of class I and II ERVs by the SETDB1/KAP1 complex. ERV and retroviral vector sequences are recognized by KRAB-ZFPs and other zinc finger proteins such as YY1, some of which may bind the PBS region. KRAB-ZFPs recruit their co-repressor KAP1, which in turn recruits SETDB1 to deposit H3K9me3 and maintain proviral silencing. Although KAP1 is known to be SUMOylated at 3 major lysines, which enables its direct binding to SETDB1 *in vitro* it is unclear what role KAP1 SUMOylation plays in SETDB1 recruitment and proviral silencing in mESCs. Furthermore, it is unknown whether additional co-repressors are required for proviral silencing with SETDB1/KAP1. Note: additional proteins are known to bind to ERV sequences and/or interact with KAP1, including HP1 proteins, HDACs and the chromatin remodeler ATRX, which are not shown above, as their role in the maintaining silencing is either unclear or not required. DNA methylation at ERVs is promoted by H3K9me3, but is generally dispensable for proviral silencing.

3.2 Proteomic analysis of SETDB1 complexes in mESCs

To identify novel factors involved in SETDB1-dependent ERV repression, I first characterized endogenous SETDB1-containing complexes from mESCs by IP-MS (**Figure 1**). In the first approach, I utilized conditions that minimize de-SUMOylation of proteins given the SUMO-dependent interactions between SETDB1 and KAP1.²⁵⁸ This approach was necessary because the SUMO-specific proteases SENP1 and SENP7 are known to de-SUMOylate KAP1^{261,262} and are expressed in mESCs.²⁶³ To enrich for candidate SUMO-dependent binding partners of SETDB1, I performed an anionic exchange step which efficiently depleted SENP1 followed by IP of endogenous SETDB1 with a specific N-terminal antibody²⁶⁴ (**Figures 1A and 1B**). MS analysis revealed the specific enrichment of

KAP1 along with the previously described SETDB1 co-factor MCAF1 (also called mAM/Atf7ip) (**Figure 1C**), which directly interacts with SETDB1 independent of SUMOylation.^{33,258} Detection of MCAF1 and KAP1 supported the validity of this approach to identify candidate SUMO-independent and SUMO-dependent binding partners.

In addition, among the high-confidence hits were several novel factors including: hnRNP K, BAF155, TRIP12, Nup155, MCM5, ZFP161 and PRMT1 (**Figure 1C**). Among these, BAF155, hnRNP K, ZFP161 (a BTB/POZ domain zinc finger protein) and arginine methyltransferase PRMT1 are known to play roles in transcriptional regulation^{265,266} (work on hnRNP K reviewed by Han *et al.*¹³⁸, work on PRMT1 reviewed by Nicholson *et al.*²⁶⁷). In contrast, MCM5 is a subunit of the eukaryotic MCM2-7 helicase complex active during S-phase,²⁶⁸ Nup155 is a conserved nuclear pore complex protein²⁶⁹ and TRIP12 an essential E3 ubiquitin ligase.²⁷⁰ With the exception of BAF155, which can interact with KAP1 as a subunit of the mESC-specific BAF chromatin remodelling complex,^{271,272} none of these factors had been shown to interact with SETDB1 or KAP1 or to be involved in ERV silencing in any cell type.

To identify factors associated with SETDB1 in mESCs under conditions where SUMOylation is not protected, I performed a SETDB1 IP directly from mESC nuclear extract without prior SENP depletion (**Figure 1D**). MS analysis revealed a different set of polypeptides associated with SETDB1 (**Figure 1E**). While MCAF1 was identified among the high-confidence hits in this direct IP approach, KAP1 was not (**Figure 1E**), indicating that the presence of SENPs in mESC nuclear extracts precludes the association of SETDB1 with

its SUMO-dependent binding partners, including KAP1. In addition, there were several unique interactors including: Nucleolin, DDX21, PAPBC1, TOP2A, RUVBL1, ILF2, NOLC1, and hnRNP A/B (**Figure 1E**). With the exception of TOP2A and RUVBL1, all of these factors are known to bind RNA and to regulate RNA processing and/or trafficking^{138,273–275} (roles of Nucleolin are reviewed by Durut and Sáez-Vásquez²⁷⁶). ILF2 also can regulate transcription.²⁷⁷ TOP2A is a DNA topoisomerase with functions in genome stability (reviewed by Chen *et al.*²⁷⁸) and RUVBL1 is an ATP-dependent helicase and subunit of the NuA4 histone acetyltransferase complex and the SWR1-like complex (reviewed by Nano and Houry²⁷⁹).

I hypothesized that among these novel interactors, the factors associated with SETDB1 in the SENP-depleted IP but not the direct IP would be good candidates for involvement in the SETDB1/KAP1 ERV silencing pathway, since the pathway would likely require SUMO-dependent interactions between SETDB1 and KAP1. Among these, I focused on the most enriched protein after KAP1, hnRNP K (**Figure 1C**), which is a broadly expressed nuclear protein originally identified as part of the ~20 subunit heterogeneous nuclear RNA protein complex¹³⁶ that binds to DNA and RNA, regulates mRNA stability and translation and functions as a transcriptional co-activator or co-repressor in different contexts.¹³⁸ Notably, *Hnrnpk* is highly expressed in the inner cell mass and in mESCs relative to earlier stages of development in the preimplantation embryo²⁶³ and was previously reported to directly interact with the KRAB-ZFPs Zik1 and Kid1.^{139,280} HnRNP K domain structure includes three KH domains, which bind C-rich DNA and RNA sequences,^{281,282} and a core protein-

protein interaction RGG box, comprised of repeats of Arg-Gly-Gly, which mediates most of its protein-protein interactions (**Figure 1F**, reviewed by Bomszyk *et al.*¹³⁹).

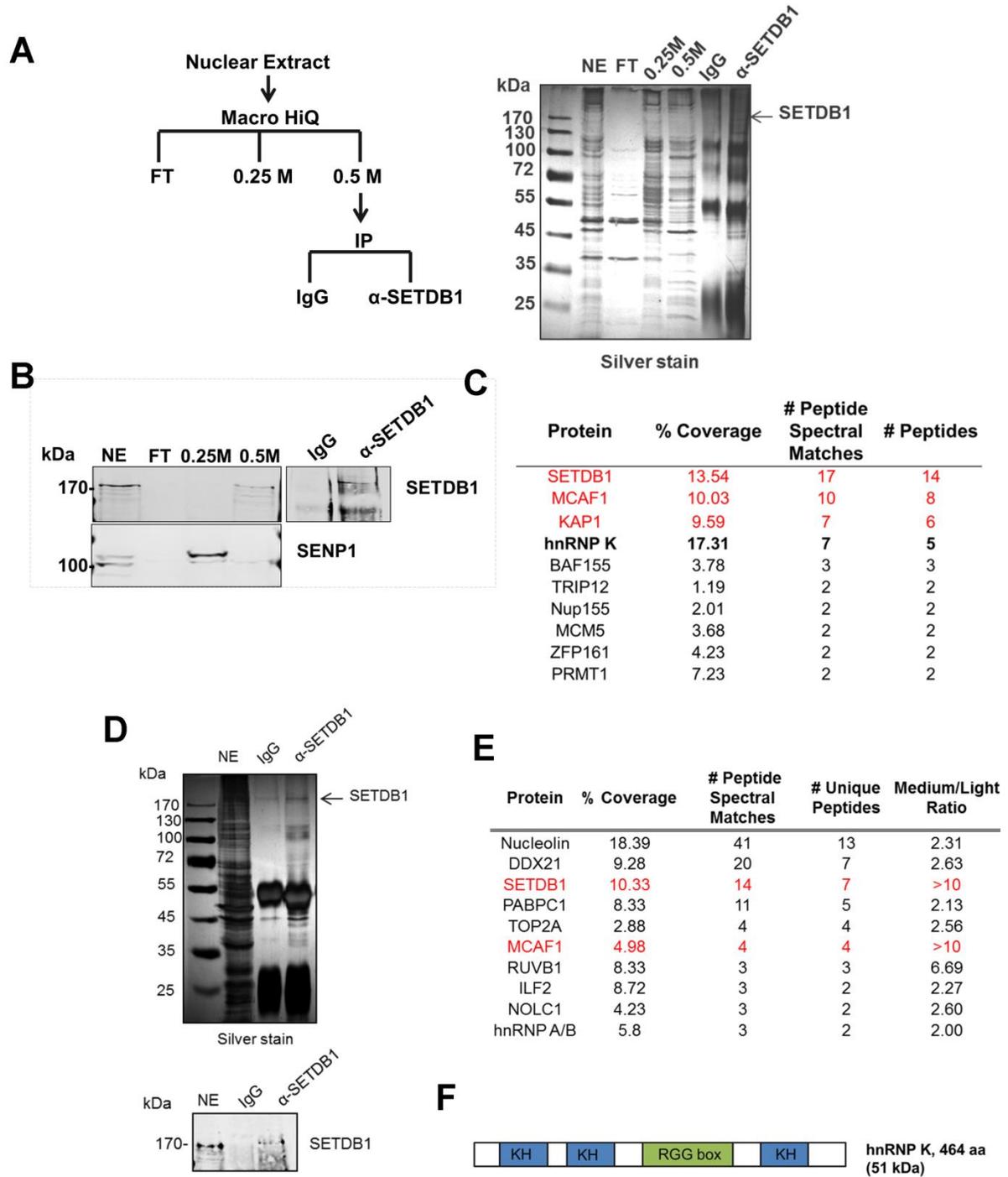


Figure 1. Proteomic analysis of SETDB1 complexes in mESCs. (A) Left panel: Purification scheme to enrich for SUMO-dependent SETDB1-interacting proteins. FT, flow-through, 0.25 M and 0.5 M KCl are elution fractions. IgG is a negative control IP. Details are given in the materials and methods section. Right panel: Silver stained gel showing the protein content of the indicated fractions (B) Western blot analysis showing the presence of SETDB1 and SENP1 during the purification. (C) Table of top 10 nuclear proteins in the SETDB1 IP with enrichment values >2 relative to IgG detected by MS. Shown in red are SETDB1 and known binding partners MCAF1 and KAP1. (D) Silver stained gel of a SETDB1 IP directly from mESC nuclear extract without prior anionic column fractionation. Western blot below confirms the presence of SETDB1 in the IP material. (E) Table of top 10 proteins in the SETDB1 IP with enrichment values >2 relative to IgG detected by MS. (F) Domain structure of hnRNP K showing three DNA/RNA binding KH domains and a internal protein-protein interaction domain, the RGG box.

3.3 hnRNP K is associated with the SETDB1/KAP1 complex in mESCs

I next performed experiments to corroborate the interactions between hnRNP K and SETDB1 and further determine whether hnRNP K also interacts with KAP1 using a combination of co-IP, immunostaining and co-sedimentation assays. Both KAP1 and hnRNP K were detected in 3XFLAG-tagged SETDB1 complexes immunopurified from mESCs in the presence of the general alkylating agent and cysteine protease inhibitor NEM, which blocks SENP activity²⁵² (**Figure 2A**). Moreover, using a specific antibody raised against an internal epitope of SETDB1,²⁸³ both KAP1 and hnRNP K co-precipitated with SETDB1 from mESC nuclear extract only in the presence of NEM (**Figure 2B**). Notably, hnRNP K exists as two different sized isoforms in human cell lines²⁸⁴ and mESCs and the faster migrating band is known to be generated by alternative splicing that removes a.a. 111-134 (Uniprot accession P61978-3). However, the full-length hnRNP K was consistent enriched in the SETDB1 IP (**Figures 2A and 2B**). The association of hnRNP K and SETDB1 was also apparent by immunostaining, which revealed that hnRNP K and SETDB1 colocalize in the nucleus and to

a lesser extent the cytoplasm of mESCs upon a short incubation with NEM.²⁴⁵ Reciprocally, SETDB1 co-precipitated with both KAP1 and hnRNP K in the presence of NEM and hnRNP K and KAP1 also co-precipitated with each other (**Figure 2C**), indicating that these proteins are present in a single complex. Notably, the IP of KAP1 was clearly more efficient in the presence of NEM (**Figure 2C**), revealing that SENP inhibition may stabilize KAP1 oligomeric state, as KAP1 is known to form oligomers.²⁸⁵ In addition, although hnRNP K binds to both DNA and RNA sequences,^{282,286} the interaction between SETDB1 and hnRNP K was not perturbed in the presence of RNase A and DNase I (**Figure 2D**) suggesting that it is not dependent upon nucleic acid.

To characterize the complexes formed by SETDB1, KAP1 and hnRNP K in mESCs, sucrose gradient ultracentrifugation was performed on purified KAP1 and hnRNP K proteins and compared with sedimentation profiles of the endogenous proteins in nuclear extracts in the presence and absence of NEM (**Figures 2E and 2F**). Purified GST-tagged KAP1 and hnRNP K proteins sedimented around a relative mobility of ~4.3S (**Figure 2E**) which likely represents monomeric species. In the absence of NEM, SETDB1, KAP1 and hnRNP K all co-sedimented at a relative mobility around ~11.3S in fractions 5-9 (**Figure 2F**) which could be consistent with them remaining uncomplexed or forming a small complex. In the presence of SENP inhibitor NEM, the stability of SETDB1/KAP1/hnRNP K complexes was increased and the three proteins co-sedimented at a higher density (larger size) at a relative mobility of ~19.4S in fractions 9-11, as compared with their profiles in the absence of NEM and the profiles of purified GST-KAP1 and GST-hnRNP K (**Figures 2E and 2F**). KAP1 showed the most dramatic change in mobility in NEM and sedimented across a broad range at high

densities in fractions 11-19 (**Figure 2F**), consistent with the better efficiency of KAP1 IP in the presence of NEM (**Figure 2C**). Interestingly, in both the presence and absence of NEM, the majority of hnRNP K sedimented at a relative mobility of ~4.3S-11.3S (**Figure 2E**), which is much smaller than the previously determined size of the heterogeneous nuclear RNA-protein spliceosomal complexes at ~40S,¹³⁷ suggesting that only a minority of total hnRNP K exists in these large complexes in mESCs. Together these results support the conclusion that hnRNP K is associated with the SETDB1/KAP1 complex under SUMOylation-enriched (or SENP-protected) conditions in mESCs.

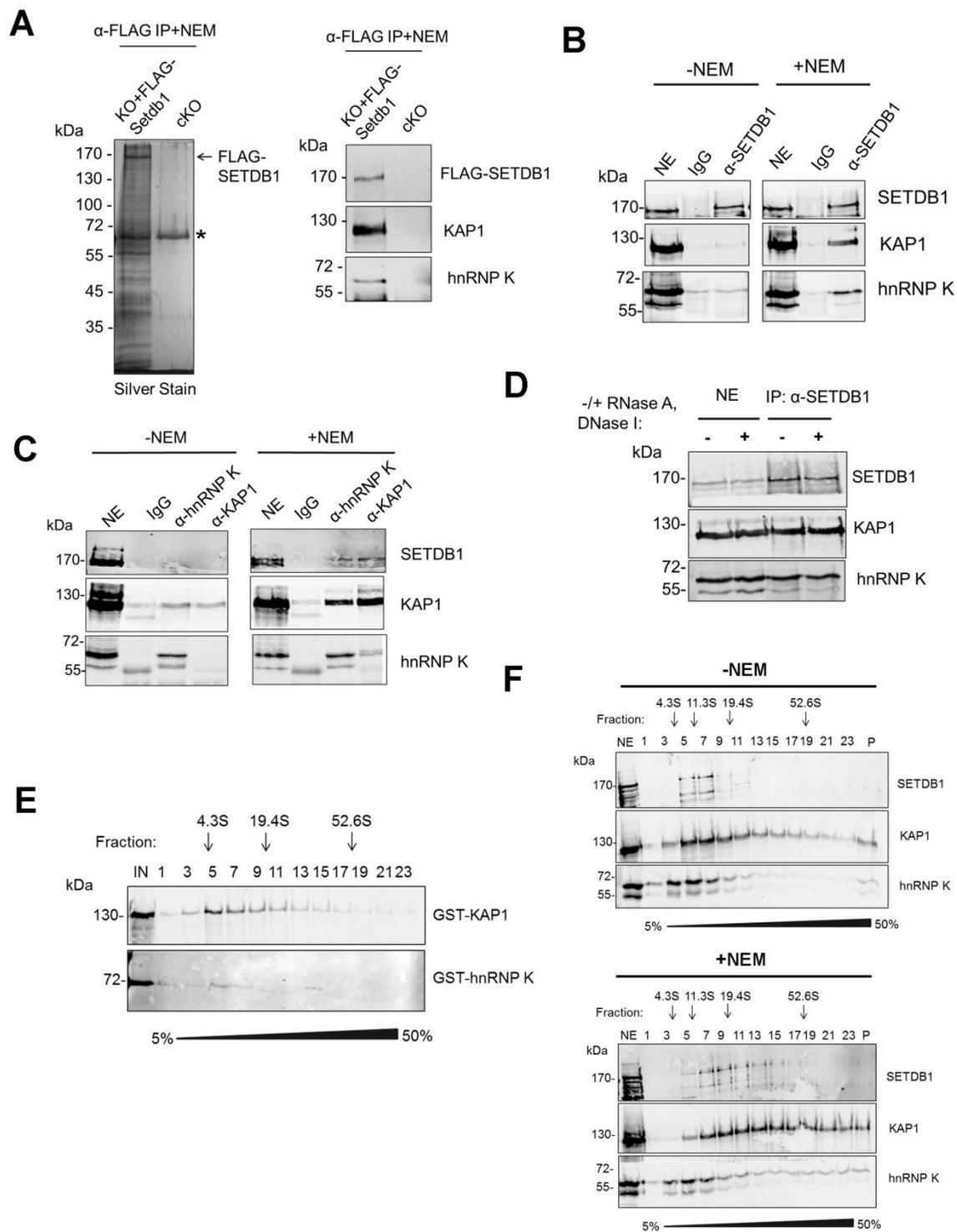


Figure 2. hnRNP K interacts with the SETDB1/KAP1 complex in mESCs. (A) Silver stain and western blot of an IP of FLAG-tagged SETDB1 from nuclear extracts of a *Setdb1* CKO mESC line deleted for endogenous *Setdb1* and expressing a 3XFLAG-*Setdb1* transgene (KO+FLAG-SETDB1). The uninduced *Setdb1* CKO line was a negative control. IP was performed in the presence of SENP inhibitor NEM. KAP1 and hnRNP K were detected by western blot in the FLAG-SETDB1 IP. Asterisk marks a non-specific IP band in the silver stained gel. (B) Co-IP assay and western blot of endogenous SETDB1 with KAP1 and hnRNP K from mESC nuclear extract in the presence or absence of NEM, using an independent antibody as compared with data shown in Fig 1. IgG was a negative control IP. Note: hnRNP K is expressed as two isoforms, the shorter isoform was termed hnRNP J (~60 kDa by SDS-PAGE) and lacks an ~30 a.a. internal region.²⁸⁴ SETDB1 interacts with the full-length hnRNP K. (C) Reciprocal co-IP assays of KAP1 and hnRNP K and western for SETDB1 in the presence or absence of NEM. (D) Co-IP assay and western blot of KAP1 and hnRNP K with SETDB1 in the presence of NEM on mESC nuclear extract with or without DNase I and RNase A. (E) Sucrose gradient ultracentrifugation of recombinant GST-tagged KAP1 and hnRNP K and western blot showing their sedimentation profiles. Sedimentation of purified standards are shown above. (F) Sucrose gradient ultracentrifugation and western blot as in (E) except on endogenous SETDB1, KAP1 and hnRNP K in mESC nuclear extracts in the absence or presence of NEM.

3.4 hnRNP K directly interacts with KAP1

To determine whether hnRNP K directly binds to SETDB1, GST pulldown assays were performed with recombinant SETDB1 or Ubc9 as a positive control protein for hnRNP K.^{242,287} Although KAP1 is SUMOylated in the SETDB1 complex under standard tissue culture conditions,²⁵⁸ hnRNP K was also reported to be SUMOylated but only following DNA damage.^{242,287} Indeed, whereas SETDB1 complexes from mESCs contained both SUMOylated and unmodified KAP1, I found no evidence of SUMOylated hnRNP K (**Figure 3A**) and thus used unmodified hnRNP K in subsequent pulldown assays. In contrast with Ubc9, which bound to SUMO2 and hnRNP K, SETDB1 bound to SUMO2 but not hnRNP K (**Figure 3B**). In addition, no interaction was detected between FLAG-tagged SETDB1 and

T7-tagged hnRNP K upon co-expression and FLAG IP from 293T cells (**Figure 3C**).

Together these data indicated that hnRNP K does not directly interact with SETDB1.

To determine whether hnRNP K directly binds SUMOylated and/or unmodified KAP1, *in vitro* SUMO1-conjugated GST-tagged KAP1, GST-p53 as a positive control binding partner of hnRNP K,²⁸⁷ or a GST-tagged fragment of RanGAP1 as a model SUMO1 substrate were prepared and used as baits in pulldown assays with recombinant SETDB1 or hnRNP K preys. Using purified SUMOylation cascade components, RanGAP1 was efficiently mono-SUMOylated at K526²⁸⁸ and KAP1 was mono-, di-, tri- and tetra-SUMOylated (**Figure 3D**) at its major SUMO acceptor lysines including K554, K676, K779 and K804.^{258,259} p53 was mono-SUMOylated which occurs at K386²⁸⁹ although this was less efficient in the absence of a SUMO E3 ligase (**Figure 3D**). While SETDB1 directly bound to p53 independently of SUMOylation, it bound to KAP1 in a SUMO1-dependent manner (**Figure 3E**), consistent with previous observations.²⁵⁸ HnRNP K binding to p53 was enhanced by SUMOylation but surprisingly, its binding to KAP1 was decreased upon SUMOylation (**Figure 3F**).

KAP1 harbours several functional domains that participate in protein-protein interactions, including an N-terminal RING-B-box-coiled-coil (RBCC) domain, which mediates binding to KRAB-ZFPs and other proteins,^{213,290,291} a proline-x-valine-x-leucine (PxVxL) motif, which binds to HP1 proteins²¹⁶ and a C-terminal PHD finger-bromodomain that binds to Ubc9 and chromatin-modifying factors, including SETDB1 and CHD3 upon its SUMOylation.^{214,215,258} While hnRNP K bound to wt full-length KAP1, it failed to bind to the deletion fragments containing only the PxVxL motif or only the PHD finger-bromodomain

(**Figure 3G**), indicating that hnRNP K binding requires the N-terminal RBCC domain. Taken together, these observations indicate that hnRNP K and SETDB1 indirectly interact with each other via their binding to unmodified or SUMOylated KAP1 subunits. Consistent with this model, SETDB1 complexes in mESCs contain both unmodified and SUMOylated KAP1 (**Figure 3A**), despite SETDB1 exhibiting binding affinity for only SUMOylated KAP1 (**Figure 3C**). In addition, endogenous KAP1 also co-precipitated with hnRNP K from 293T cell extracts (**Figure 3H**), indicating that this interaction is not limited to mESCs.

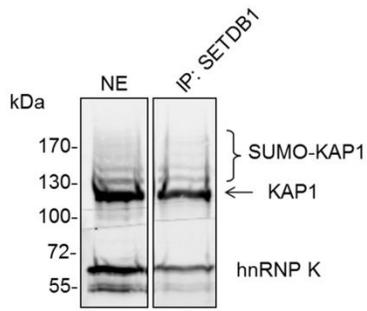
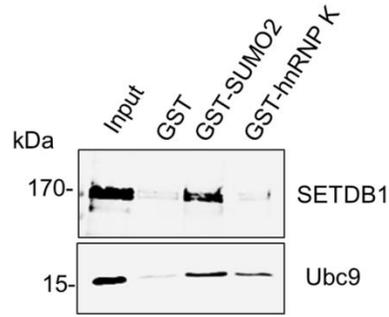
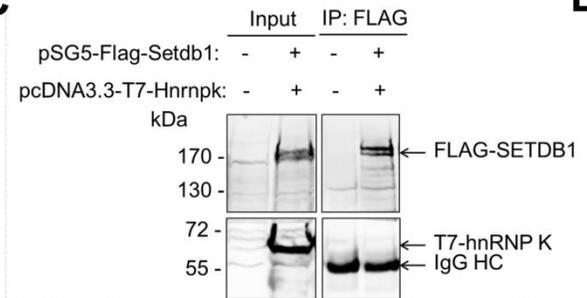
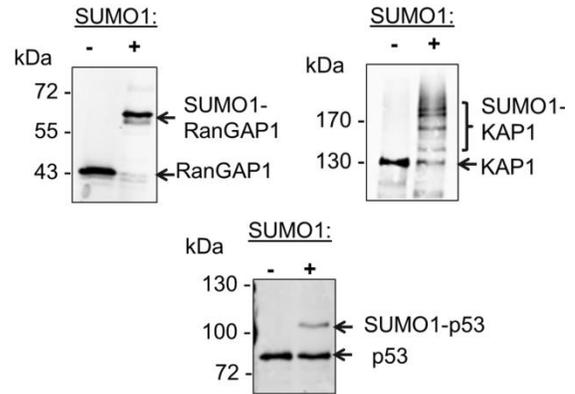
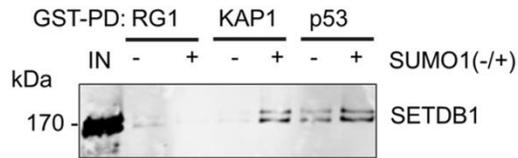
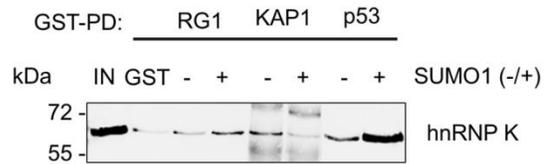
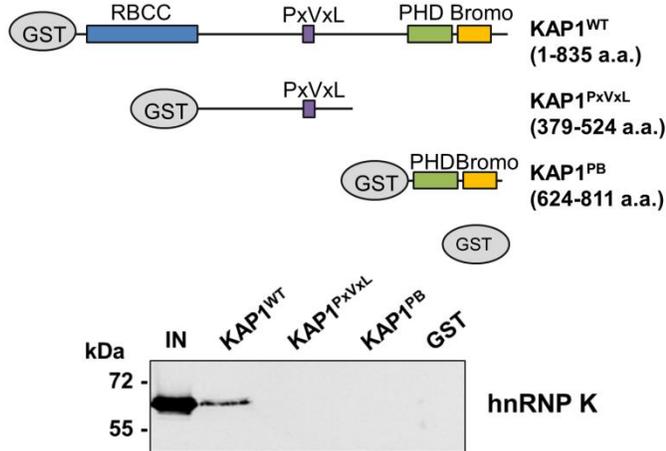
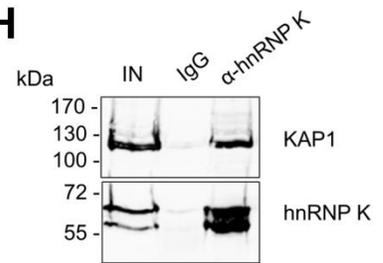
A**B****C****D****E****F****G****H**

Figure 3. hnRNP K directly interacts with KAP1. Western blot detection of KAP1 and hnRNP K in a SETDB1 IP from mESC nuclear extract in the presence of 20 mM NEM to visualize protein SUMOylation. (B) GST pulldown assays using GST-SUMO2, GST-hnRNP K and GST alone as baits and recombinant SETDB1 or Ubc9 as prey proteins. (C) Co-IP assay of FLAG-tagged SETDB1 and T7-tagged hnRNP K upon co-expression in 293T cells. (D) Western blot validation of *in vitro* SUMOylation reactions for GST-RanGAP1 (C-terminal fragment), GST-KAP1 and GST-p53. GST antibodies were used to detect RanGAP1 and p53 SUMOylation, while KAP1 antibodies were used to detect KAP1 SUMOylation. (E) GST pulldown (PD) assays using *in vitro* SUMOylated bait proteins indicated with recombinant SETDB1 prey protein. Negative control PDs were performed by omitting SUMO1. (F) GST PD assays, as in (E) except using 6XHis-tagged recombinant hnRNP K prey protein. GST was also included as an additional negative control. (G) GST PD assays with GST-tagged KAP1 wt or mutants indicated and recombinant hnRNP K prey protein. (H) Co-IP assay of KAP1 with hnRNP K from 293T cell protein extracts. Mouse IgG IP was a negative control.

3.5 Depletion of hnRNP K compromises mESC self-renewal

Having established that hnRNP K interacts with the SETDB1/KAP1 complex, likely via direct binding to KAP1, I next investigated whether loss of hnRNP K compromises SETDB1-dependent proviral silencing. During the course of this work, an *Hnrnpk* homozygous genetrapped mESC line (*Hnrnpk*^{Gt/Gt}) became available based on the work of Horie *et al.*²⁴¹ Using a random retroviral mutagenesis approach in a mESC line deficient in the DNA repair protein Blm, a mESC line was recovered in which a retroviral genetrapped vector integrated into intron 1 of *Hnrnpk* (**Figure 4A**). This mutation was predicted to disrupt *Hnrnpk* mRNA leading to null mutations at both alleles. However, using qRT-PCR and western blot analysis, I found that the levels of full-length hnRNP K protein and mRNA were not affected in the mutant line as compared to wt mESCs (**Figures 4B and data not shown**). The presence of the genetrapped was verified by PCR (data not shown). Therefore it was concluded that the *Hnrnpk* gene likely has one or more downstream TSSs, which are capable

of compensating for loss of expression from the genetrapped alleles in mESCs. Since *Hnrnpk* KO mESCs were not available for my genetic analysis, I instead utilized RNAi to deplete hnRNP K.

Using siRNA-mediated knockdown (KD), hnRNP K was efficiently depleted at the protein level by 24 h post-transfection (**Figure 4C**). Notably, KD of hnRNP K in mESCs significantly reduced their proliferation by 72 h post-transfection (**Figure 4D**). However, there was no gross effect on cell cycle distribution at this time-point and only minimal effects on expression of the pluripotency marker SSEA1 (**Figures 4E and 4F**). Furthermore, reduced proliferation was not associated with induction of apoptosis, as determined by Annexin V staining (**Figure 4F**). Thus hnRNP K KD does not result in overt differentiation or apoptosis at this time-point. These results are consistent with the recent report by Lin *et al.*¹⁴¹ who showed that shRNA-mediated depletion of hnRNP K leads to eventual apoptosis and loss of pluripotency in mESCs. To mitigate the impacts of cell death and differentiation on proviral silencing phenotypes resulting from the depletion of hnRNP K, cells were generally harvested at 72-96 h or earlier.

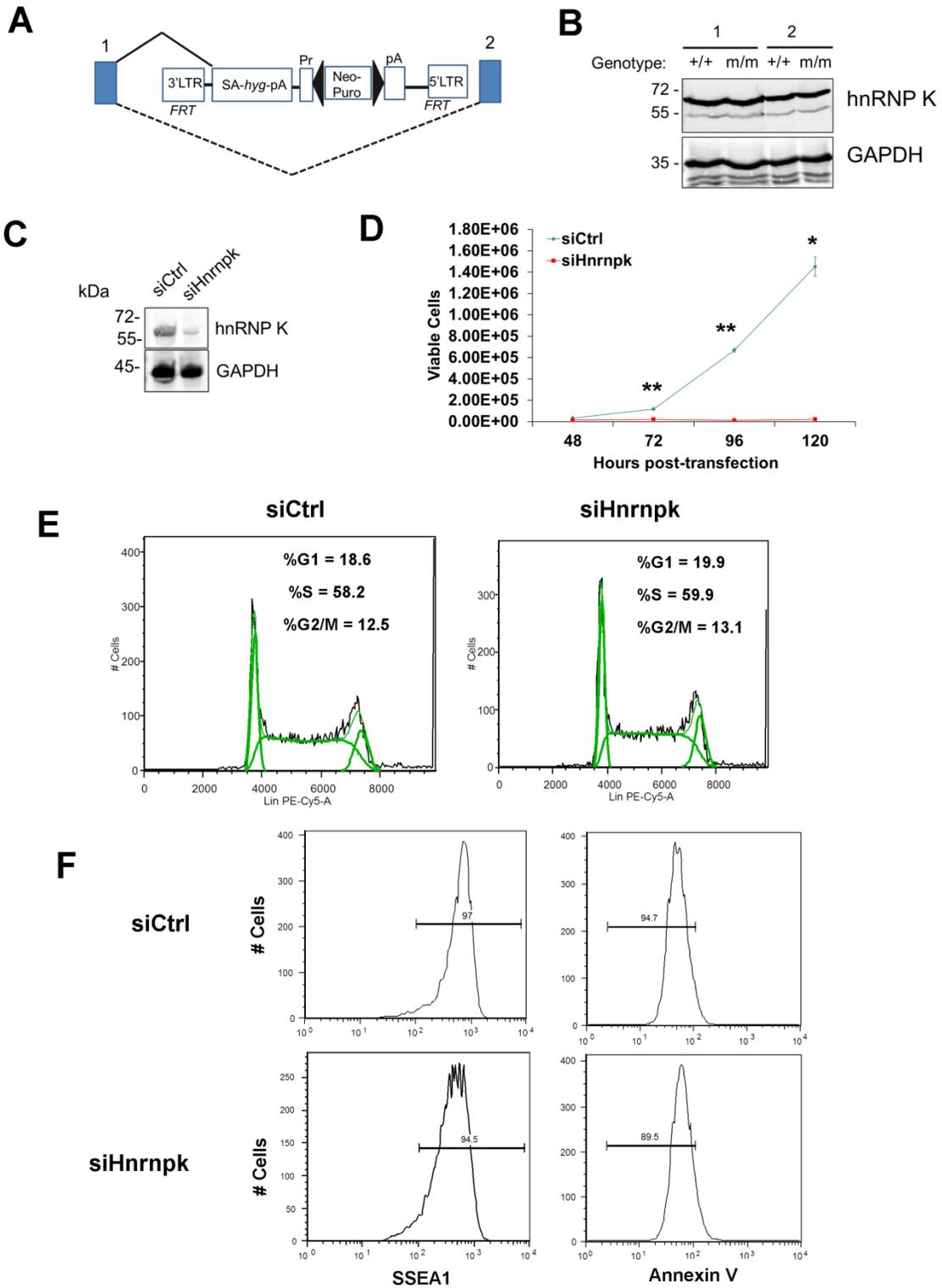


Figure 4. Depletion of hnRNP K compromises mESC self-renewal. (A) Schematic of the genetrapped *Hnrnpk* locus generated by Horie *et al.*²⁴¹ zoomed in on exons 1-2 where the retroviral vector genetrapp is in intron 1. A splice acceptor (SA) cassette splices from the annotated exon 1 precluding the production of full-length transcripts from this exon. (B) Western blot analysis of hnRNP K in the parental wt (+/+) and *Hnrnpk* homozygous genetrapp mutant (m/m) mESC line. GAPDH was a loading control. (C) Western blot of hnRNP K in TT2 mESCs transfected with control or *Hnrnpk* siRNAs at 24 h post-transfection. GAPDH was a loading control. (D) Growth curve assay for the cells from (C). *p < 0.001, **p < 0.0001 two-tailed T-test. (E) Cell cycle analysis using PI staining for control and *Hnrnpk* siRNA-transfected cells at 72 h post-transfection. (F) SSEA1 and Annexin V flow cytometry analysis of mESCs transfected with control or *Hnrnpk* siRNAs at 72 h.

3.6 hnRNP K is required for maintenance of SETDB1-dependent proviral reporter and ERV silencing

To determine the effect of hnRNP K depletion on proviral silencing, I used previously established proviral GFP reporter mESC lines, including the MSCV-PBS^{Gln} GFP line (hereafter referred to as MSCV-GFP)⁸⁹ and the HA36 mESC line, which harbours a silent IAP LTR-PBS-5'UTR region driving *GFP*, which is integrated at a defined genomic locus²³⁷ (hereafter referred to as IAP-GFP, **Figure 5A**). These mESC lines allowed for the interrogation of the role of hnRNP K in proviral silencing at randomly integrated sites (as with the MSCV-GFP) and at a defined locus (as for the IAP-GFP). Importantly, in both lines, proviral silencing is dependent upon H3K9me3 deposited by the SETDB1/KAP1 complex.^{89,237} Furthermore, the silencing of the MSCV-GFP construct is not perturbed in the DNMT TKO mESC line and deletion of *Setdbl* leads to loss of DNA methylation at the LTR,⁸⁹ indicating that its silencing is totally dependent on the SETDB1/KAP1 complex.

Transfection of siRNAs specific for *Setdbl* or *Hnrnpk* effectively reduced expression of the relevant mRNAs to ~10-25% of the control siRNA-transfected cells (**Figure 5B**). While only

~2-3% of SSEA1⁺ cells were also GFP⁺ in the untransfected (MSCV and IAP) and siRNA transfected controls, KD of *Setdb1* de-repressed both reporters, resulting in ~37% and ~20% SSEA1⁺; GFP⁺ cells, respectively (**Figure 5C**). Strikingly, KD of *Hnrnpk* also consistently de-repressed both the MSCV and IAP reporters, resulting in an average of ~29% and ~20% SSEA1⁺; GFP⁺ cells, respectively (**Figure 5C**).

I next determined whether KD of hnRNP K disrupts silencing of ERVs. In contrast to the proviral reporter lines, KD of SETDB1 or hnRNP K in TT2 wt mESCs resulted in only modest de-repression (~2-fold) of class I and II ERVs by 72 h post-transfection (**Figures 5D and 5E**), with the exception of robust induction of MMERVK10C elements in SETDB1 KD cells, despite efficient depletion of the protein. Surprisingly, class III MERVL elements, which are repressed by KAP1 in a SETDB1-independent manner⁸⁶, were strongly induced in hnRNP K KD cells (**Figure 5E**).

A previous study revealed that DNA methylation, although not strictly necessary for ERV silencing in mESCs, still plays a role in transcriptional repression of ERVs, particularly of IAP elements in mESCs cultured in serum.²⁴⁶ Thus the presence of DNA methylation at ERVs upon transient depletion of SETDB1 or hnRNP K could be sufficient to maintain their silencing. To preclude the influence of DNA methylation, I knocked down *Setdb1* or *Hnrnpk* in the *Dnmt* TKO mESC line²³⁸ (**Figure 5F**), which is devoid of DNA methylation but maintain SETDB1 binding and H3K9me3 at ERVs⁸⁹ and thus solely rely on the SETDB1/H3K9me3 pathway for silencing of these elements. The absence of DNA methylation alone did not perturb silencing of MLV, MMERVK10C and MusD elements, but

yielded a ~6-fold upregulation of IAP elements (**Figure 5G**), consistent with the finding that IAP elements are modestly upregulated in *Dnmt* TKO cells.⁸⁹ In contrast, KD of *Setdb1* expression resulted in a substantial induction of ERVs in these cells (**Figure 5G**). These ERVs were also de-repressed upon *Hnrnpk* KD, with IAP elements showing an increase in expression of ~18-fold, ~3-fold greater than the control KD in the *Dnmt* TKO line (**Figure 5G**). Taken together, these results reveal that depletion of hnRNP K disrupts SETDB1/H3K9me3-mediated silencing of ERVs in mESCs.

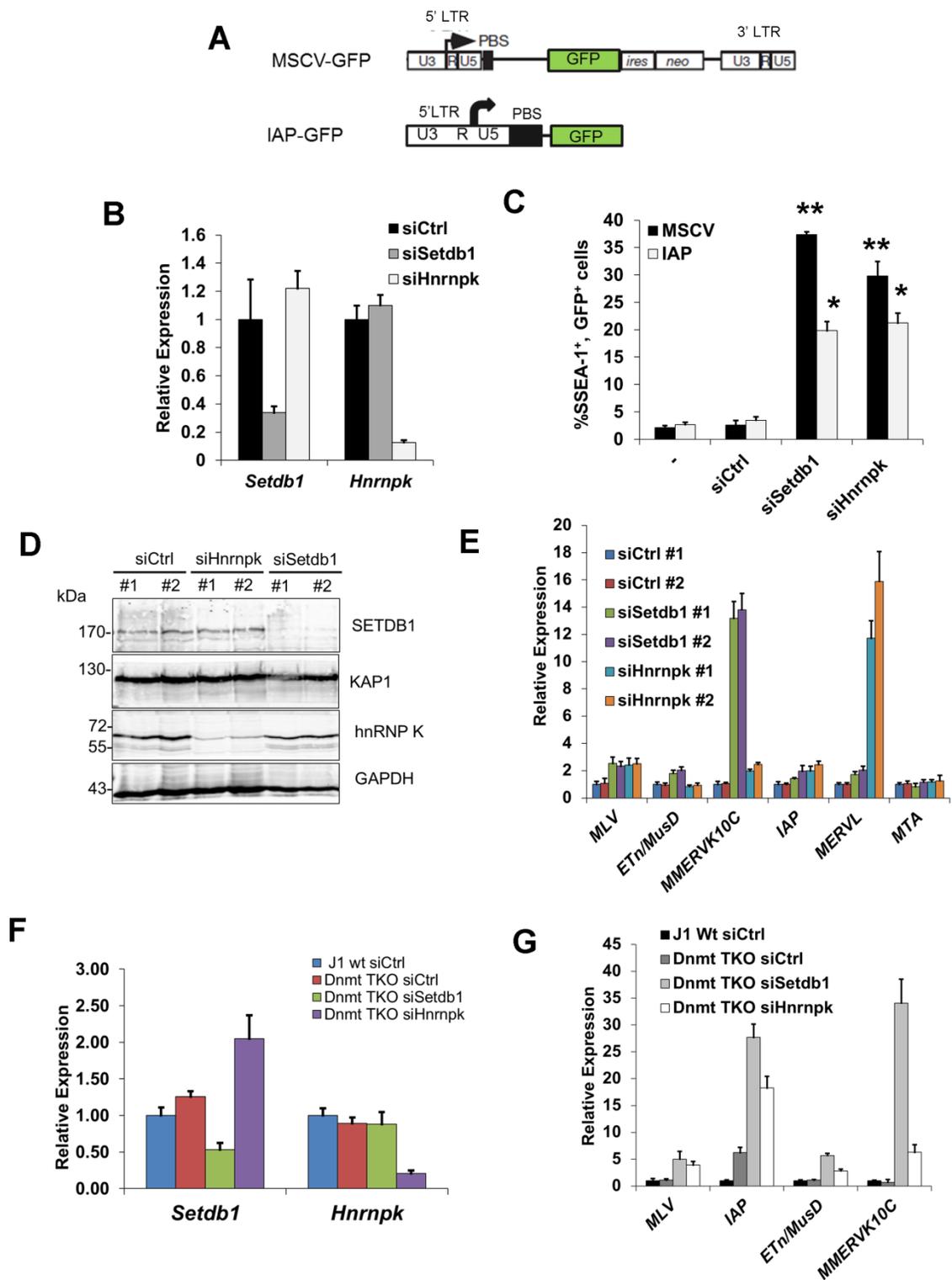


Figure 5. hnRNP K is required for SETDB1-dependent silencing of proviral reporters and ERVs. (A) Schematic of the MSCV-PBS^{Gln}-GFP reporter construct and the IAP LTR-PBS-GFP reporter construct. (B) qRT-PCR validation of *Setdb1* and *Hnrnpk* KDs at 24 h post-transfection. (C) Flow cytometry analysis of SSEA1⁺;GFP⁺ cells upon transfection with control, *Setdb1* or *Hnrnpk* siRNAs at 72 h post-transfection, relative to the untransfected parental line negative control(-). Data are mean of n = 3 biological replicates, error bars are s.d. *p < 0.01, **p < 0.001, two-tailed T-test relative to the control siRNA lines. (D) Western blot analysis of SETDB1, KAP1 and hnRNP K in biological replicates at 24 h post-transfection with the indicated siRNAs. GAPDH was a loading control. (E) qRT-PCR analysis of ERV expression in the indicated KD cells, relative to the control siRNA, at 72 h post-transfection. (F) qRT-PCR validation of *Setdb1* and *Hnrnpk* KDs in the *Dnmt* TKO mESC line at 24 h post-transfection. (G) qRT-PCR analysis of ERV expression in J1 wt or *Dnmt* TKO cells at 96 h post-transfection with the indicated siRNAs. All qRT-PCR data are the mean of n = 3 technical replicates from a representative experiment, error bars are s.d.

3.7 hnRNP K represses a cohort of SETDB1-targeted male germline-specific genes

To investigate whether depletion of hnRNP K disrupts SETDB1-dependent repression of genes, I performed mRNA-seq from two biological replicates of TT2 mESCs transfected with control or *Hnrnpk* siRNA (**Figure 6A**). A total of 290 genes were consistently misregulated upon hnRNP K KD, 264 genes were upregulated ≥ 2 -fold in both KD lines while only 26 were downregulated by $\geq 50\%$ (**Figure 6A**).²⁴⁵ Gene ontology (GO) analysis revealed that the upregulated genes were enriched for “apoptosis”²⁴⁵ indicating that although these KD cells do not show high levels of Annexin V staining at this time-point, their progressive proliferation block (**Figure 3D**) may coincide with induction of the apoptotic pathway. Although hnRNP K regulates the expression of pro-apoptotic genes *Bcl-Xs* and *Bik* under certain conditions,²⁹² these genes were not upregulated in hnRNP K KD mESCs. Nevertheless, *Btg2*, *Anxa8*, *Perp*, *Trp73*, *Cdkn1a* and *Casp14* were among the 16 apoptosis-associated genes identified by GO analysis. In addition there was an enrichment of genes involved in “lung and respiratory system development” including the primitive endoderm and mesoderm lineage transcription factors *Gata6*, *Gata3*, *Tbx3*, *Tbx20*, *Foxa1*, *Nkx2-9* and

Nkx2-2.²⁴⁵ Previous ChIP-seq data indicates that these transcription factor genes harbour the bivalent chromatin state of H3K4me3 and H3K27me3^{9,10,93} and are subject to PRC2-mediated silencing.⁵¹ The de-repression of *Gata6* and *Gata3*, which were upregulated ~15-fold and ~10-fold in hnRNP K KD cells, respectively as validated by qRT-PCR (**Figure 6B**), indicates that hnRNP K KD could eventually lead to a loss of pluripotency, since the overexpression of these transcription factors is sufficient to drive endoderm lineage differentiation.^{293,294} RNA-seq analysis of ERVs in the TT2 hnRNP K KD cells generally confirmed our qRT-PCR analysis from the same cells (**Figure 5B**) in that class I and II elements were only modestly de-repressed (≤ 2 -fold) while MERVL elements were strongly de-repressed (≥ 14 -fold).²⁴⁵

Comparison of the list of genes upregulated in hnRNP K KD mESCs to the list of upregulated genes in *Setdb1* KO mESCs²⁴⁶ revealed 54 genes in common.²⁴⁵ A cohort of 33 male germline lineage genes were previously determined to be repressed by SETDB1-dependent H3K9me3 and DNA methylation.²⁴⁶ Notably, many of these direct SETDB1 target genes were consistently upregulated >2-fold in hnRNP K KD cells (15 of these genes are shown in **Table 2**). In addition, of the 134 SETDB1-bound genes that are upregulated in *Setdb1* KO mESCs,²⁴⁶ 30 were consistently de-repressed in hnRNP K KD cells (**Figure 6C**). Quantitative RT-PCR analysis of a subset of these genes, including the male germline-specific genes *Dazl*, *Fkbp6*, *Mael* and *Taf7l* confirmed that they are indeed upregulated in hnRNP K KD cells (**Figure 6B**). Furthermore, hnRNP K was required for H3K9me3 at these genes, since the levels of H3K9me3 at the promoters of these genes were reduced in hnRNP K KD cells to a similar extent as in SETDB1 KD cells (**Figure 6D**).

A comparison of the genes upregulated in hnRNP K KD and *Kap1* KO cells⁹⁰ also revealed a significant overlap (**Figure 6E**) and included other lineage-restricted genes such as *Gata6*, *Arg2* and *Dkk1*, which were recently shown to be repressed by the cooperative actions of KAP1 and PRC1.²⁹⁵ There were 33 genes commonly de-repressed in *Setdb1* KO, *Kap1* KO and hnRNP K KD mESCs, most of which were expressed in a lineage-dependent fashion, including the imprinted gene *Igf2* and liver-specific gene *Cml2*.²⁴⁵ The promoter of *Cml2* lies immediately downstream of an intact ETn family retroelement that is bound by SETDB1 and marked by SETDB1-dependent H3K9me3, which spreads into the *Cml2* promoter (**Figure 6F**) indicating that this gene is silenced by the spreading of H3K9me3 from the intact ERV. In conclusion, these results support a role for hnRNP K in transcriptional repression of genes regulated by SETDB1 and KAP1 as well as PRC2/PRC1, the latter via an undefined pathway.

Table 2. RNA-seq fold-changes of a subset of SETDB1-targeted tissue-specific genes upregulated in common in *Setdb1* KO and *Hnrnpk* KD cells.[#]

Gene	Average Fold-Change hnRNP K KD	Fold-Change <i>Setdb1</i> KO	Lineage/Tissue expression (BioGPS)
<i>Fkbp6</i>	2.4	11.7	Testis
<i>Dpep3</i>	2.15	25.8	Testis
<i>Cox7b2</i>	4.1	23.6	Testis, neuronal
<i>Wfdc15a</i>	8	11.3	Testis
<i>Slc25a31</i>	7.8	12.9	Testis
<i>Taf7l</i>	9.4	8.3	Testis, Placenta
<i>Ankrd45</i>	4.6	8	Testis, Neuronal
<i>Mael</i>	3.2	5.8	Testis
<i>1190002H23Rik</i>	3.1	3.7	Adipose, Connective,
<i>Tex101</i>	5	84.4	Testis
<i>Gtsf1</i>	2.2	5.3	Testis
<i>Dazl</i>	4.3	6.5	Testis
<i>Hormad1</i>	2.3	6.1	Testis
<i>Stk31</i>	2.1	5.5	Testis, placenta
<i>1700007K13Rik</i>	7.3	5	ESCs, Ovary, Testis

- Genes shown in yellow were analyzed by qRT-PCR and NChIP for H3K9me3, as shown in Figures 6B and 6D, respectively.

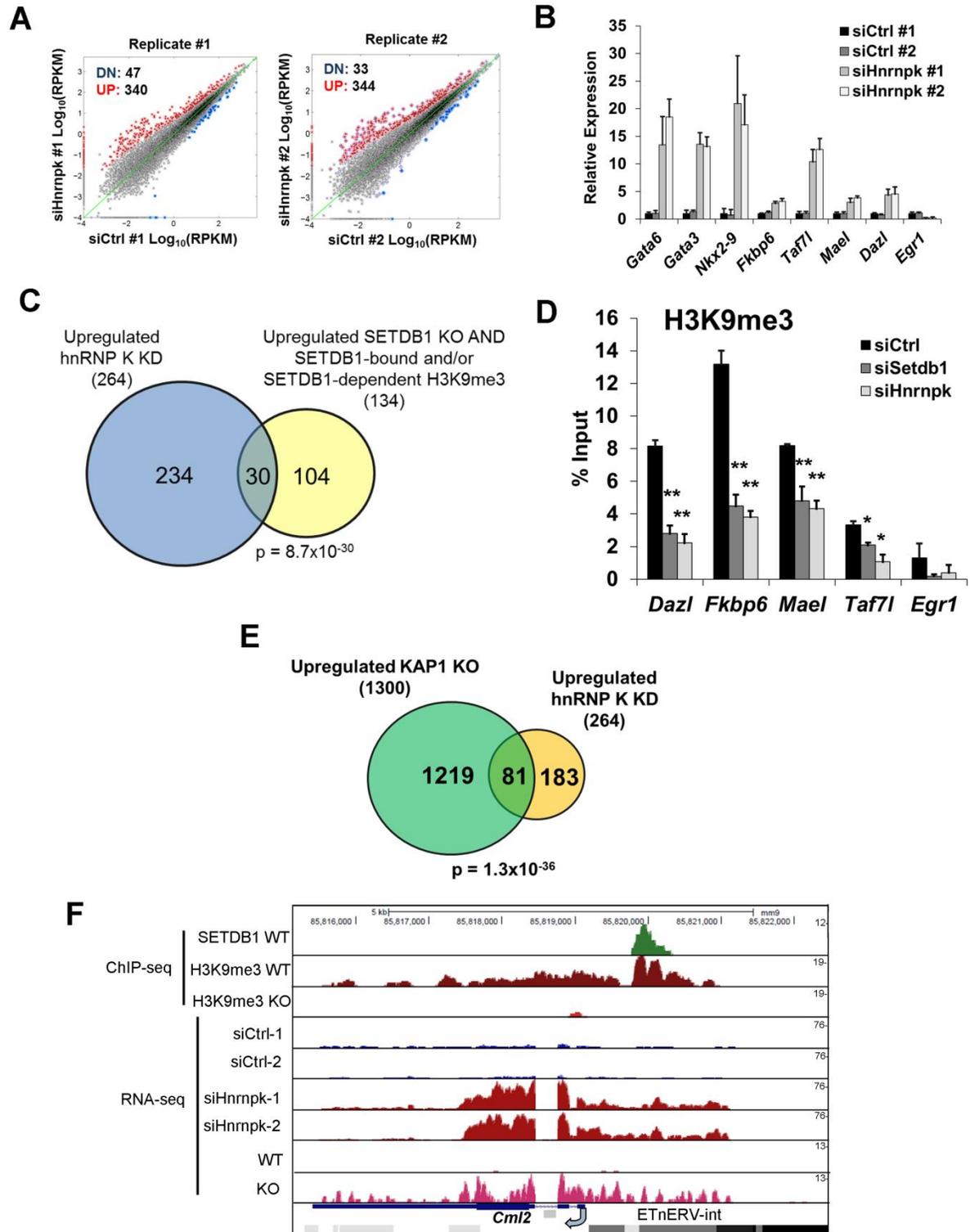


Figure 6. hnRNP K co-represses genes with SETDB1 and KAP1. (A) mRNA-seq from two biological replicates of control or *Hnrnpk* KD cells at 72 h post-transfection, showing the expression of $n = 22,138$ ENSEMBL annotated genes and those upregulated (≥ 2 -fold) or downregulated (≤ 0.5 -fold) relative to the control. (B) qRT-PCR validation of a panel of genes found to be misregulated by mRNA-seq: PRC2-repressed differentiation genes (*Gata6*, *Gata3*, *Nkx2-9*), SETDB1-targeted germline genes (*Dazl*, *Mael*, *Fkbp6*, *Taf7l*) or a downregulated gene (*Egr1*) in control and *Hnrnpk* KD cells at 72 h post-transfection. (C) Venn diagram of the overlap between genes commonly upregulated (≥ 2 -fold) in hnRNP K KD cells (264) and genes bound by SETDB1 or marked by SETDB1-dependent H3K9me3 and upregulated in *Setdb1* KO mESCs (134).²⁴⁶ $p = 8.7 \times 10^{-30}$, Fisher's exact test from $n = 22,138$ ENSEMBL annotated genes. (D) NChIP for H3K9me3 in control, *Setdb1* or *Hnrnpk* KD TT2 cells at 72 h post-transfection using primers to amplify the promoters of the indicated genes. The *Egr1* promoter is active in mESCs and was a negative control for the H3K9me3 mark. Data are mean enrichment relative to input chromatin from $n = 3$ technical replicates, error bars are s.d. * $p < 0.05$, ** $p < 0.005$ two-tailed T-test. (E) Venn diagram overlap between genes upregulated in *Kap1* KO cells (1300)⁹⁰ and *Hnrnpk* KD cells (264). $p = 1.3 \times 10^{-36}$, Fisher's exact test, on $n = 22,138$ ENSEMBL annotated genes. (F) Data track files for SETDB1 ChIP-seq,²⁸³ H3K9me3 ChIP-seq in WT and *Setdb1* KO cells and RNA-seq from WT and *Setdb1* KO cells²⁴⁶ alongside RNA-seq from control and *Hnrnpk* KD cells biological replicates. Data are shown over the *Cml2* locus showing an upstream ETn family retroelement marked with SETDB1-dependent H3K9me3 that spreads into the adjacent *Cml2* promoter. *Cml2* is upregulated in *Setdb1* KO, *Hnrnpk* KD and *Kap1* KO mESCs.

3.8 hnRNP K is bound at ERVs and its depletion leads to reduced levels of H3K9me3 at proviral chromatin

To determine whether hnRNP K is bound at ERVs, I performed ChIP analysis with a specific antibody.²⁹⁶ HnRNP K was enriched at the promoters of the SETDB1-bound, H3K9me3-marked germline genes *Fkbp6*, *Dazl*, *Mael* and *Taf7l*, with the highest level of enrichment detected at *Mael* (**Figure 7A**), indicating that these loci are direct targets of hnRNP K in mESCs. Relative to the germline gene promoters, the 5'LTRs of class I and II ERVs showed lower enrichment of hnRNP K, with ETn/MusD and MLV elements showing the highest and lowest levels, respectively (**Figure 7A**). Importantly, the signal at ERVs and the germline gene promoters was specific, since it was reduced upon hnRNP K KD. In contrast, there was no enrichment of hnRNP K at the *Egr1* promoter (**Figure 7A**), which is active in mESCs and

was shown to be bound by hnRNP K only upon serum stimulation in the HCT116 colon cancer cell line.^{296,297}

I next determined whether hnRNP K is required for SETDB1-dependent H3K9me3 deposition at ERVs. As shown previously,⁸⁹ SETDB1 KD resulted in depletion of H3K9me3 at MLV, IAP, MMERVK10C and ETn/MusD 5'LTRs and a similar depletion of H3K9me3 was observed in hnRNP K KD cells (**Figure 7B**). Furthermore, as early as 24 h after hnRNP K KD, there was a clear reduction of H3K9me3 at the MSCV 5'LTR-PBS and *Gfp* regions (**Figures 7C and 7D**). The levels of H4K20me3, a mark deposited by SUV420H1/2 enzymes in a SETDB1/H3K9me3-dependent manner⁸⁹, were also reduced at the MSCV provirus in hnRNP K-depleted cells (**Figure 7E**). Importantly, siRNA-mediated depletion of hnRNP K or SETDB1 did not affect global H3K9me2 or H3K9me3 levels at the 72 h time-point (**Figure 7F**), indicating that the effect of hnRNP K depletion on H3K9me3 at ERVs is not the result of a general reduction of H3K9me2/3. Thus hnRNP K is required for SETDB1-dependent H3K9me3 deposition at proviral chromatin.

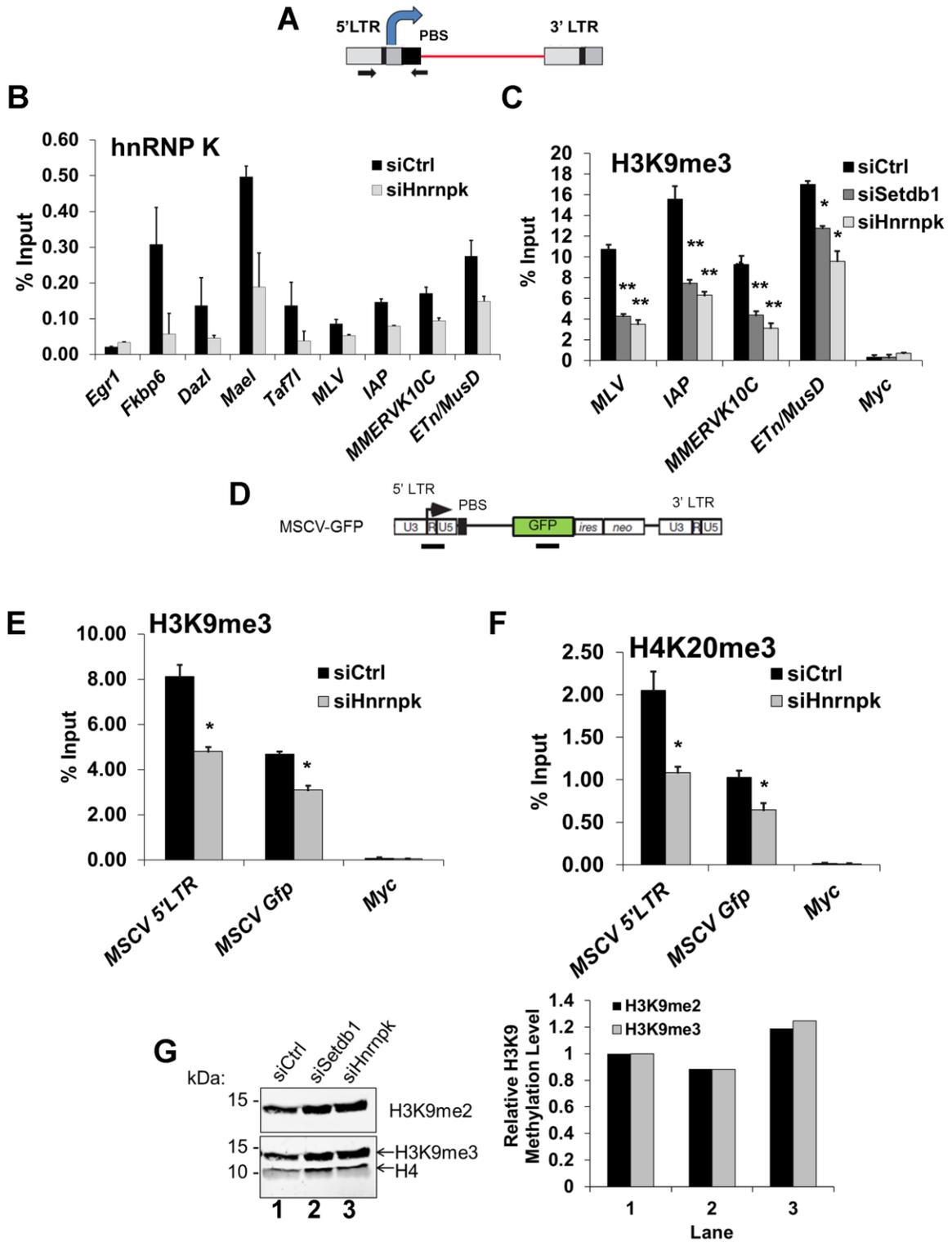


Figure 7. hnRNP K is required for H3K9me3 at proviral chromatin. (A) ERV schematic showing the position of the primers used for qPCR at the 5'LTR-PBS-*gag* region to amplify intact elements. (B) ChIP of hnRNP K at the indicated gene promoters and ERV 5'LTRs in TT2 control or *Hnrnpk* KD cells at 24 h post-transfection. The *Egr1* promoter was a negative control locus. Data are mean enrichment normalized into input chromatin from n = 3 technical replicates error bars are s.d. (C) NChIP for H3K9me3 at the indicated ERV 5'LTRs in control, *Setdb1* or *Hnrnpk* KD TT2 cells at 72 h post-transfection. Data are mean enrichment relative to input chromatin from n = 3 technical replicates, error bars are s.d. *p < 0.001, **p < 0.0001, two-tailed T-test. (D) Schematic of the MSCV-GFP reporter construct indicating positions of the LTR-PBS and *Gfp* amplicons for qPCR. (E and F) NChIP for H3K9me3 or H4K20me3 as in (C) except in unsorted (GFP⁺ and GFP⁻) MSCV-GFP reporter mESCs at 72 h post-transfection. (G) Western blot analysis of H3K9me2 and H3K9me3 on whole cell lysates of control, *Setdb1* and *Hnrnpk* KD cells at 72 h post-transfection. Histone H4 was a loading control. To the right of the blot is the quantification of H3K9me2 and H3K9me3, relative to total H4 for each.

3.9 hnRNP K is required for SETDB1 but not KAP1 recruitment to ERVs

The reduced levels of H3K9me3 at proviral chromatin upon hnRNP K depletion could be explained if hnRNP K promotes SETDB1 recruitment. To determine whether hnRNP K is required for SETDB1 recruitment, I conducted ChIP analysis of SETDB1 in cells depleted of hnRNP K (**Figures 8A and 8B**). A reduction of SETDB1 enrichment was apparent at all class I and II ERV LTRs in SETDB1 KD cells, confirming the specificity of the ChIP antibody (**Figure 8B**). Strikingly, KD of hnRNP K also reduced the level of SETDB1 enrichment at ERVs to an even greater extent (**Figure 8B**). SETDB1 enrichment was also reduced at the MSCV 5'LTR-PBS and *Gfp* internal region upon depletion of hnRNP K (**Figure 8C**) revealing a link between loss of H3K9me3, de-repression of the MSCV proviral reporter and reduced SETDB1 recruitment. Importantly, neither SETDB1 nor KAP1 protein levels were reduced in hnRNP K KD cells (**Figure 8A**) and SETDB1 still localizes to the nucleus in hnRNP K-depleted cells.²⁴⁵

KAP1 is the only factor known to be required for SETDB1 recruitment to proviral chromatin,^{89,90} and therefore I hypothesized that hnRNP K might promote SETDB1 recruitment indirectly by KAP1. CHIP analysis showed that while KAP1-depleted cells showed reduced levels of KAP1 enrichment at ERVs confirming antibody specificity, KD of hnRNP K did not affect KAP1 enrichment levels (**Figures 8C and 8D**). In contrast, KD of KAP1 substantially reduced hnRNP K enrichment at ERVs (**Figure 8E**). Taken together, these data reveal that hnRNP K is recruited in a KAP1-dependent manner and facilitates subsequent SETDB1 binding at proviral chromatin.

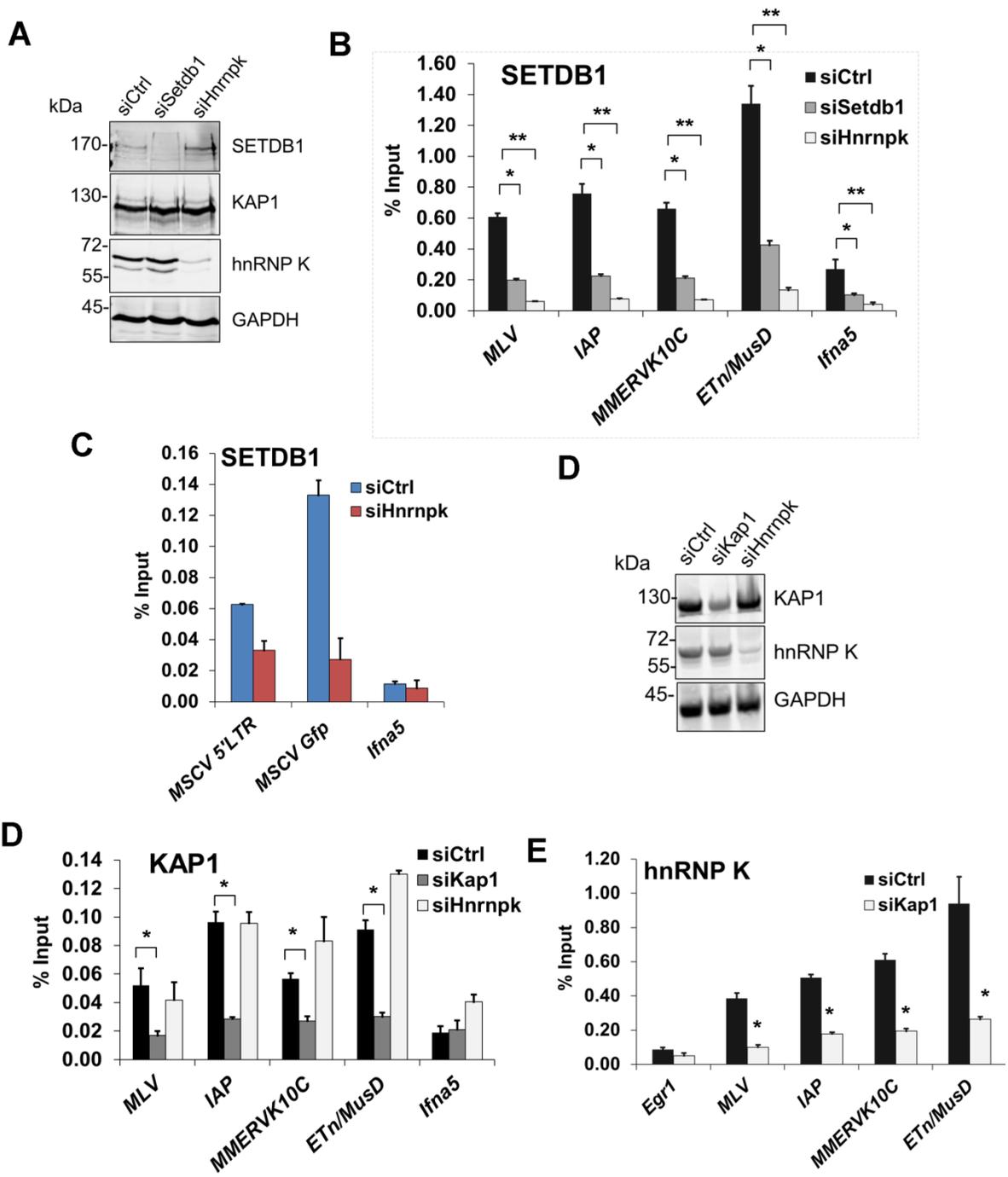


Figure 8. hnRNP K is required for SETDB1 recruitment to proviral chromatin and is recruited in a KAP1-dependent manner. (A) Western blot validation of *Setdb1* and *Hnrnpk* KDs on TT2 mESCs at 24 h post-transfection. GAPDH was a loading control. (B) ChIP of SETDB1 in control, *Setdb1* and *Hnrnpk* KD cells at 72 h post-transfection. (C) ChIP of SETDB1 as in (B) except on MSCV-GFP cells transfected with control or *Hnrnpk* siRNAs at 24 h post-transfection. (C) Western blot validation of *Kap1* and *Hnrnpk* KDs on TT2 cells at 24 h post-transfection. (D) ChIP of KAP1 in the cells from (C) except at 72 h post-transfection. (E) ChIP of hnRNP K in control and *Kap1* KD cells at 72 h post-transfection. For all ChIP-qPCR, data are mean enrichment relative to input chromatin from n = 3 technical replicates, error bars are s.d. The *Ifna5* promoter was a negative control locus for SETDB1 and KAP1, while the *Egr1* promoter was a negative control for hnRNP K. *p < 0.001, **p < 0.0001 two-tailed T-test.

3.10 Depletion of hnRNP K phenocopies inhibition of SUMO conjugation, which interferes with SETDB1 recruitment and proviral silencing

Previous studies have shown that KAP1 SUMOylation is necessary for recruitment of SETDB1 and H3K9 methylation to promote transcriptional silencing of heterologous promoters in transformed cell lines.^{258,259,298} To determine whether a functional SUMOylation pathway is also necessary for SETDB1 recruitment to ERVs in pluripotent stem cells, I used either anacardic acid to inhibit the SUMO E1 activating enzyme²⁹⁹ or siRNAs to deplete *Ubc9* (also called *Ube2i*) transcripts in the MSCV-GFP cell line and assayed for de-repression of the proviral LTR by flow cytometry (**Figure 9A**). In accord with the inhibitory effect of anacardic acid on the activity of SUMO E1 activating enzyme Aos1/Uba2 and histone H3 lysine acetyltransferases such as p300,³⁰⁰ this compound blocked both KAP1 SUMOylation and bulk histone H3 acetylation (**Figure 9B**). While, anacardic acid treatment did not affect bulk H3K9me3 (**Figure 9B**), it consistently de-repressed the proviral reporter in a dose-dependent manner, resulting in ~15% GFP⁺ cells at 100 μ M (**Figure 9C**).

Using siRNA KD, *Ubc9* mRNA was depleted to ~35% of the control (**Figure 9D**, inset graph). As *Ubc9* is essential for early embryogenesis,¹⁴⁶ I monitored changes in MSCV expression at 48 h post-transfection. KD of *Ubc9* expression consistently de-repressed the proviral reporter resulting in an average of 23% GFP⁺ cells (**Figure 9D**). Notably, ChIP analysis revealed that SUMO1 levels at the MSCV 5' LTR were dramatically reduced in *Ubc9*-depleted cells (**Figure 9E**) indicating that the loss of SUMOylation on proviral chromatin correlates with de-repression. Moreover, there was a reduction of SETDB1 enrichment at the MSCV provirus in *Ubc9* KD cells (**Figure 9F**), confirming that SUMOylation of chromatin proteins associated with ERVs enhances SETDB1 recruitment. Strikingly, SUMO1 levels were greatly reduced at the 5' LTRs of MLV, IAP, MMERVK10C and ETn/MusD elements by 24 h post-transfection of hnRNP K siRNAs (**Figure 9G**). Moreover, this effect also occurred at the MSCV provirus (**Figures 9H**) and coincided with the timeframe in which SETDB1 recruitment to proviral chromatin was compromised (**Figure 8B**). Although the loss of SUMOylation at ERV chromatin upon hnRNP K KD could be a consequence rather than a cause of reduced SETDB1 recruitment, KD of SETDB1, which was sufficient to de-repress the MSCV LTR, did not concomitantly attenuate SUMOylation on proviral chromatin.²⁴⁵ Taken together these results are consistent with the model that hnRNP K is necessary for SUMOylation of proteins such as KAP1 on ERV chromatin, which is required for SETDB1 recruitment and in turn proviral silencing.

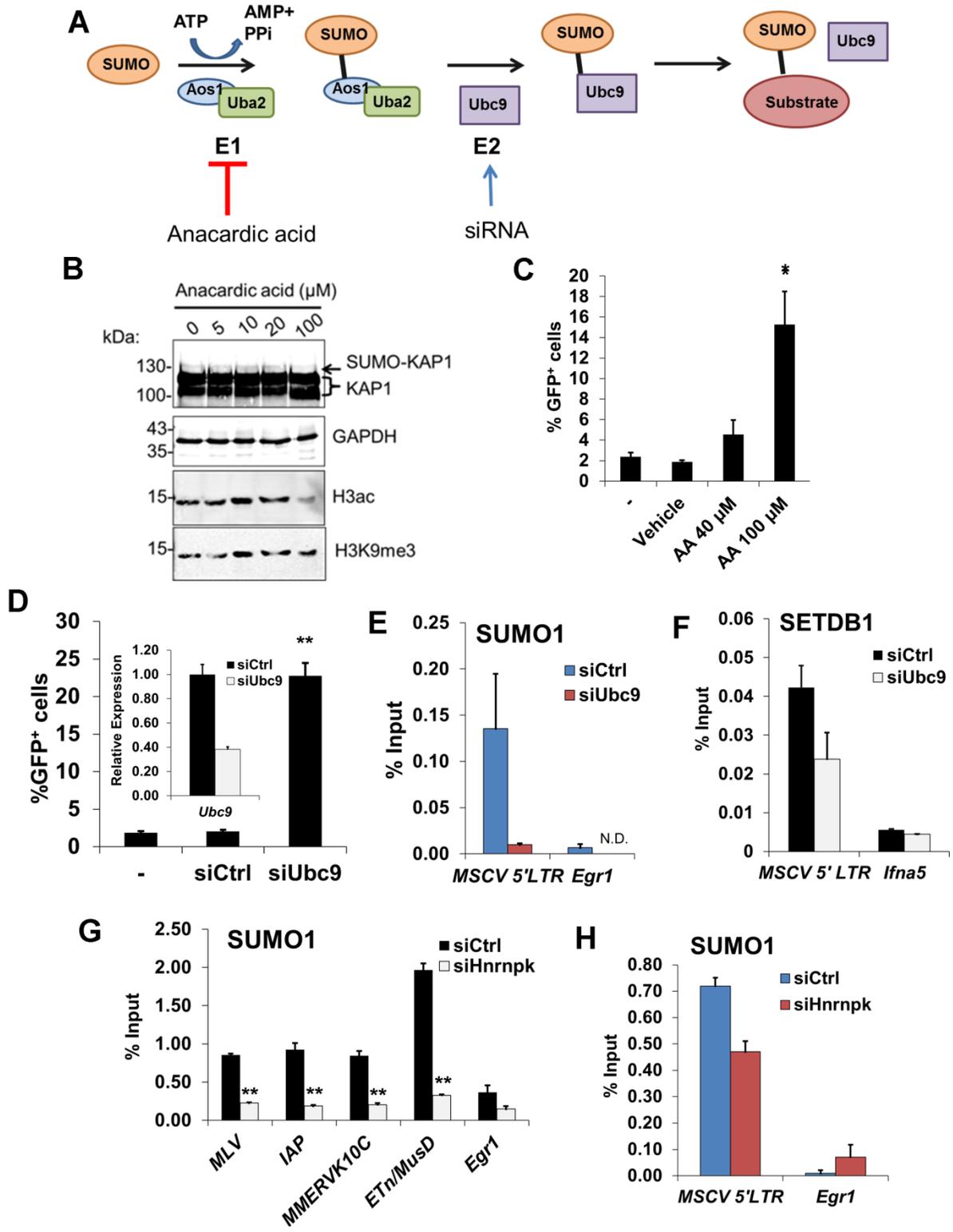


Figure 9. Depletion of hnRNP K leads to loss of SUMOylation on ERV chromatin, which is necessary for SETDB1 recruitment and proviral silencing. (A) Schematic of the SUMO conjugation pathway showing the activities of the SUMO E1 activating heterodimer enzyme (Aos1/Uba2 also called SAE1/2) and the SUMO conjugating enzyme Ubc9. Anacardic acid was used to inhibit the SUMO E1 activating enzyme and siRNAs were used to target *Ubc9* mRNA. (B) Western blot analysis of KAP1 mono-SUMOylation upon treatment of MSCV-GFP mESCs with indicated concentrations of anacardic acid (AA). GAPDH was a loading control. AA also inhibits some H3 lysine acetyltransferases, as indicated by the lower levels of H3 acetylation during AA treatment but total H3K9me3 was unaffected. (C) Flow cytometry of MSCV-GFP cells untreated (-) or treated with vehicle (DMSO) or indicated concentrations of AA. Data are mean of n = 3 biological replicates, error bars are s.d. (D) Flow cytometry as in (C) except for untransfected (-), control or *Ubc9* siRNA transfected cells at 48 h post-transfection. Inset, qRT-PCR validation of *Ubc9* KD at 24 h post-transfection. (E) ChIP of SUMO1 on unsorted MSCV-GFP control and *Ubc9* KD cells at 48 h post-transfection. (F) ChIP of SETDB1 on the same cells as in (E). ND = not detected in 40 cycles of qPCR. (G) ChIP of SUMO1 on TT2 control and *Hnrnpk* KD cells at 24 h post-transfection. (H) ChIP of SUMO1 on unsorted MSCV-GFP control and *Hnrnpk* KD cells at 72 h post-transfection. All ChIP-qPCR data are mean enrichment relative to input from n = 3 technical replicates, error bars are s.d. *p < 0.01, **p < 0.001 two-tailed T-test.

3.11 Depletion of hnRNP K does not affect bulk KAP1 mono-SUMOylation and hnRNP K does not directly regulate KAP1 SUMOylation in vitro

The results thus far suggested a model in which hnRNP K regulates the SUMOylation of ERV-bound proteins, such as KAP1. To test this idea, total KAP1 mono-SUMOylation was assayed by western blot in hnRNP K KD cells (**Figure 10A**). Although depletion of Ubc9 led to reduced SUMOylation, at both 24 and 72 h post-transfection, KAP1 mono- and di-SUMOylation was not changed in hnRNP K KD cells relative to the control KD cells (**Figure 10A**). To determine whether hnRNP K directly stimulates KAP1 SUMOylation conjugation, *in vitro* SUMOylation assays were performed using recombinant GST-tagged KAP1, recombinant KRAB-ZFP ZIK1, which was shown to directly interact with hnRNP K²⁸⁰ (**Figure 10B**). However, the presence of either ZIK1 and/or hnRNP K did not substantially stimulate KAP1 SUMOylation under efficient SUMOylation conditions (**Figure**

10B). Under conditions where SUMOylation cascade components were limiting and only mono-SUMOylation of KAP1 was achieved, only a very modest increase was detected in the presence of hnRNP K and ZIK1, but this did not follow a dose-dependent trend (**Figure 10C**). Similarly, hnRNP K did not apparently stimulate mono-SUMOylation of another direct binding partner p53 (**Figure 10D**). As KAP1 can be de-SUMOylated by SENP1,²⁶¹ I next tested whether hnRNP K would be refractory to SENP1-mediated de-SUMOylation. To this end, GST-KAP1 was SUMOylated *in vitro* and incubated recombinant SENP1 catalytic domain (SENP1^{CD}) to catalyze de-SUMOylation (**Figure 10E**). SENP1^{CD} de-SUMOylation was efficient and specific as evidenced by the ability of the alkylating agent and SENP inhibitor NEM to block its activity (**Figure 10E**). However, the pre-incubation of the SUMOylated KAP1 with its binding partners including ZIK1, SETDB1 and/or hnRNP K had no effect of de-SUMOylation by SENP1^{CD} (**Figure 10E**). Although a role for hnRNP K in regulating KAP1 SUMOylation *in vivo* cannot be excluded on the basis of these experiments, they nevertheless suggest that hnRNP K may exert a subtle and localized effect in promoting KAP1 SUMOylation when bound to chromatin and/or that hnRNP K regulates SUMOylation of an alternative factor(s) at ERV chromatin.

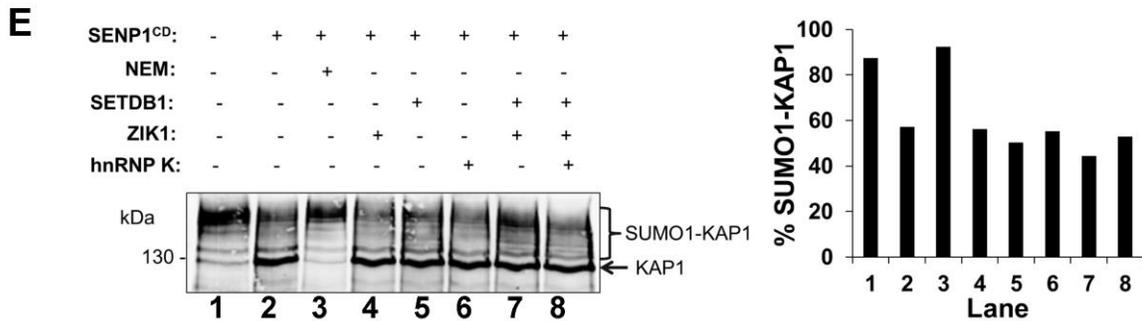
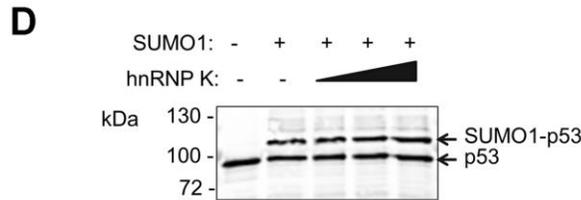
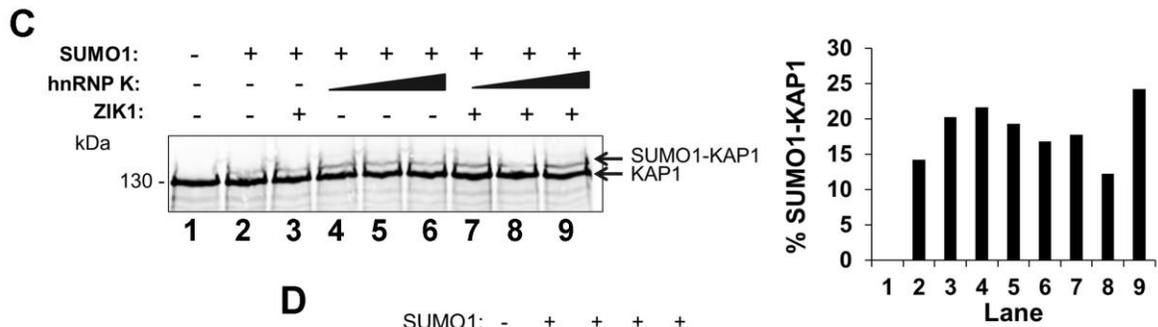
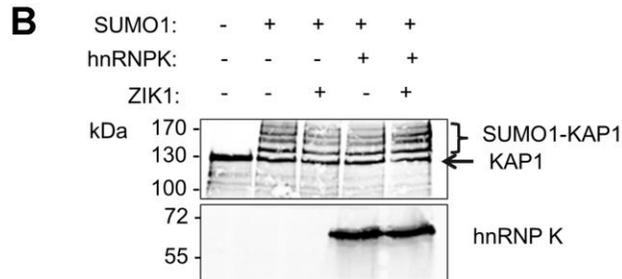
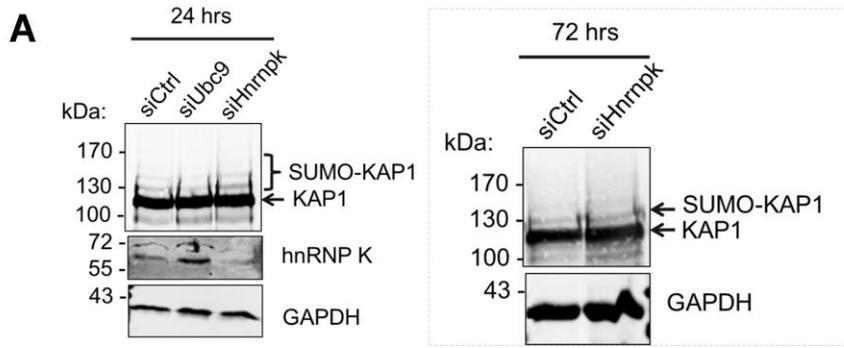


Figure 10. Depletion of hnRNP K does not abolish KAP1 mono-SUMOylation and hnRNP K does not directly regulate KAP1 SUMOylation *in vitro*. (A) Western blot analysis of SUMOylated KAP1 in control, Ubc9, or hnRNP K KD TT2 cells at 24 h post-transfection and the control and hnRNP K KD at 72 h. (B) *In vitro* KAP1 SUMOylation assay and western blot of KAP1 under reaction conditions with high SUMOylation cascade components and the presence of ZIK1 and/or hnRNP K. (C) *In vitro* KAP1 SUMOylation assay as in (B) except using limiting concentrations of SUMO E1 and E2, with ZIK1 and increasing amounts of hnRNP K. Quantification shown on the right, where the proportion of amount of SUMO1-KAP1 is normalized to the total KAP1 (SUMO1-KAP1 + unmodified KAP1). (D) *In vitro* p53 SUMOylation assay with increasing amounts of hnRNP K as in (C). (E) *In vitro* KAP1 de-SUMOylation assay, with the presence or absence of indicated recombinant proteins. SENP1 catalytic domain (SENP1^{CD}) was used to de-SUMOylate KAP1 and was specifically inhibited with NEM. To the right is shown the quantification, as in panel (C).

3.12 MCAF1/mAM is required for SETDB1-dependent proviral silencing

Among the other SETDB1-associated proteins in mESCs, MCAF1 was found to be the only factor detected in both IP approaches (**Figures 1C and 1E**). MCAF1 forms a core heteromeric complex with SETDB1 and promotes its catalytic activity towards H3K9me2 to produce H3K9me3.³³ I therefore interrogated the role of MCAF1 in SETDB1-mediated proviral silencing. Interestingly, *Setdb1* KD cells showed a ~4-fold upregulation of *Mcaf1* expression (**Figure 11A**) indicating that the level of *Mcaf1* expression is sensitive to the level of SETDB1. Notably, depletion of *Mcaf1* transcripts led to de-repression of the MSCV proviral reporter (**Figures 11B**), revealing that this catalytic co-factor of SETDB1 also plays a role in proviral silencing. Similarly, KD of *Mcaf1* in the *Dnmt* TKO cells also resulted in upregulation of class I and II ERVs beyond what was observed in the *Dnmt* TKO line alone with the most dramatic upregulation observed for MMERVK10C elements (**Figure 11C**) which are also highly sensitive to SETDB1 KD (**Figure 5E**). Furthermore, H3K9me3 was dramatically reduced at both ERVs and the MSCV-GFP proviral reporter in *Mcaf1* KD cells,

phenocopying the *Setdb1* KD (**Figure 11D**). However, in contrast with hnRNP K, KD of MCAF1 did not perturb SETDB1 enrichment at ERVs or the MSCV provirus (**Figure 11E**). Together these results support a role for MCAF1 in maintenance of H3K9me3 by SETDB1 at proviral chromatin.

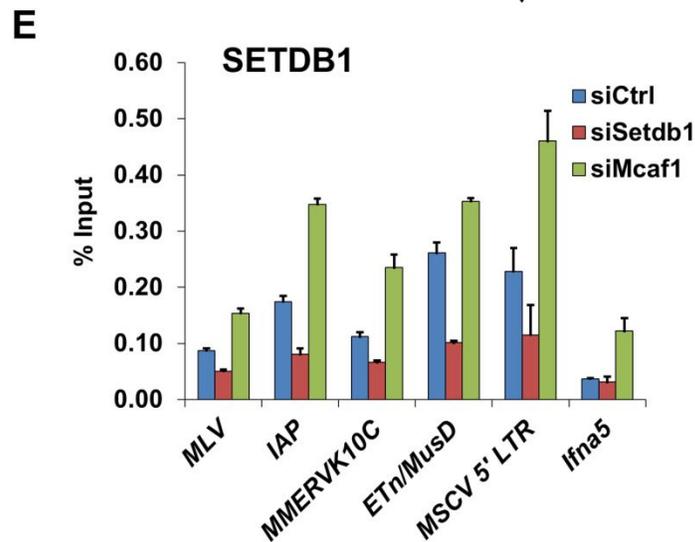
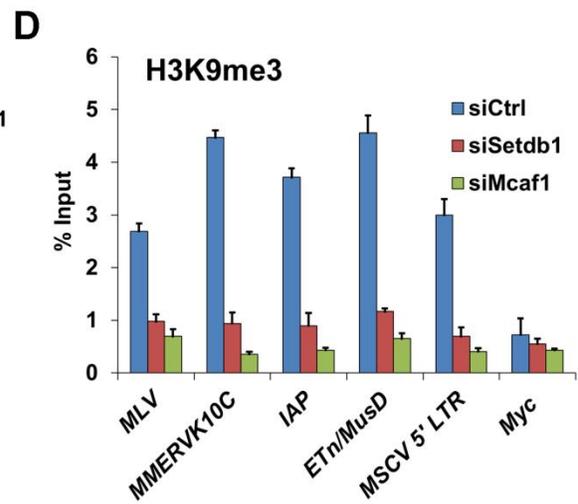
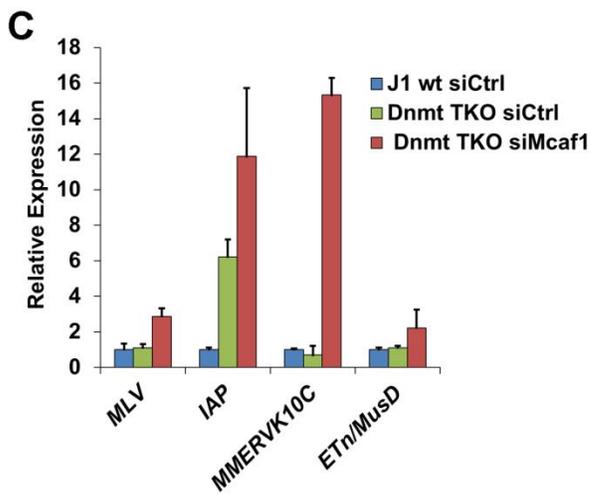
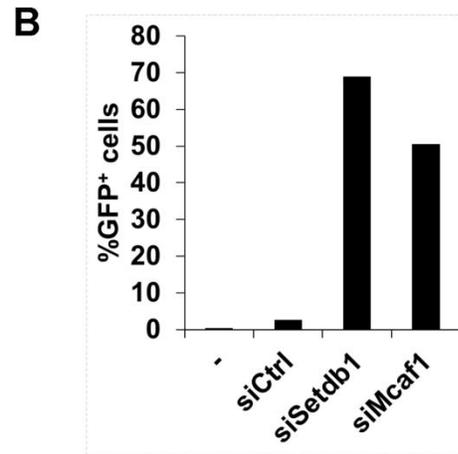
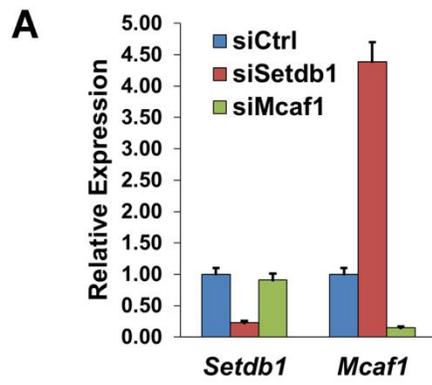


Figure 11. MCAF1 is required for proviral silencing. (A) qRT-PCR validation of *Setdb1* and *Mcaf1* KDs in the MSCV-GFP cell line. (B) Flow cytometry of GFP⁺ cells from the untransfected parent line (-) and cells transfected with control, *Setdb1* or *Mcaf1* siRNAs at 72 h post-transfection. Data are percentage of GFP⁺ cells from one replicate of 10,000 cells. (C) qRT-PCR analysis of ERV expression in J1 wt and Dnmt TKO cells transfected with control or *Mcaf1* siRNAs at 96 h post-transfection. (D) NChIP for H3K9me3 in MSCV-GFP cells transfected with indicated siRNAs at 72 h post-transfection. The Myc promoter was a negative control for H3K9me3 (E) ChIP for SETDB1 in the cells from (D) at 72 h post-transfection. The *Ifna5* promoter was a negative control for SETDB1. All ChIP data are mean enrichment normalized to input chromatin from n = 3 technical replicates, error bars are s.d.

3.13 Discussion

KRAB-ZFP/KAP1 complexes^{220,301} are thought to play a central role in repression of ERV transcription in pluripotent stem cells via SETDB1 recruitment.^{89,90} Although the roles of KAP1 and SETDB1 in this proviral silencing pathway have been defined, the role of SUMOylation and the presence of additional co-repressors had not been addressed. In this work, I used biochemical approaches to identify novel candidate co-repressors for the SETDB1/KAP1 silencing machinery. The results of further genetic analysis support the conclusion that hnRNP K is a co-repressor required for recruitment of SETDB1 to proviral chromatin, H3K9me3 and efficient proviral silencing. In contrast, the SETDB1 catalytic co-factor MCAF1 is required for H3K9me3 and silencing, but not SETDB1 recruitment at proviral chromatin. HnRNP K is a highly conserved, multi-functional protein involved in transcription regulation, mRNA splicing and translation.¹³⁹ Although no null mutant has been characterized in the mouse, studies in flies, yeast and in mammalian cell lines support the view that hnRNP K plays important roles in development and gene regulation.^{302,303} Early work using yeast-two-hybrid screens revealed that hnRNP K could directly interact with chromatin regulatory proteins, such as the PRC2 subunit EED³⁰⁴ and KRAB-ZFPs Zik1 and Kid1,^{139,280} suggesting that it may regulate Polycomb and/or KRAB-ZFP/KAP1 complexes.

Moreover, recent work has converged upon a novel function for hnRNP K in binding lncRNAs to regulate gene expression and recruitment of active or repressive histone modifications at gene promoters in different developmental and cellular contexts.^{126,140,141,305,306}

My work significantly builds upon the known functions of hnRNP K by demonstrating an indispensable role for this factor in KRAB-ZFP/KAP1/SETDB1-dependent proviral silencing. Based on the results presented here, I propose a novel model for the SETDB1/KAP1 proviral silencing pathway incorporating hnRNP K (**Figure 12A**). In wt mESCs, KRAB-ZFPs recruit KAP1 in an oligomeric state, possibly as a homotrimer,^{213,285} to proviral chromatin and unmodified KAP1 may recruit hnRNP K. HnRNP K may promote the SUMOylation of KAP1 and/or other factors on chromatin, which then serves as a ligand for the SETDB1/MCAF1 complex,^{258,307} eliciting H3K9me3 deposition at SUMOylated KAP1-bound proviral LTRs (**Figure 12A**). In hnRNP K-deficient cells, SUMOylation on chromatin may be compromised, leading to reduced SETDB1 recruitment at ERVs, diminution of H3K9me3 and eventual transcriptional de-repression (**Figure 12B**). This model is consistent with recent ChIP-seq analyses of SUMO1, SUMO2 and Ubc9 in human fibroblasts³⁰⁸, which show co-occupancy with sites of KAP1, SETDB1 and H3K9me3 at the 3' ends of KRAB-ZFP genes³⁰⁹ indicating that these SETDB1/KAP1-bound, SUMOylated loci are sites of active KAP1 SUMOylation on chromatin.³⁰⁸ Although a possible contraindication to this model is the observation of the differing binding affinities of SETDB1 and hnRNP K for SUMOylated KAP1 *in vitro*, this could be rationalized by: 1) the existence of multiple KAP1 subunits in each complex such that some are SUMOylated while others are unmodified,

providing binding sites for both SETDB1 and hnRNP K simultaneously, and/or 2) the observation that hnRNP K directly binds to certain KRAB-ZFPs such as ZIK1^{139,280} and therefore may still indirectly interact with SUMOylated KAP1. Indeed, consistent with the former possibility, rather than solely containing SUMOylated KAP1, SETDB1 complexes contained predominantly unmodified KAP1 with only a minority SUMOylated KAP1 under conditions where mono- and di-SUMOylated KAP1 was preserved in mESC nuclear extracts (**Figure 2A**). This observation indicates that SETDB1 binding to KRAB-ZFP/KAP1 complexes *in vivo* may only require a small proportion of the total KAP1 in the complex to be SUMOylated. Germane to this possibility is the observation that KAP1 SUMOylation is highly dynamic and previous investigations have relied on overexpression of SUMO paralogues to detect it.^{259,261,291,310} Therefore, it is also possible that hnRNP K facilitates transient KAP1 SUMOylation events in a cell cycle-dependent manner, such as during S-phase when chromatin modifications must be re-established. How hnRNP K promotes the SUMOylation of KAP1 or other proteins on chromatin remains to be determined, although given that hnRNP K is a SUMO target itself and can directly interact with Ubc9,^{242,287} it may facilitate recruitment of this SUMO E2 enzyme to KAP1-bound loci. Alternatively, hnRNP K might also counteract SENP activity toward KAP1 providing an additional layer of regulation over KAP1de-SUMOylation. Since KAP1 is constitutively phosphorylated at Ser824 in pluripotent stem cells,²⁷¹ another intriguing possibility is that hnRNP K counteracts the activity of the SUMO-targeted ubiquitin ligase RNF4, which conjugates ubiquitin to Lys676 SUMOylated, Ser824 phosphorylated KAP1 promoting its degradation.³¹⁰

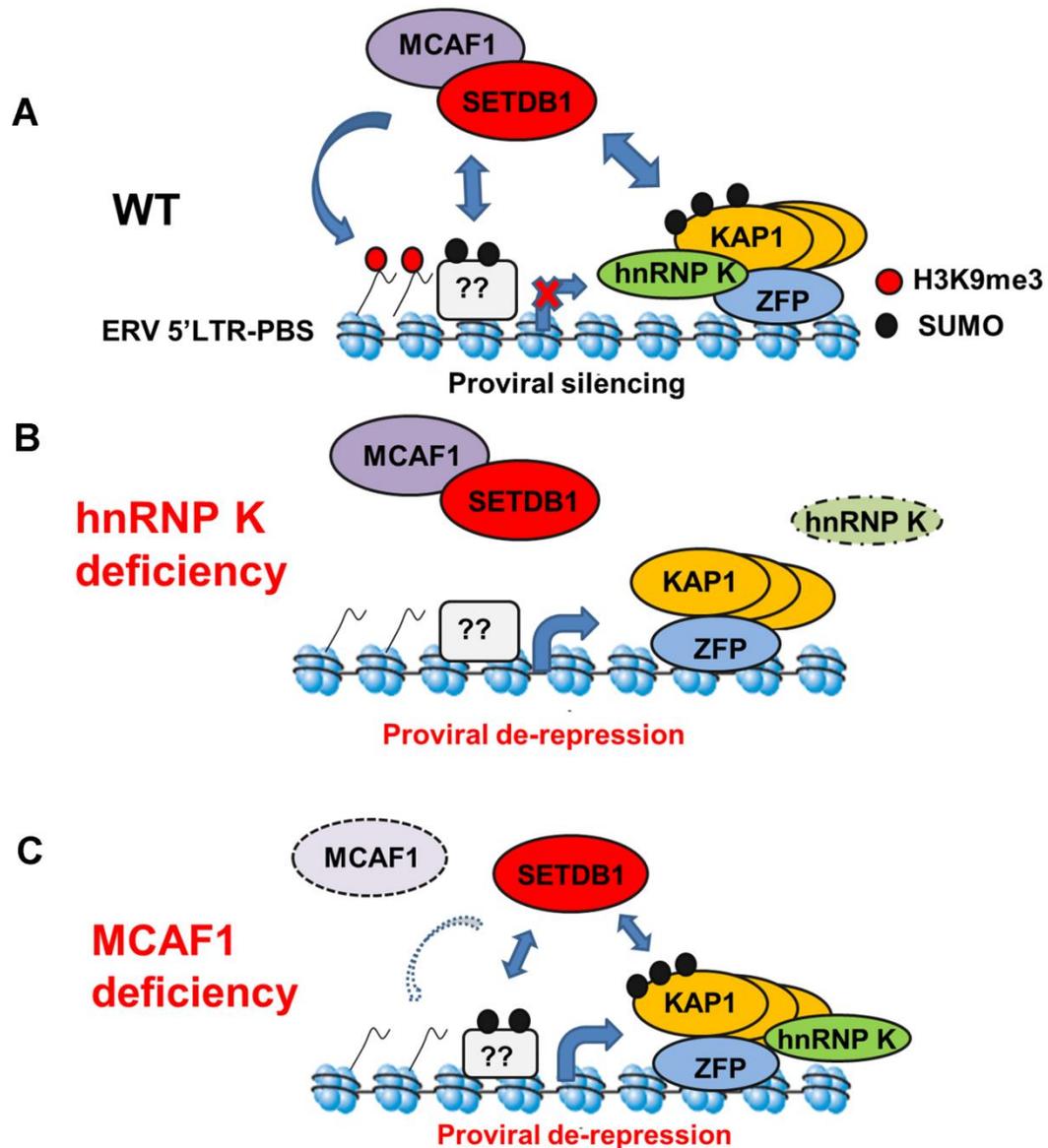


Figure 12. Revised model for SETDB1/KAP1-mediated proviral silencing incorporating the functions of hnRNP K and MCAF1. (A) In wt mESCs, KRAB-ZFPs and other zinc finger proteins bind proviral sequences to recruit KAP1. KAP1 recruits hnRNP K, which promotes or enhances the SUMOylation of KAP1 and/or other chromatin bound factors (?). SUMOylated proteins including KAP1 provide a binding site for SETDB1/MCAF1 which catalyzes H3K9me3 to enforce a silenced state. (B) In hnRNP K deficient mESCs, SUMOylation of KAP1 and/or other chromatin-bound factors is reduced, leading to reductions in SETDB1 binding and H3K9me3 levels and eventual proviral de-repression. (C) In MCAF1 deficient mESCs, SETDB1 can still be recruited by SUMOylated KAP1 and other factors, but its catalytic activity is insufficient to support H3K9me3, leading to proviral de-repression.

Similar to SETDB1 KD mESCs,²⁴⁶ KD of hnRNP K only resulted in modest upregulation of class I and II ERVs in wt mESCs cultured in serum. A likely explanation for this observation is the relatively high level of DNA methylation in mESCs cultured in serum relative to two-inhibitor (2i) media. Under the latter conditions, mESCs adopt a “naïve” hypomethylated state, more reflective of the inner cell mass of the E3.5 blastocyst from which they are derived.⁹³ Consistent with this model, depletion of hnRNP K in DNA methylation-deficient cells led to a more robust upregulation of class I and II ERVs as compared with wt cells and previous work has shown that IAP elements are synergistically upregulated upon KD of both SETDB1 and DNMT1 in serum-cultured mESCs.²⁴⁶ Thus siRNA KDs in serum-cultured mESCs are likely not robust enough to elicit loss of DNA methylation at ERVs controlled by SETDB1, despite losses of H3K9me3. In contrast with ERVs, for reasons that are not entirely clear, knocking down SETDB1 in serum-cultured mESCs harbouring a newly integrated silent MSCV provirus results in losses of both H3K9me3 and DNA methylation at the 5'LTR concomitant with de-repression.^{89,237}

My work has also defined a crucial role for MCAF1 in SETDB1-mediated proviral silencing, consistent with its role in enhancing SETDB1 catalytic activity towards H3K9me2 to generate H3K9me3.³³ This is the first report of a developmental role for MCAF1 with SETDB1 in mammals and consistent with this observation, MCAF1 was also identified in a recent genome-wide screen of factors required for the establishment of proviral silencing.³¹¹ In contrast to hnRNP K-depleted cells, SETDB1 recruitment is maintained but H3K9me3 is no longer efficiently deposited at proviral chromatin in MCAF1-deficient mESCs (**Figure 12C**). Thus although MCAF1 like SETDB1 can directly bind SUMOylated proteins,¹⁴⁵ this

activity is apparently not required for recruitment of the SETDB1/MCAF1 complex to ERVs. The MCAF1 KD phenotype is consistent with the observation that the catalytic activity of SETDB1 is crucial for full ERV repression.^{89,237} Notably, a previous report showed that the MCAF1 orthologue Windei is necessary for dSETDB1/Eggless function in the *Drosophila* germline,³¹² revealing a conserved role for this co-factor in SETDB1 function.

In addition to hnRNP K and MCAF1, several other proteins were associated with SETDB1 in mESCs and might also play roles in proviral silencing. Key candidates would include the factors found in the SETDB1 IP under SUMO-enriched conditions with known roles in transcriptional regulation such as the DNA-binding transcriptional repressor ZFP161, E3 ubiquitin ligase TRIP12 and the arginine methyltransferase PRMT1. Indeed, in the case of the latter, a recent study reported that histone H2A/H4R3 methylation by a related enzyme, PRMT5, marks IAP in early stage PGCs and deletion of *Prmt5* leads to modest (~2-fold) reactivation of ERVs in this context.³¹³ Future work could validate these interactions and determine whether for instance, ZFP161 promotes SETDB1/KAP1 recruitment to ERVs and TRIP12 ubiquitin ligase activity regulates protein turnover at ERVs to maintain their silencing.

While this work being completed, Tchasovnikarova and colleagues³¹⁴ reported the identification three proteins TASOR, MPP8 and Periphilin, which were required to maintain the silencing of SETDB1-targeted, H3K9me3-marked genes in human cell lines. These factors were found via a forward genetic screen in the near haploid human cell line KBM7 and subsequently shown to form a complex that interacts with SETDB1, binds to H3K9me3-

rich loci to maintain this mark and enforce a silent state.³¹⁴ Apart from technical differences in the IP procedure employed by this study³¹⁴ and my work, the absence of these proteins in my SETDB1 IP-MS analyses in mESCs suggests that they may only associate with SETDB1 in other cell types, such as adult somatic cell lines. While the role of TASOR and Periphilin in SETDB1-dependent proviral silencing has not been determined, siRNA KD of MPP8, which directly binds H3K9me3,³¹⁵ was found to perturb silencing of specific proviral elements in mESCs in one study³¹¹ but not another,²³⁷ indicating that the functions of this complex could also be involved in this pathway.

Intriguingly, class III MERVL elements, which are marked by H3K9me2 and de-repressed in mESCs deficient for *G9a* or *Glp*,⁸⁶ were strongly induced in hnRNP K KD cells. Since these elements are also de-repressed in *Kap1* KO but not *Setdb1* KO mESCs,^{86,90} hnRNP K may play a role in SETDB1-independent chromatin regulatory pathways with KAP1 and G9a/GLP. This possibility is investigated further in chapter 4.

In addition to ERVs, a cohort of SETDB1/H3K9me3-repressed male germline-specific genes²⁴⁶ are bound at their promoters by hnRNP K, show reduced H3K9me3 and increased expression upon hnRNP K KD, indicating a role for hnRNP K in SETDB1/H3K9me3-mediated gene repression. Maeda *et al.*³¹⁶ showed that the repression of a cohort of these male germline genes in mESCs requires the TF Max, which recruits G9a/GLP and H3K9me2. Therefore, these genes are repressed by the combined activities of G9a/GLP and SETDB1, which could be explained by G9a/GLP producing H3K9me2 to be converted to H3K9me3 by SETDB1. How SETDB1 and H3K9me3 may be targeted to these promoters by

hnRNP K remains unclear but it is likely to be KAP1-independent, since these genes are not upregulated in *Kap1* KO cells.⁹⁰ While this work was underway, Bao *et al.*³⁰⁶ reported that hnRNP K forms a complex with SETDB1 mediated by *lncRNA-p21* in MEFs undergoing somatic cell reprogramming, which provides a block to reprogramming by deposition of H3K9me3 at pluripotency gene promoters. Despite the interactions between hnRNP K and SETDB1 in mESCs not being sensitive to RNase A digestion (**Figure 3D**), this study nevertheless raises the possibility that some proportion of the nuclear hnRNP K and SETDB1 might participate in a KAP1-independent complex that involves lncRNAs. Another possibility is that hnRNP K promotes SUMOylation of proteins other than KAP1 on chromatin, leading to SETDB1 binding and transcriptional silencing. In addition to SETDB1/H3K9me3, hnRNP K may also promote PRC2/H3K27me3-mediated gene repression in mESCs, since a cohort of PRC2 target genes, including *Gata6* and *Nkx2-9* were strongly upregulated in hnRNP K KD cells. This is consistent with a previous report showing that hnRNP K can promote PRC2-dependent repression via recruitment of the subunit EED to a heterologous promoter.³⁰⁴ Also, hnRNP K was recently shown to function in *Xist*-mediated H3K27me3 and transcriptional silencing of genes on the inactive X in mouse cells,¹²⁶ providing another link between hnRNP K and PRC2 in gene repression. Further studies will be necessary to clarify the contributions of hnRNP K to SETDB1- and PRC2-mediated transcriptional silencing at specific genes in mESCs.

4. Role of hnRNP K, G9a/GLP and H3K9me2 in silencing of MERVL elements in mESCs

4.1 Background and Summary

Class III MERVL elements are among the oldest ERVs in placental mammals and are present as ~37,000 solitary LTRs and ~600-700 full-length copies in the C57BL/6 mouse genome ~350 of which carry intact protein-coding *gag*, *pol* and *dUTPase* genes³¹⁷ (reviewed by Schoorlemmer *et al.*³¹⁸) Notably, a recent burst of retrotransposition occurred in the mouse lineage, indicative of emergence of a “new” variant of MERVL elements in this lineage.³¹⁷ MERVL transcripts are among the most abundant in the 2-cell embryo³¹⁹ and sequences derived from these young ERVs have been co-opted as regulatory elements during embryogenesis with many 2-cell stage-specific genes in the mouse driven by MERVL LTRs as their promoters.^{85,319–321} Indeed, MERVL sense transcripts are necessary for progression beyond the 2-cell stage, as antisense oligo interference leads to arrest at the 4-cell stage.³²² Following the 2-cell stage, MERVL is subject to transcriptional repression, which persists in the blastocyst and mESCs.^{86,323}

In contrast to class I and II ERVs, which are silenced by histone H3K9 trimethylation (H3K9me3) mediated by the SETDB1/KAP1 complex, MERVL elements are not marked with H3K9me3 in mESCs, nor are they de-repressed upon *Setdb1* deletion.^{86,95,246}

Furthermore, although MERVL sequences are DNA methylated, the loss of DNA methylation alone has a modest effect on MERVL expression in mESCs^{246,321,324} perhaps due to their relatively low C-G density in comparison with many class I and II ERVs. Also, MERVL elements differ from class I and II ERVs in that the former harbour lower levels of SUMO1,²⁴⁵ pointing to a SUMO-independent silencing mechanism at these ERVs. Instead

MERVL elements are generally enriched in H3K9me2 and bound by the G9a/GLP complex in mESCs^{86,321} (**Illustration 5**). The loss of G9a, GLP or the catalytically active G9a leads to de-repression of MERVL,⁸⁶ suggesting a direct role for G9a/GLP-mediated H3K9me2 in MERVL silencing (**Illustration 5**). The LSD1/CoREST complex³²⁵ is also required for MERVL silencing via deacetylation of histones and demethylation of H3K4me1/2, since both HDAC inhibition and loss of LSD1 or its catalytic activity perturbs MERVL silencing.³²¹ Notably, MERVL is also de-repressed upon the loss of KAP1, HP1 α and HP1 β although their effects are likely indirect as these factors are not bound at MERVL in mESCs.^{86,90} Conditional knockout of the PRC1 subunit RYBP also leads to de-repression of MERVL elements in mESCs³²⁶ but the mechanism by which RYBP influences MERVL silencing is unknown. Interestingly, the depletion of each of these chromatin modifiers leads not only to MERVL upregulation, but also to the upregulation of a cohort of 2-cell stage genes, some of which encode TFs promoted by MERVL LTRs.^{85,86,321,326} Therefore, an intriguing possibility which remains to be investigated is that MERVL elements may be activated by 2-cell stage TFs and that the transcriptional inactivity of MERVL after the 2-cell stage is a direct result of the absence of activating TFs. Thus while MERVL repression correlates with the presence of H3K9me2, it remains to be determined whether the transcriptional inactivity of MERVL elements is inexorably dependent on H3K9me2.

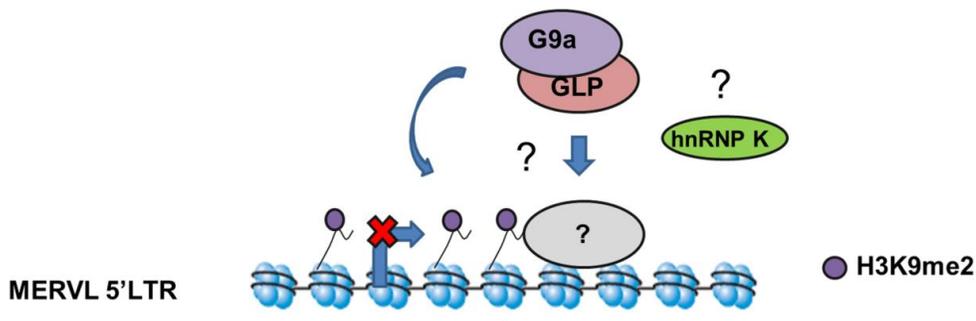


Illustration 5. Current model for the transcriptional silencing of MERVL elements by G9a/GLP. MERVL elements are marked by H3K9me2 and bound by G9a/GLP, which are required for maintaining their silencing in mESCs. It is unknown whether MERVLs can autonomously recruit G9a/GLP via repressive TFs or whether this KMT complex marks these elements as an indirect consequence of their genomic location. HnRNP K KD mESCs de-repress MERVL elements as determined by RNA-seq and qRT-PCR, but the role of hnRNP K in MERVL repression is unknown.

There are several outstanding questions concerning the mechanism of MERVL repression that are yet to be addressed. First, previous studies have not unequivocally shown whether MERVL sequences can independently recruit chromatin modifiers to achieve silencing or whether silencing is a function of genomic location (**Illustration 5**). For instance, in contrast to the silencing of MLV-based retroviral vectors, which require the PBS^{Pro} sequence^{89,220} and specific IAP elements, which require a ~500 bp 5'UTR sequence or a ~160 bp *gag* sequence,^{90,257} there are no known MERVL sequences that have been shown to autonomously confer transgene repression. Notably, a ~250 bp region of the LTR is necessary and sufficient to drive luciferase expression upon transient transfection in *Kdm1a* (LSD1) KO mESCs,³²¹ raising the prospect that sequences within the LTR that can act as promoters may also direct silencing. Similarly, while KRAB-ZFPs and other zinc finger transcription factors such as YY1 are known to promote the silencing of class I and II ERVs and their retroviral vector counterparts^{220,254,301} the specific TFs that direct MERVL silencing have not been defined. The pluripotency-associated TF Rex1 (also called Zfp42) was suggested to contribute to MERVL repression in mESCs.³²⁴ Rex1 is bound at MERVL in mESCs, however no upregulation of MERVL is observed in *Rex1*^{-/-} cells and the degree of de-repression observed upon RNAi knockdown of *Rex1* is relatively modest (~2-fold)³²⁴ in comparison with loss of G9a or GLP (~8-12-fold)⁸⁶ suggesting that it plays only a minor role.

As shown in chapter 3, in addition to the reduction in SETDB1 recruitment and H3K9me3 at class I and II ERVs and MSCV proviral reporter constructs, hnRNP K KD cells also dramatically de-repress MERVL elements²⁴⁵ suggesting that hnRNP K may also participate in MERVL silencing. Indeed, hnRNP K is thought to regulate transcription in several different ways including indirectly as a co-activator³²⁷ and directly by binding to single-stranded DNA sequences in the promoter regions of genes in collaboration with other transcription factors and the preinitiation complex.^{328–330} HnRNP K also binds along the body of immediate-early genes, such as *EGR1* upon serum stimulation in a co-transcriptional fashion to promote efficient termination.^{296,297} HnRNP K can also regulate transcription and chromatin states via binding to lncRNAs such as *lncRNA-p21*, *Tunar*, *Xist* and *EWSAT1* to direct the deposition of chromatin modifications and regulate cohorts of genes.^{126,140,141,305,306,331} As with *G9a* and *Kap1*, *Hnrnpk* expression is also anti-correlated with MERVL in the preimplantation embryo,^{263,319} consistent with a potential role for hnRNP K in repressing MERVL *in vivo*.

In this study, I found that MERVL elements in the genome can be generally divided into those that occur in large H3K9me2-rich domains and those that are in regions lacking H3K9me2. H3K9me2 can be acquired upon inhibition of transcriptionally active elements, suggesting that it marks a subset of silent elements, rather than establishing a silent state. Furthermore, a newly integrated MERVL LTR was incapable of autonomously recruiting H3K9me2 and did not require H3K9me2 to be transcriptionally silent. HnRNP K formed a novel RNA-dependent complex with G9a/GLP and is required for a proportion of global H3K9me2 in mESCs, which impacts the repression of MERVLs enriched in H3K9me2.

However, depletion of hnRNP K also led to upregulation of a MERVL LTR reporter that was not marked by H3K9me2. Interestingly, hnRNP K and G9a/GLP were bound to MERVL nuclear RNAs and a point mutation in hnRNP K that abrogates binding of its KH3 domain to nucleic acids led to de-repression of MERVL in a G9a-dependent manner, indicating that hnRNP K binding to RNA and/or DNA is important for MERVL repression by G9a/GLP. Taken together, these data reveal novel insights into the mechanisms governing MERVL silencing during development and point to an important regulatory role for 2-cell stage TFs in driving expression of MERVL.

4.2 MERVL elements are marked by H3K9me2 as a consequence of genomic location and transcriptional inactivity

Since we previously reported that MERVL is enriched in H3K9me2 and de-repressed in *G9a* and *Glp* KO mESCs,⁸⁶ I first investigated the relationship between MERVL and H3K9me2 enrichment genome-wide using unpublished H3K9me2 native ChIP-seq datasets from TT2 wt and *Glp*^{-/-} mESCs generated in the Lorincz lab by Carol Chen. Consistent with previous mapping of H3K9me2 by ChIP-chip in mESCs, this mark forms broad megabase-sized domains across transcriptionally inactive regions encompassing ~30-50% of the genome.^{82,332}

To determine whether H3K9me2 is recruited autonomously by MERVL sequences and spreads into the flanking regions, the genome-wide H3K9me2 distribution was analyzed in the 6 kb up or downstream of full-length MERVL elements in TT2 and *Glp* KO cells by Dr. M. Karimi (**Figure 13A**). IAP elements, which recruit H3K9me3 in an autonomous fashion by SETDB1/KAP1,^{90,237,257} frequently exhibit spreading of H3K9me3 ~1-2 kb into flanking

regions.^{79,86} In contrast, there was no substantial change in the H3K9me2 distribution over the flanks of MERVL elements in wt or *Glp* KO cells (**Figure 13A**) suggesting that MERVLs do not exhibit a difference in the levels of H3K9me2 over the transition from the edges of the elements into flanking regions. H3K9me2 was increased in the flanks of IAPs in *Glp* KO cells (**Figure 13A**), which normally harbour low levels of H3K9me2 but high H3K9me3. Since *Glp* KO cells globally lose H3K9me2, this effect may be due to increased antibody binding that captures the H3K9me2 intermediate in the process of producing H3K9me3. Importantly, there were clear examples of MERVL elements reactivated in *Glp* and *G9a* KO mESCs that were located in broad, late-replicating H3K9me2 domains (**Figure 13B**) consistent with the absence of focal H3K9me2 enrichment in the flanks of MERVLs that decreases to background levels. To examine this relationship at all MERVLs, H3K9me2 levels were sampled using unique reads over 6 kb upstream from 5' flank of 656 individual full-length MERVL elements (**Figure 13C**). This analysis confirmed that H3K9me2 was not substantially higher towards the immediate 5' upstream flank of MERVL elements over the 6 kb region (**Figure 13C**), which suggests against direct recruitment of H3K9me2 and spreading into the flanks. More importantly, the majority of MERVLs were located in uniquely mapped regions, as evidenced by the mappability scores in the 5' flanking region and these elements were located in high, intermediate and low H3K9me2 regions (**Figure 13C**). There was a loss of H3K9me2 in the 5' flanking region in *Glp* KO cells (**Figure 13C**), confirming the general requirement for *G9a/GLP* in generating H3K9me2 regions near MERVL elements. Similar results were obtained when H3K9me2 was sampled over 6 kb downstream of the 3' end of these 656 MERVLs (data not shown).

Under normal serum culture conditions, a small proportion of cells in a mESC population cycle in and out of a 2-cell-like state where they express MERVL.⁸⁵ To determine whether H3K9me2 deposition on MERVL elements occurs as a consequence of transcriptional inactivity, Pol II initiation was inhibited by the small-molecule inhibitor triptolide (Trp), which blocks TFIID function.³³³ Consistent with previous results,⁵⁹ incubation of mESCs for 3.5 h in Trp led to depletion of transcriptionally engaged Pol II and total Pol II due to proteasomal degradation (**Figure 13D**). After 8.5 h Trp, ChIP for H3K9me2 revealed increased levels at MERVL LTRs (**Figure 13E**) suggesting that H3K9me2 deposition can occur as a consequence of Pol II inactivity. Taken together these data indicate that MERVLs do not show evidence of spreading of H3K9me2 into their flanks, which is a hallmark of autonomous recruitment of H3K9 methylation and that a subset of MERVLs are located in genomic regions with low H3K9me2. In addition, the H3K9me2 mark may be ectopically deposited at a subset of MERVL LTRs as a result of transcriptional inactivity.

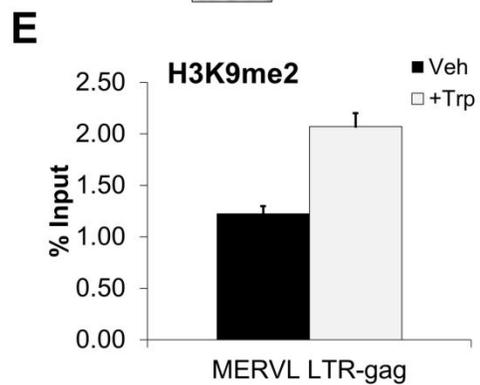
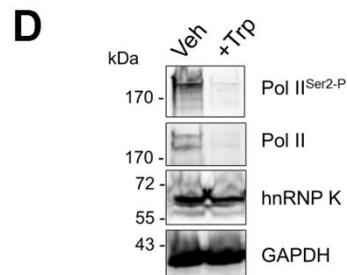
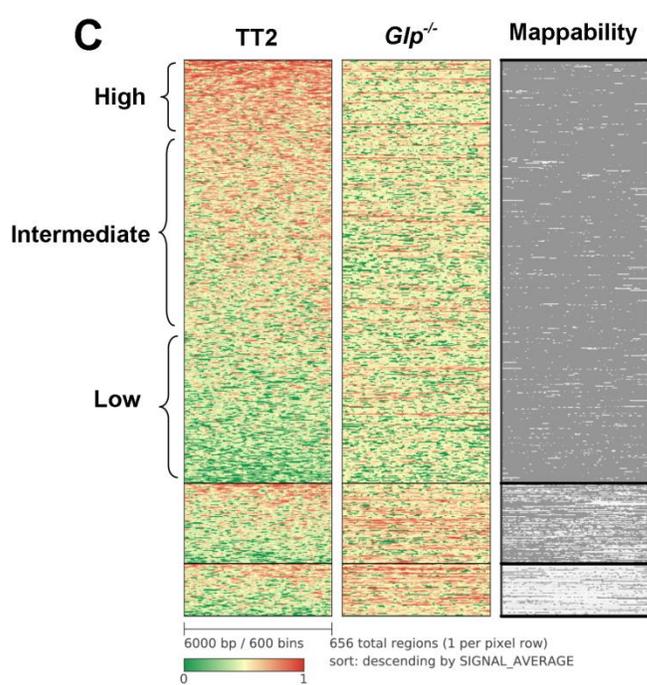
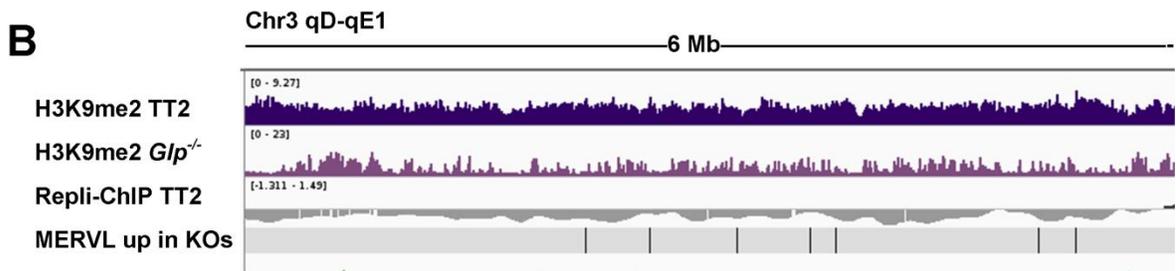
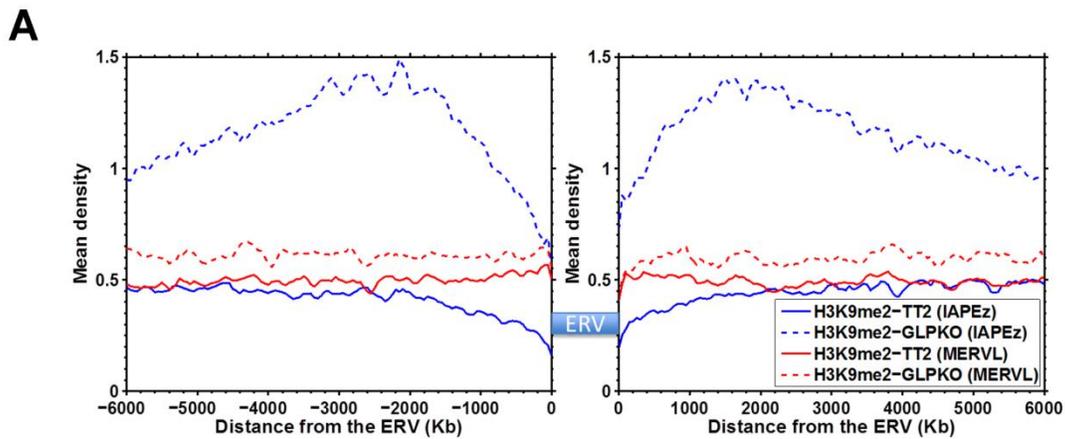


Figure 13. MERVL elements are marked by H3K9me2 as a consequence of genomic location and transcriptional inactivity. (A) H3K9me2 binned ChIP-seq coverage over the -6 kb upstream or +6 kb downstream of the flanks of MERVL or IAP elements in TT2 wt or *Glp*^{-/-} mESCs. Note: in the absence of GLP, global H3K9me2 is lost, which likely results in an erroneously high level of H3K9me2 ChIP signal at regions where H3K9me2 is converted to H3K9me3 (e.g. IAPs). (B) IGV screenshot with trackfiles for ChIP-seq of H3K9me2 in TT2 wt and *Glp* KO, annotation for RepliChIP showing early and late replicating regions in mESCs from ENCODE and annotations for MERVLs upregulated (≥ 2 -fold) in *G9a* and/or *Glp* KO mESCs. (C) Heatmap plot showing normalized H3K9me2 coverage binned over -6 kb upstream of 656 individual MERVL elements (each line represents a 6 kb region on the 5' flank a single element) in TT2 wt and *Glp*^{-/-} cells. Mappability is a measure of the uniqueness of the genomic sequence, with dark grey representing high mappability (unique) and white representing lower mappability (repetitive). (D) Western blot analysis of transcriptionally engaged Pol II^{Ser2-P} and total Pol II in TT2 mESCs 3.5 h after incubation in vehicle (DMSO) or Pol II inhibitor triptolide (Trp). HnRNP K was a nuclear loading control, GAPDH was a cytoplasmic loading control. (E) NChIP for H3K9me2 at MERVL 5'LTR-gag regions in vehicle or Trp-treated cells after 8.5 h. Data represent mean enrichment relative to input chromatin from n = 3 technical replicates, error bars are s.d.

4.3 hnRNP K promotes global H3K9me2 and represses MERVLs in H3K9me2-rich domains

RNA-seq analysis of hnRNP K KD mESCs revealed that MERVL LTRs and internal regions (annotated as MT2_Mm and MERVL-int, respectively) are the most highly de-repressed retrotransposon (~12-16-fold),²⁴⁵ suggesting a role for hnRNP K in transcriptional repression of MERVL. To determine whether hnRNP K has an effect on MERVL silencing via its binding at these ERVs and promoting H3K9me2, ChIP assays were conducted on KD cells (**Figure 14A**). HnRNP K showed a similar level of enrichment at MERVL 5'LTRs and *pol* regions as compared with IAP elements, which was decreased in the KD cells, confirming specificity (**Figure 14B**). Strikingly, depletion of hnRNP K led to dramatic reduction of H3K9me2 levels at many loci by 24 h post-transfection including at the positive control *Magea2* promoter, which is bound by G9a/GLP and enriched in H3K9me2 in mESCs,⁸⁰ in addition to MERVL LTRs and two different intergenic regions (**Figure 14C**). These results

suggested that depletion of hnRNP K generally reduces H3K9me2. To quantify the changes in total H3K9me2, western blot analysis was performed on hnRNP K KD cells in parallel with *G9a* and *Glp* KO cells (**Figure 14D**). While *G9a* and *Glp* KO cells had a ~75-80% reduction in total H3K9me2 as previously reported⁸⁰ hnRNP K KD cells showed a consistent ~25% reduction in H3K9me2 at 24 h post-transfection (**Figure 14E**). Importantly, there were examples of MERVLs located in H3K9me2-rich regions that were de-repressed upon KD of hnRNP K similarly to KO of *G9a* or *Glp* (**Figure 14F**), indicating that the loss of hnRNP K impacts the repression of H3K9me2-marked MERVL elements. Therefore hnRNP K helps to maintain global H3K9me2 and contributes to repression of MERVLs in H3K9me2-rich domains.

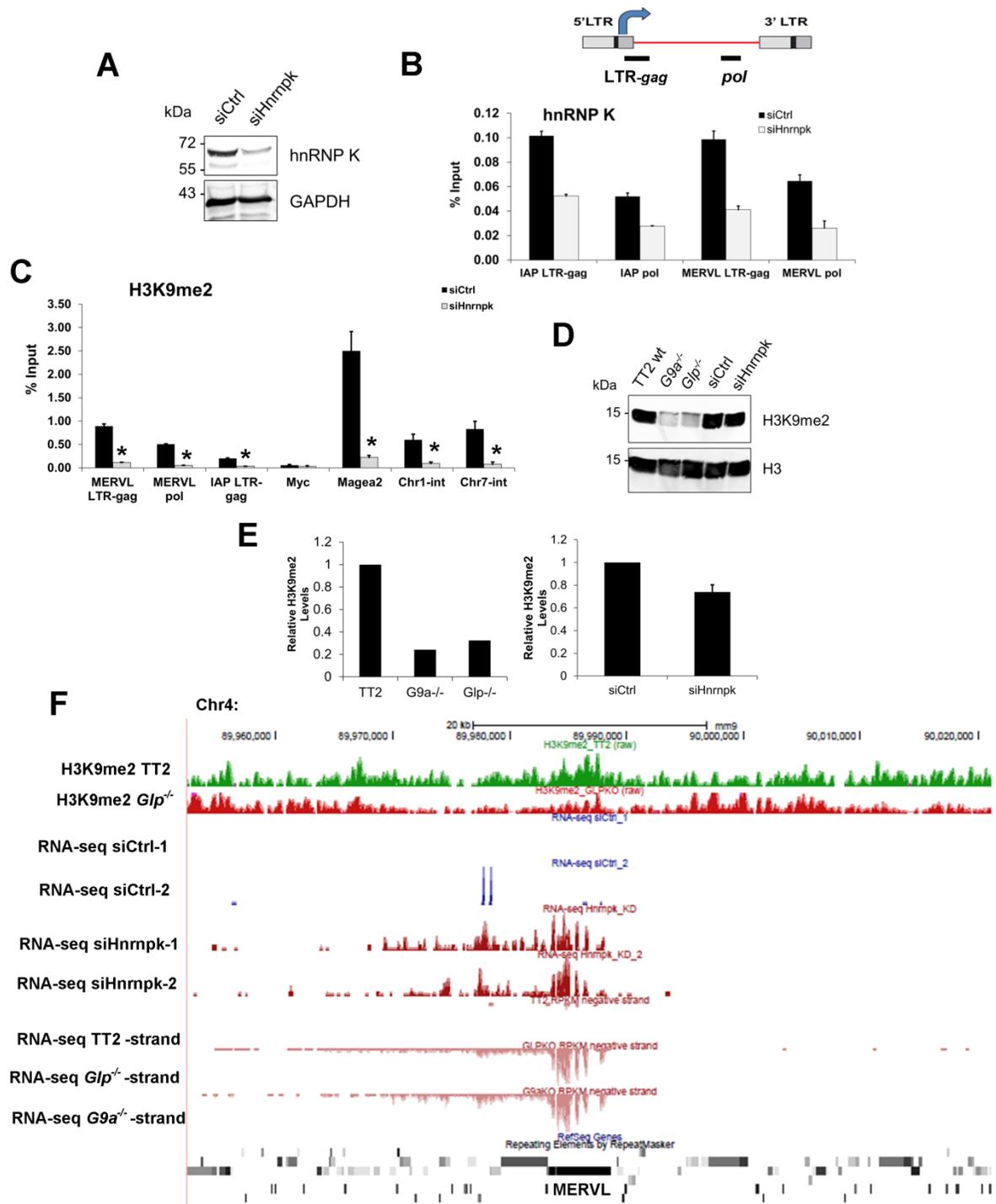


Figure 14. hnRNP K contributes to global H3K9me2 and promotes repression of H3K9me2-marked MERVL elements. (A) Western blot validation of hnRNP K siRNA KD in TT2 mESCs at 24 h post-transfection. GAPDH was a loading control. (B) ChIP of hnRNP K in the cells from (A). Schematic above shows the positions of amplicons for LTR-*gag* and *pol* regions on ERVs. (C) NChIP for H3K9me2 in the cells from (A). ChIP data are mean enrichment relative to input chromatin from n = 3 technical replicates, error bars are s.d. *p < 0.0001 two-tailed T-test. (D) Western blot analysis of H3K9me2 and total H3 on acid-extracted histones from the indicated mESC lines. (E) Quantification of H3K9me2 bands from the experiment in (D) for *G9a* and *Glp* KO cells relative to the TT2 wt control line, or *Hnrnpk* KD cells relative to the control siRNA KD. H3K9me2 band intensity was normalized to total H3 band intensity. Data for the hnRNP K KD cells is from two independent KD experiments, error bar is s.e.m. (G) UCSC genome browser trackfiles H3K9me2 ChIP-seq from TT2 and *Glp* KO cells, alongside RNA-seq from *Hnrnpk* KD cell replicates (see Chapter 3) and RNA-seq from TT2, *G9a* and *Glp* KO cells (generated in the Lorincz lab). Data for each wt/control and KD/KO pair are shown on equal scales and displayed over a region harbouring an intact full-length MERVL element on the antisense strand. Note: the TT2, *G9a*^{-/-} and *Glp*^{-/-} RNA-seq datasets are strand-specific and show the read coverage over the antisense strand, while the strand-specific *Hnrnpk* KD and control KD datasets have been collapsed into single coverage track files. All datasets were built from uniquely mapping reads.

4.4 hnRNP K is required for repression of a MERVL LTR reporter that lacks H3K9me2

To determine how newly integrated MERVL LTRs are repressed, a clonal mESC line stably transfected with a MERVL LTR reporter construct was derived by the Macfarlan lab, as described previously.^{85,321} This mESC line, termed 2C::*Gfp*, harbours a ~700 bp fragment containing a near consensus MERVL 5'LTR, PBS^{Leu} and a portion of the *gag* coding sequence cloned upstream of a *GFP-T2A-Puro* cDNA (**Figure 15A**) stably transfected into A2lox mESCs.²³⁹ In transient transfection assays this MERVL LTR reporter is not active in wt mESCs but is highly active in *Kdm1a*, *Kap1* and *G9a* KO mESCs, which also express 2-cell stage genes,⁸⁵ suggesting that the cellular milieu becomes permissive for LTR activity in the absence of these factors. However, it has not been determined whether the transcriptionally inactive state of the integrated reporter is dependent on H3K9me2.

To determine whether the integrated MERVL LTR construct can autonomously recruit H3K9me2, ChIP was performed on the *2C::Gfp* cell line using primers that amplify the unique *gag-Gfp* or *Gfp-T2A-Puro* sequences of this construct (**Figure 15A**). To preclude the use of MEF feeder cells in these experiments, this cell line was cultured in 2i media, which also leads to lower basal expression of MERVL,⁸⁵ probably as a consequence of a homogeneous cell population in ‘ground-state’ pluripotency.⁹³ Unexpectedly, ChIP analysis demonstrated that while endogenous MERVL elements had high levels of H3K9me2 enrichment in the *2C::Gfp* line, the integrated MERVL LTR reporter construct was devoid of H3K9me2 as judged by primers spanning the unique *gag-Gfp* and *Gfp-T2A* regions (**Figure 15B**). This indicates that despite its transcriptional inactivity, the integrated reporter construct is not marked by H3K9me2, nor is it integrated into an H3K9me2-rich domain.

To determine the requirement for hnRNP K in repression of such MERVL LTRs that lack H3K9me2, hnRNP K was depleted by siRNAs alongside KAP1 as a positive control.⁸⁶ Both proteins were depleted by 24 h post-transfection, as judged by western blot and the levels of KAP1, LSD1 or G9a were unaffected by hnRNP K KD (**Figure 15C**). ChIP in the control and hnRNP K KD *2C::Gfp* cells demonstrated that hnRNP K was specifically enriched at the *2C::Gfp* reporter at a level similar to that observed for IAP and endogenous MERVL ERVs (**Figure 15D**). Strikingly, flow cytometry analysis revealed a dramatic increase in the average proportion of live GFP⁺ cells upon depletion of either KAP1 or hnRNP K (**Figure 15E**) and qRT-PCR confirmed increased mRNA levels from the *2C::Gfp* reporter in both KDs (**Figure 15F**). Taken together, these data confirm that newly integrated MERVL LTRs can remain transcriptionally inactive even in the absence of H3K9me2 enrichment and reveal

that hnRNP K exerts a repressive effect on MERVL LTRs independently of the presence of H3K9me2.

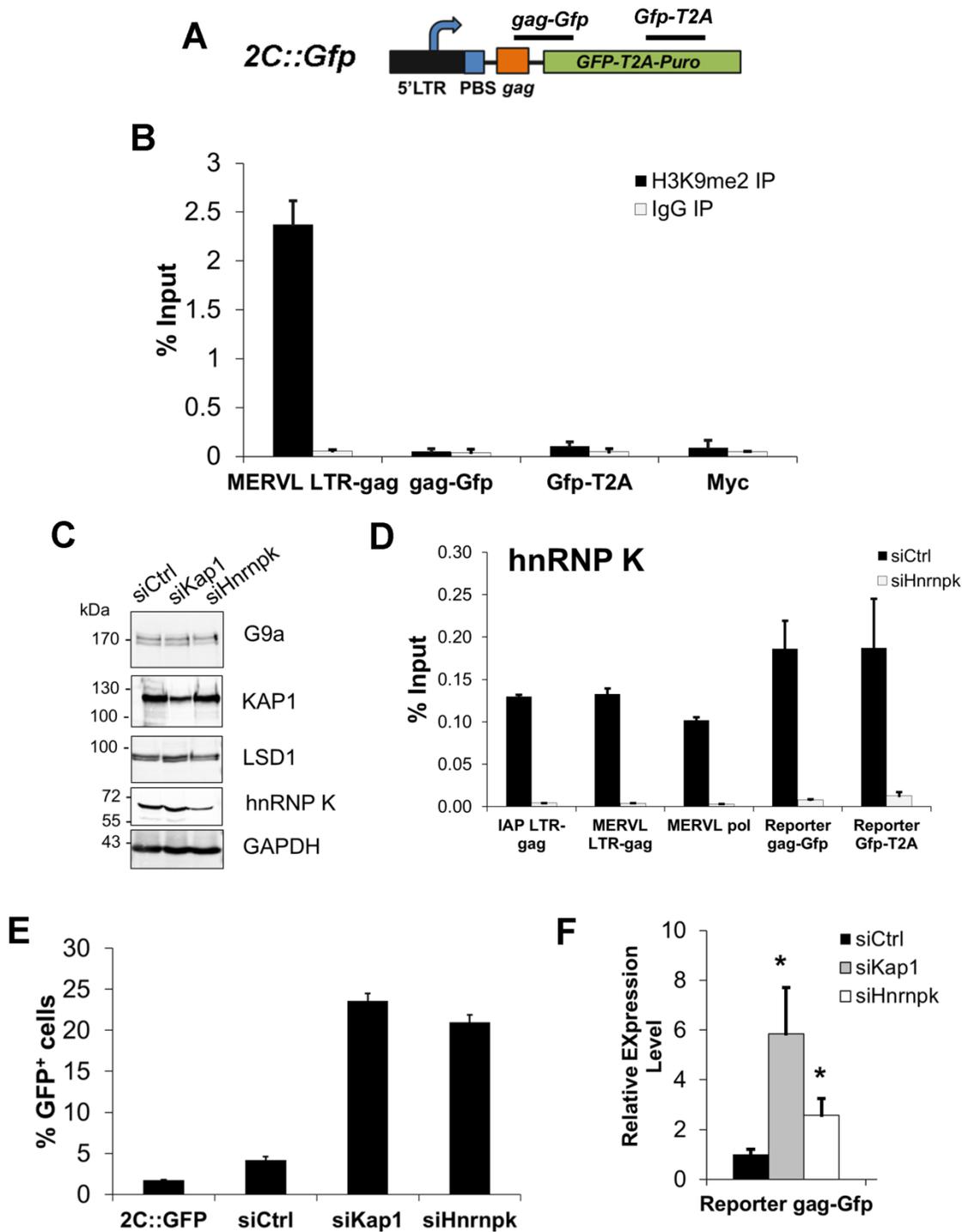


Figure 15. hnRNP K is required for repression of a newly integrated MERVL LTR reporter lacking H3K9me2. (A) Schematic of the *2C::Gfp* MERVL LTR reporter, showing positions of the unique *gag-Gfp* and *Gfp-T2A* amplicons. (B) NChIP for H3K9me2 in the *2C::Gfp* mESC line, at endogenous MERVL LTR-*gag* sequences and at the integrated *2C::Gfp* reporter construct. The *Myc* promoter was a negative control locus for H3K9me2 and mouse IgG was a negative control ChIP. (C) Western blot validation of KAP1 and hnRNP K KDs in *2C::Gfp* cells, also showing levels of G9a and LSD1 in the KDs. GAPDH was a loading control. (D) ChIP for hnRNP K in control or *Hnrnpk* KD *2C::Gfp* cells at endogenous IAP and MERVL 5'LTR-*gag* and *pol* regions and the *2C::Gfp* construct at 24 h post-transfection. (E) Flow cytometry analysis of GFP⁺ cells in the parental *2C::Gfp* line (untransfected) and cells transfected with indicated siRNAs at 72 h post-transfection. Data are mean percentage of GFP⁺ cells from n = 3 biological replicates, error bars are s.d. (E) qRT-PCR of mRNA from the *2C::Gfp* construct, amplifying the *gag-Gfp* region in *2C::Gfp* control siRNA or *Kap1* and *Hnrnpk* KDs at 72 h post-transfection. Data are mean fold-change relative to control siRNA from n = 3 technical replicates, error bars are s.d. *p < 0.01, two-tailed T-test. ChIP data are the mean enrichment relative to input from n = 3 technical replicates, error bars are s.d.

4.5 hnRNP K forms an RNA-dependent complex with G9a/GLP

The depletion of H3K9me2 upon hnRNP K KD (**Figures 14C and 14D**) suggested a functional interaction between hnRNP K and G9a/GLP. To determine whether hnRNP K physically interacts with G9a/GLP in mESCs, endogenous hnRNP K complexes were immunoprecipitated for western blot analyses (**Figure 16A**). KAP1, which directly interacts with hnRNP K in association with SETDB1²⁴⁵ was detected in the hnRNP K IP along with G9a and GLP, but LSD1 was not (**Figure 16A**). Reciprocally, hnRNP K co-precipitated with GLP and the interactions between hnRNP K and G9a/GLP were sensitive to phosphatase activity, since in the absence of phosphatase inhibitor an interaction was not detected (**Figure 16B**). Since hnRNP K interacts with several of its binding partners in RNA-dependent complexes^{141,306,334} and G9a/GLP also associates with lncRNAs,^{131,132} I tested whether this interaction was sensitive to RNase activity. While GLP co-precipitated with hnRNP K from mESC nuclear extracts in the absence of RNases, the presence of single-stranded RNases A

and T1 clearly weakened their interaction (**Figure 16C**), suggesting that intact RNA stabilizes this complex. Although G9a/GLP forms core complexes with DNA-binding zinc finger proteins Wiz and ZNF644,^{26,27} due to lack of antibodies that would recognize the mouse proteins, I could not determine whether hnRNP K is also present in a complex with Wiz and Zfp644 (the murine orthologue of ZNF644) (data not shown).

Consistent with the physical interactions between hnRNP K and G9a/GLP and the loss of H3K9me2 upon hnRNP K depletion, G9a enrichment was significantly decreased at the *Magea2* promoter and MERVL LTRs as early as 24 h post-transfection of *Hnrnpk* siRNAs (**Figure 16D**), indicating that hnRNP K is required for G9a/GLP recruitment to its chromatin targets. Surprisingly, hnRNP K enrichment at a variety of targets including MERVL and H3K9me3-marked loci IAP and *Mael*^{245,246} was significantly reduced in *G9a* KO cells (**Figure 16E**). The same results were observed within 5 days upon acute deletion of *G9a* in a conditional KO mESC line (data not shown), indicating that this is not a result of prolonged absence of G9a or accumulated secondary changes. Furthermore, depletion of hnRNP K or G9a/GLP had similar effects on the mESC transcriptome. Comparison of genes upregulated >2-fold by mRNA-seq analysis of hnRNP K KD cells²⁴⁵ and *G9a* KO or *Glp* KO cells (generated previously in the Lorincz Lab) revealed that a significant proportion (~49%, 130/264) of the hnRNP K-repressed genes were also upregulated upon loss of G9a and/or GLP (**Figure 16F**). This included many of the previously reported 2-cell stage-specific genes, some of which are driven by MERVL LTRs.^{86,245}

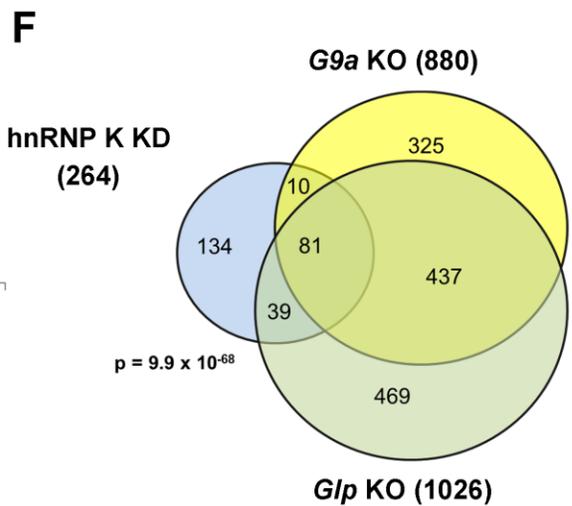
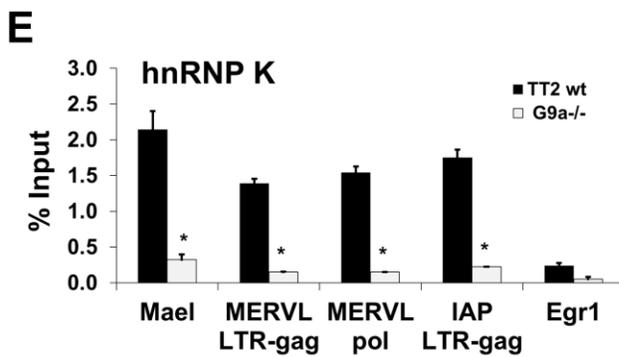
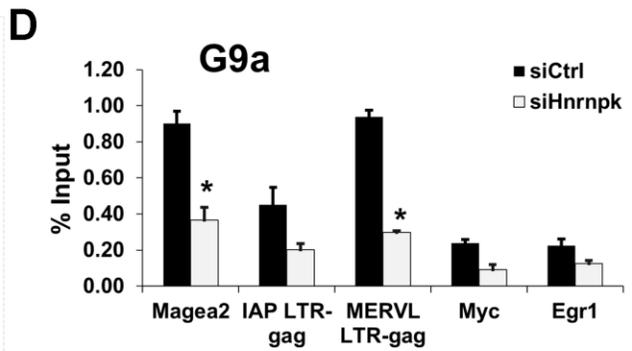
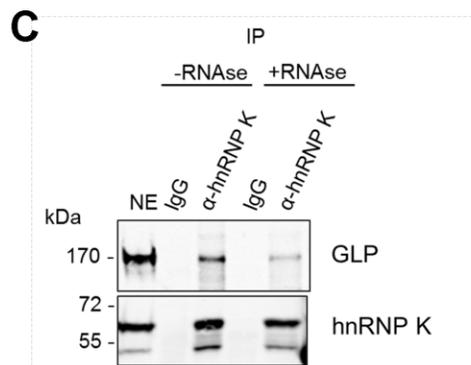
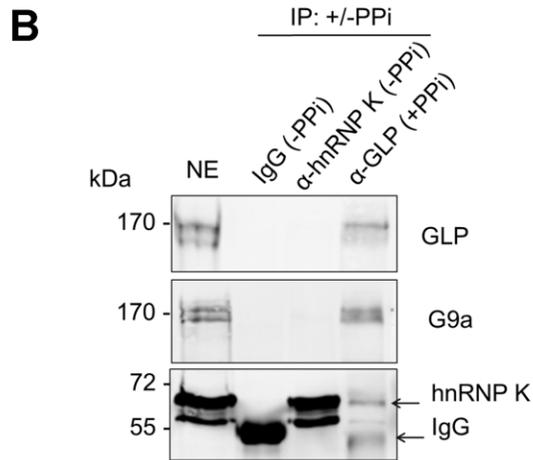
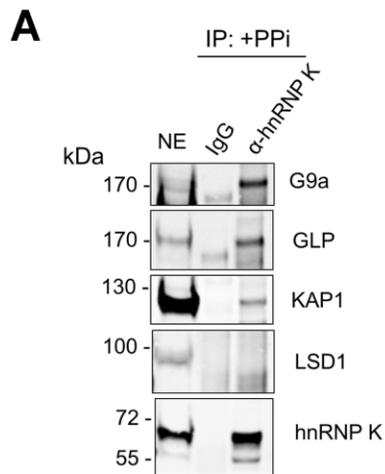


Figure 16. hnRNP K forms an RNA-dependent complex with G9a/GLP. (A) Co-IP assay and western blot for hnRNP K with G9a, GLP, KAP1 and LSD1 in TT2 mESC nuclear extracts. IP of hnRNP K complexes was performed where nuclear extract was prepared with phosphatase inhibitor (+PPi). In all IP experiments, IgG is a mouse IgG negative control IP. (B) Co-IP assay and western blot for GLP, G9a and hnRNP K. IP was performed for hnRNP K and GLP complexes from TT2 mESC nuclear extracts prepared with or without PPi, respectively. (C) Co-IP assay and western blot for hnRNP K with GLP where TT2 mESC nuclear extract was prepared in the absence or presence of single-stranded RNAses A and T1 (-/+RNase). (D) ChIP for G9a in TT2 cells transfected with control or *Hnrnpk* siRNAs at 24 h post-transfection. (E) ChIP for hnRNP K in TT2 wt or *G9a* KO mESCs. For both ChIP-qPCRs, data are mean enrichment relative to input chromatin from n = 3 technical replicates, error bars are s.d. *p < 0.0001 two-tailed T-test. (F) Venn diagram of the overlap between genes upregulated in *Hnrnpk* KD (264), *G9a* KO (880) and *Glp* KO (1026). p = 9.9 x 10⁻⁶⁸, Fisher's exact test using n = 22,138 ENSEMBL annotated genes.

To rule out effects on hnRNP K protein levels and localization in *G9a* KO cells, immunofluorescence staining and western blot analyses were performed. Notably, the loss of G9a did not substantially alter hnRNP K protein levels (**Figure 17A**) and similarly, immunostaining confirmed that hnRNP K remains exclusively localized to the nucleus in *G9a*-deficient cells (**Figure 17B**). Since hnRNP K is required for SETDB1 recruitment to class I and II ERVs,²⁴⁵ I determined whether SETDB1 binding at ERVs was altered upon deletion of *G9a*. Strikingly, SETDB1 enrichment was dramatically increased in *G9a*^{-/-} cells compared with TT2 wt at IAP and ETn/MusD 5'LTRs, but not at the germline gene *Mael* (**Figure 17C**) indicating that SETDB1 binding at these ERVs is antagonized by G9a. Taken together these data demonstrate a complex regulatory relationship between hnRNP K and the G9a/GLP complex in mESCs.

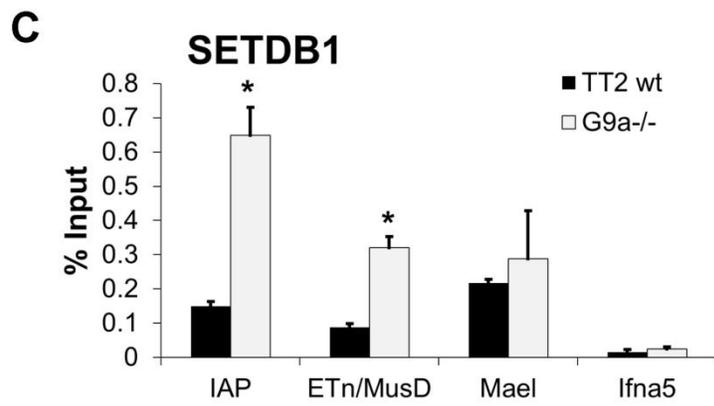
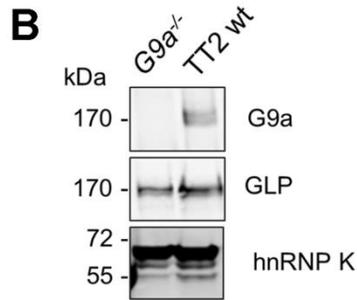
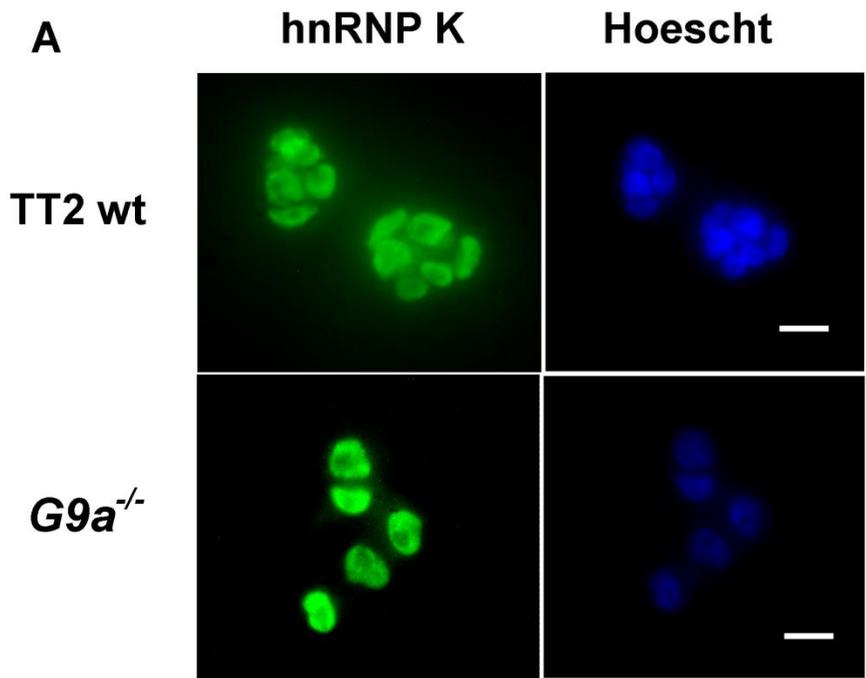


Figure 17. *G9a* KO cells do not show changes in hnRNP K localization or protein levels, but have increased SETDB1 binding at ERVs. (A) Immunofluorescence staining to detect hnRNP K in TT2 and *G9a* KO cells. Hoescht was used to counterstain nuclei. Scale bar is 30 μm . (B) Western blot analysis of hnRNP K in TT2 and *G9a*^{-/-} cells. GLP was a loading control. (C) ChIP for SETDB1 in TT2 and *G9a*^{-/-} mESCs. The *Ifna5* promoter was a negative control. Data are mean enrichment relative to input from n = 3 replicates, error bars are s.d. *p < 0.0001, two-tailed T-test.

4.6 hnRNP K and G9a/GLP are associated with nuclear MERVL transcripts in mESCs

In order to identify hnRNP K-G9a/GLP-associated RNAs, I took a candidate approach. I hypothesized that hnRNP K-G9a/GLP complexes may bind to MERVL-derived RNAs, which may be involved in targeting genomic regions harbouring MERVL sequences.

To determine whether MERVL transcripts are present in the nucleus in wt mESCs, RT-PCR assays were performed on total, nuclear and cytoplasmic RNA. Efficient cell fractionation was confirmed in that 45S pre-rRNA was detected in only in the nuclear fraction while 28S rRNA was detected in both nuclear and cytoplasmic fractions (**Figure 18A**). Interestingly, MERVL LTR-*gag*-spanning transcripts were detected in both the nuclear and cytoplasmic compartments (**Figure 18A**). The presence of cytoplasmic MERVL mRNA is consistent with the previous finding that a small proportion of mESCs and cells of the inner cell mass in the blastocyst cycle into a 2-cell-like state and express MERVL Gag protein.⁸⁵ Using random 15-mers, oligo d(T), sense or antisense LTR-*gag* primers to prime cDNA synthesis from nuclear RNA, MERVL transcripts were detected with each RT primer (**Figure 16B**) indicating that nuclear MERVL transcripts occur in both the sense and antisense orientations and are polyadenylated, consistent with previous RT-PCR assays on preimplantation embryos.^{319,335} Next, to test whether hnRNP K interacts with these nuclear MERVL-derived RNAs, native

RNA immunoprecipitation (RIP) was performed. HnRNP K complexes were immunoprecipitated from mESC nuclear extracts under native RNase-protected conditions (**Figure 18C**) and co-precipitated RNAs were analyzed by qRT-PCR. *Egr1* transcripts, which are bound by hnRNP K in HCT116 cells upon serum stimulation^{296,297} and are downregulated upon hnRNP K KD in mESCs²⁴⁵ were enriched in the anti-hnRNP K RIP as expected, but *Gapdh* transcripts were ~8-fold less enriched (**Figure 18C**) confirming the specificity of the anti-hnRNP K RIP. Strikingly, MERVL LTR-*gag* RNA was enriched in the anti-hnRNP K RIP to a similar extent as *Egr1* mRNA. In contrast, the mouse IgG RIP failed to enrich any of the RNAs tested (**Figure 18C**), further confirming anti-hnRNP K RIP specificity. Strand-specific RT-PCR demonstrated the presence of both sense and antisense MERVL transcripts in the RIP with hnRNP K antibodies, but not with mouse IgG (**Figure 18D**). Furthermore, RIP with GLP antibodies under native conditions also revealed the enrichment of nuclear MERVL LTR-*gag* transcripts relative to the IgG RIP (**Figure 18E**) indicating binding of MERVL RNA by G9a/GLP.

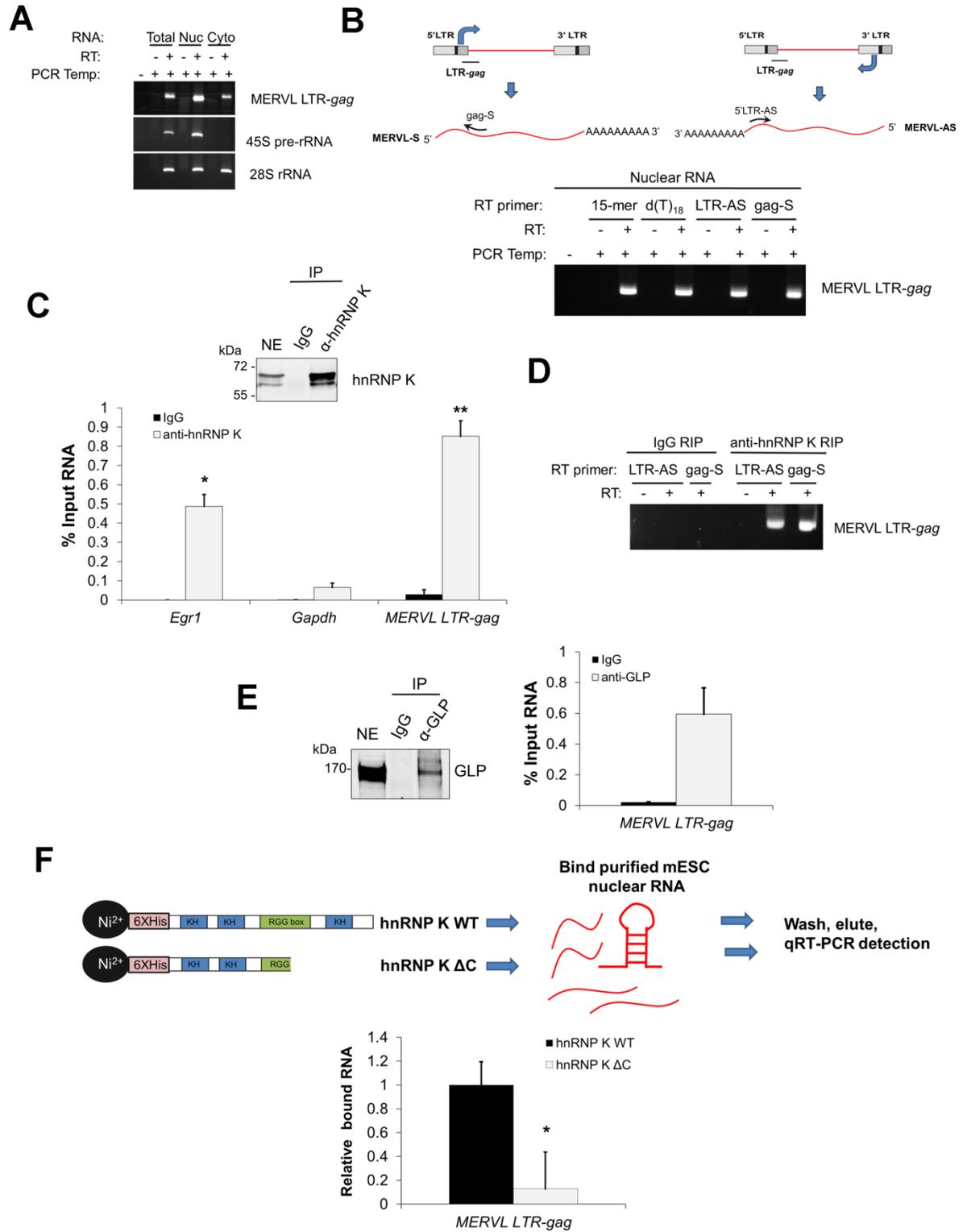


Figure 18. hnRNP K and G9a/GLP associate with nuclear MERVL transcripts. (A) RT-PCR assay on total, nuclear or cytoplasmic RNA for the indicated transcripts. 45S pre-rRNA is the unspliced precursor transcript while 28S rRNA is amplified from both the mature species in the cytoplasm and the precursor in the nucleolus. For all RT-PCR assays, RT indicated presence or absence of reverse transcriptase (+ or - respectively), while PCR temp indicated presence or absence of template cDNA. (B) Strand-specific RT-PCR assay from nuclear RNA, using random 15-mers, oligo dT (18Ts), or MERVL LTR-gag antisense (AS) or gag sense (S) primers to prime reverse transcription. (C) hnRNP K RIP assay, where hnRNP K antibodies or mouse IgG was used for the IPs. Upper panel: western blot validation of hnRNP K IP. Lower panel: qPCR of indicated RNA species in the mouse IgG or hnRNP K RIPs. For all RIPs, data are mean enrichment relative to input RNA from n = 3 technical replicates, error bars are s.d. *p < 0.001, **p < 0.0001 two-tailed T-test, relative to the *Gapdh* negative control in the hnRNP K RIP. (D) Strand-specific RT-PCR on mouse IgG or hnRNP K RIPs, using primers to amplify sense or antisense MERVL LTR-gag-spanning transcripts. (E) GLP RIP assay, where GLP antibodies or mouse IgG were used for the RIPs. Left panel: western blot validation of the IP. Right panel: qRT-PCR of MERVL LTR-gag transcripts in the IPs. Data are mean enrichment relative to input RNA from n = 3 technical replicates, error bars are s.d. Mouse IgG was a negative control for RIP. (F) Upper panel: Schematic of the nuclear RNA pulldown assay with recombinant hnRNP K proteins. wt or a C-terminal deletion mutant (hnRNP K Δ C) lacking the third KH domain were used as baits in a pulldown of purified nuclear RNAs. Lower panel: qPCR of MERVL LTR-gag RNA associated with the hnRNP K Δ C mutant relative to the wt. Data are mean enrichment from input RNA of n = 3 technical replicates, and normalized to the wt hnRNP K pulldown. Error bars are s.d. *p < 0.01, two-tailed T-test.

The KH3 domain of hnRNP K is a well-characterized nucleic acid recognition module that binds with high affinity to C-rich RNA and DNA sequences^{336,337} and contributes to the functions of hnRNP K in developmental regulation of mRNA translation^{244,338} and class switch recombination.³³⁴ To determine whether hnRNP K directly binds MERVL transcripts in a KH3 domain-dependent manner, an *in vitro* nuclear RNA pulldown assay was performed using recombinant 6X-His-tagged wt hnRNP K or a C-terminal deletion mutant that lacks the KH3 domain (hnRNPK Δ -C, **Figure 18F**). This assay showed that the hnRNPK Δ -C mutant was dramatically impaired in binding MERVL LTR-gag transcripts as compared with wt (**Figure 18F**). However, EMSA confirmed that the hnRNP K Δ -C mutant binds to the DNA analogue of the *LOX* 3'UTR differentiation control element (DICE)²⁴⁴ even more efficiently

than wt hnRNP K (**Figure 19**), indicating that this mutant is not generally deficient in nucleic acid binding, consistent with previous results.²⁸¹ Taken together these data demonstrate direct binding of hnRNP K to nuclear sense and antisense MERVL-derived RNAs in a KH3 domain dependent manner *in vitro* and in mESCs and further reveal association of the G9a/GLP complex with nuclear MERVL transcripts in mESCs.

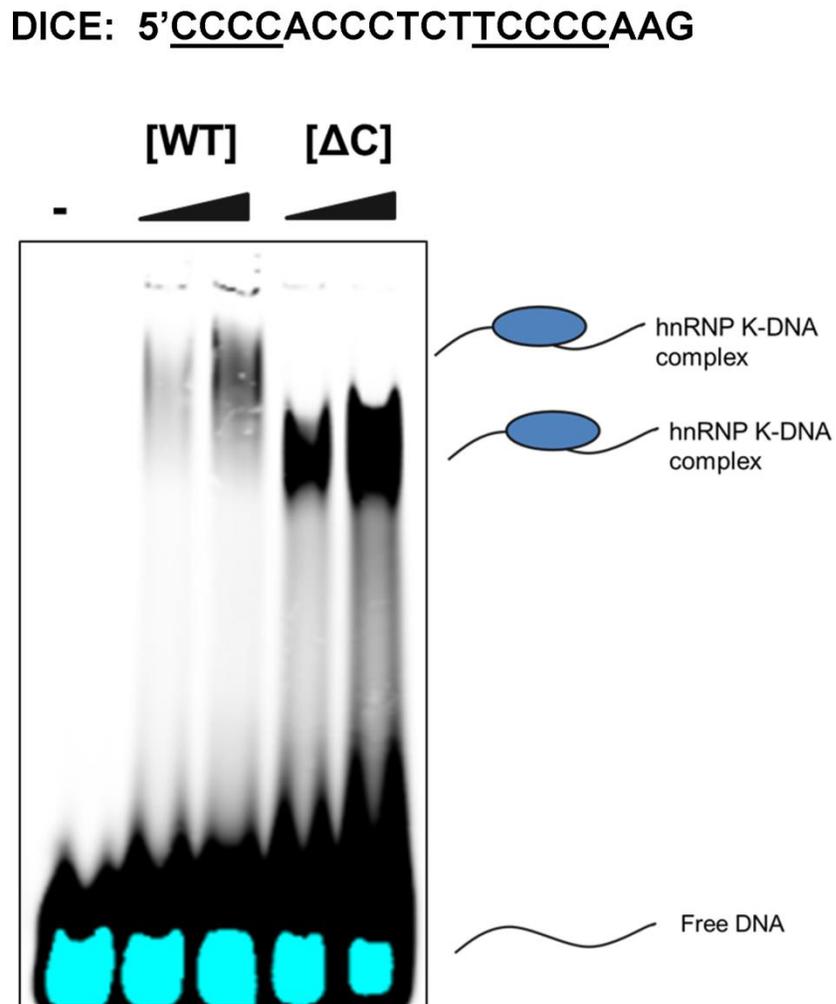


Figure 19. Comparison of the nucleic acid binding activities of wt and Δ -C mutant hnRNP K proteins. EMSA was performed using the wt or Δ -C mutant hnRNP K, as in Figure 18F, using the DNA analogue of the DICE element (shown above). Underlined regions indicate C-rich regions bound by hnRNP K.

4.7 G400R point mutation of the nucleic acid binding KH3 domain of hnRNP K perturbs MERVL silencing

To determine whether DNA/RNA binding by the KH3 domain of hnRNP K is necessary for maintaining repression of MERVL in mESCs, a dominant-negative approach was employed, since hnRNP K can form homodimers when bound to nucleic acids.^{337,339} To this end, I overexpressed the previously characterized KH3 domain point mutants G400R and Y458D, which are significantly impaired in their ability to bind nucleic acids *in vitro*^{338,340} (**Figure 20A**). For comparison, I also generated a deletion mutant lacking the RGG box/KI domain (Δ -RGG a.a. 240-338, **Figure 20A**), which is required for hnRNP K homodimeric interactions and protein-protein interactions³³⁹ but leaves its KH3 domain intact.

Although stable transfection was preferred to transient transfection with these constructs, after several attempts I found that stable overexpression of wt hnRNP K was not tolerated by mESCs, as the ectopic hnRNP K could not be detected by western blot in several independent antibiotic-resistant clones that were recovered (data not shown). Therefore, the N-terminal T7-tagged wt and mutant hnRNP K constructs were transiently transfected into TT2 wt mESCs and the exogenous and endogenous total hnRNP K were monitored by western blot (**Figure 20B**). Notably, while wt hnRNP K and the G400R mutant were stable, the Y458D mutant was not detected (**Figure 20B**) suggesting that this mutation, which mimics constitutive phosphorylation by the c-Src kinase,³³⁸ is unstable under steady-state conditions in mESCs. Similarly, the Δ -RGG mutant was also consistently expressed at a lower level compared with the wt and G400R mutant (**Figure 20B**). Due to these differences in expression, further analyses were focused on the wt and G400R mutant. In independent

experiments, expression of the wt and G400R mutant hnRNP K proteins were equivalent (**Figure 20C**).

ERV transcripts were quantified in the transfected cells by qRT-PCR (**Figure 20D**). While overexpression of the wt hnRNP K had no effect on MERVL expression, the G400R mutant induced de-repression of MERVL by ~3-fold (**Figure 20D**). In contrast, IAP elements were not similarly de-repressed in the hnRNP K G400R mutant-expressing cells (**Figure 20D**) indicating that IAP silencing is not sensitive to this mutation. To confirm that the G400R point mutant specifically interferes with endogenous hnRNP K in G9a/GLP-dependent MERVL silencing, the wt hnRNP K or G400R mutant were overexpressed in *G9a* KO cells (**Figure 20E**), where endogenous hnRNP K is no longer bound at MERVL (**Figure 16E**) and MERVL elements are already de-repressed. Consistent with previous results,⁸⁶ MERVL expression was upregulated ~5-fold in the mock-transfected *G9a* KO cells relative to TT2 wt (**Figure 20F**). However, the wt and G400R mutant hnRNP K proteins failed to significantly enhance or suppress the MERVL de-repression phenotype in *G9a* KO cells (**Figure 20F**), indicating that the MERVLs de-repressed upon loss of G9a cannot be further influenced by overexpression of wt hnRNP K or the G400R mutant. Taken together these results demonstrate that hnRNP K contributes to G9a/GLP-dependent repression of MERVL in a manner dependent on nucleic acid binding activity by the KH3 domain.

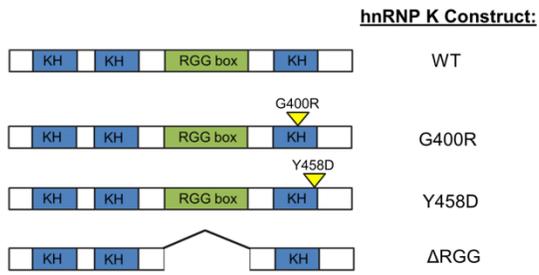
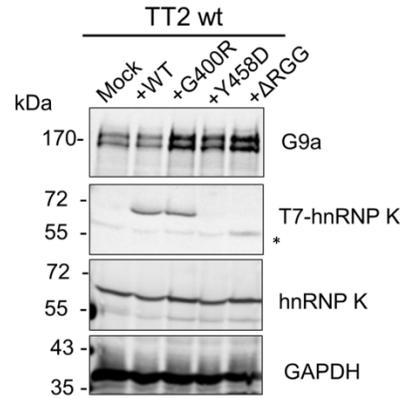
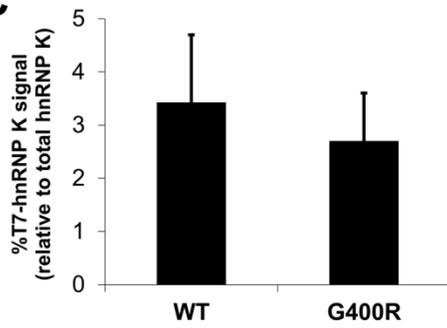
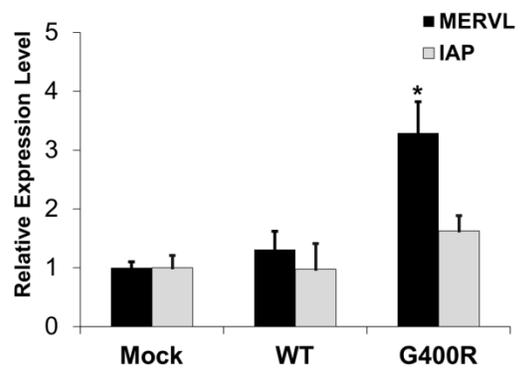
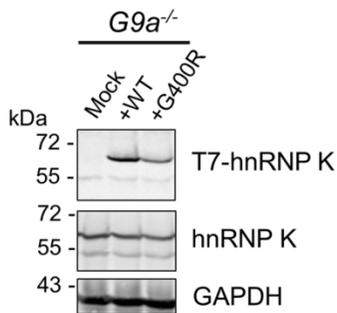
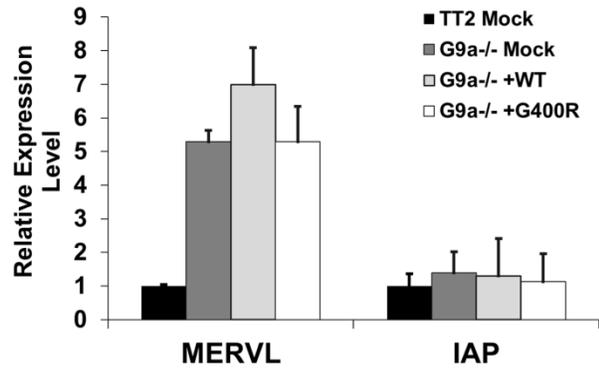
A**B****C****D****E****F**

Figure 20. G400R point mutation of the nucleic acid binding KH3 domain of hnRNP K perturbs G9a-dependent MERVL silencing. (A) Schematic of hnRNP K mutagenesis. (B) Western blot analysis of T7-tagged hnRNP K wt or indicated mutants transiently transfected into TT2 mESCs at 24 h post-transfection. Endogenous/total hnRNP K and G9a were also detected, GAPDH was a loading control. Asterisk indicates a non-specific band detected in the untransfected line. (C) Quantification of T7-hnRNP K protein levels upon transient transfection into TT2 cells. Data are mean signal for the T7-hnRNP K band, normalized to GAPDH from two independent experiments, error bars are s.e.m. (D) qRT-PCR analysis of IAP or MERVL LTR-gag amplicons in untransfected TT2 cells (Mock) or cells transfected with the wt or G400R mutant at 72 h post-transfection. Data are mean fold-changes relative to the untransfected line from n = 3 technical replicates, error bars are s.d. *p < 0.001 two-tailed T-test. (E) Western blot analysis of T7-hnRNP K expression upon transient transfection into *G9a*^{-/-} mESCs. Endogenous/total hnRNP K was also detected and GAPDH was a loading control. (F) qRT-PCR analysis of MERVL and IAP LTR-gag regions as in (D) except on TT2 or *G9a*^{-/-} cells transfected with the wt or G400R hnRNP K constructs.

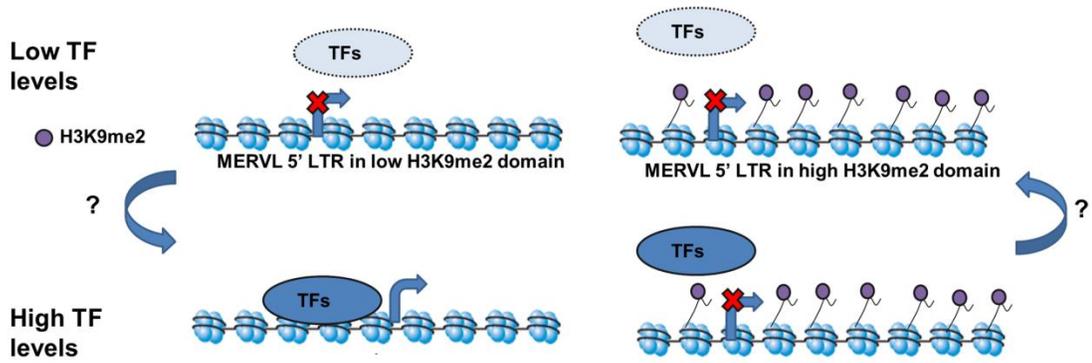
4.8 Discussion

While MERVL LTRs have been co-opted for regulation of a cohort of 2-cell stage genes,^{85,86,319,321} these ERVs are repressed prior to the first major differentiation in the embryo from totipotency to pluripotency at the late morula stage.³²³ Curiously, the repressive mechanisms acting on MERVL in mESCs are very different than those at young class I and II ERVs. This is evidenced by the observations that ~0.5-1% of mESCs in a population naturally cycle in and out of a 2-cell-like state where MERVL is expressed⁸⁵ and that MERVL repression is sensitive to depletion of several different chromatin modifiers^{86,321,324,326} but not to the loss of DNA methylation.^{246,321} In addition, unlike young class I and II ERVs, MERVL silencing may require removal of histone modifications associated with transcriptional activation.³²¹ The evolutionary basis for the distinct epigenetic mechanisms acting on MERVL ERVs are unclear, but given their co-opted regulatory role at the 2-cell stage, KRAB-ZFP recognition and associated robust SETDB1/H3K9me3-dependent silencing may not provide a selective advantage.⁸⁶

It remained unclear whether intact MERVL sequences can autonomously recruit H3K9me2 and more importantly, whether the H3K9me2 mark is always necessary for MERVL repression in mESCs. This study provides the first evidence that MERVLs are transcriptionally inactive whether or not they are marked by H3K9me2 and that these elements do not show evidence of autonomously recruiting this mark, but rather their H3K9me2 enrichment is likely to be a general consequence of their integration in H3K9me2-rich genomic domains. Furthermore, since H3K9me2 can be acquired on MERVL sequences upon Pol II inhibition, similar to what has been shown for ectopic targeting of PRC2 and H3K27me3 upon treatment of mESCs with Pol II inhibitors,⁵⁹ it is probable that some MERVLs are marked by H3K9me2 as a consequence of their transcriptional inactivity. These findings together support a model in which MERVL repression during embryogenesis may be defined by the absence of transcriptional activators, such as 2-cell stage TFs and the low levels or absence of such activators accounts for their transcriptional inactivity (**Figure 21A**). Transcriptional inactivity may or may not result in the acquisition of H3K9me2 depending on genomic context (**Figure 21A**). Once marked, H3K9me2 forms a repressive barrier that must be removed for such elements to be re-expressed, explaining the requirement of G9a/GLP in MERVL silencing for those elements marked by H3K9me2 (**Figure 21A**). However, in the case of MERVLs that are not located in H3K9me2-rich domains, these elements may be induced as a consequence of de-repression of 2-cell stage TFs that directly promote MERVL LTR transcriptional activity (**Figure 21A**). In support of this idea, a similar cohort of 2-cell stage genes are induced upon loss of KAP1, HP1 α , HP1 β , LSD1, RYBP, G9a/GLP and hnRNP K and includes TFs from the *Zscan4* gene family among others, which are not driven by MERVL LTRs.^{85,86,245,321,326} Indeed, it was recently

shown that *Zscan4* expression is transient in mESC populations and *Zscan4*-expressing cells (~1-5% of cells in culture) exhibit transcriptional activation of normally inactive 2-cell stage genes concomitant with globally increased histone acetylation.³⁴¹ Therefore, this model predicts that the MERVLs that are expressed in a cyclical fashion in a small percentage of mESCs in the population and are not marked by H3K9me2 might be driven by MERVL activator TFs (**Figure 21A**), which could include *Zscan4* proteins. Future studies should test this hypothesis by determining whether ectopic expression of *Zscan4* induces MERVL expression in mESCs and the expression of such TFs during the cell cycle and in serum versus 2i conditions could then be investigated to provide a correlation with cyclical expression of MERVL.

A



B

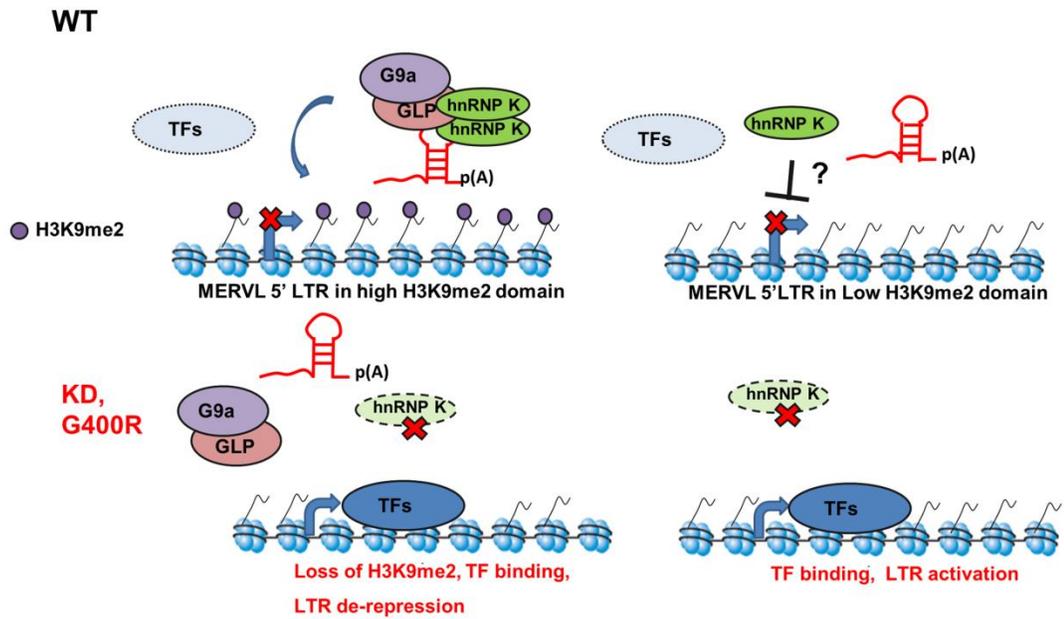


Figure 21. Revised model for the mechanisms governing MERVL silencing in mESCs.

(A) Role of 2-cell stage TF activators of MERVL in the periodic expression of MERVL elements. In wt mESCs, MERVL elements can be classified as those in large H3K9me2-rich domains and those in regions with low H3K9me2, both of which are normally transcriptionally inactive in most cells. Low levels of 2-cell stage TFs (e.g. Zscan4 proteins) precludes expression from MERVL LTRs in low H3K9me2 regions, but the levels of these TFs could increase in a small percentage of mESCs in the population, permitting expression from this subset of MERVLs. In contrast, MERVL LTRs located in H3K9me2-rich regions are protected from activation by these 2-cell stage TFs. The means by which these states may cycle back and forth is unknown. (B) Role of hnRNP K and G9a/GLP in MERVL repression. In wt mESCs, hnRNP K forms an RNA-dependent complex with G9a/GLP and promotes maintenance of H3K9me2-rich domains over MERVL elements. G9a/GLP is ultimately required for repression of MERVLs located in broad H3K9me2-rich domains. HnRNP K also maintains transcriptional inactivity at MERVL LTRs that lack H3K9me2, possibly by binding to their LTR sequences. In hnRNP K KD cells, there is a reduction of H3K9me2 due to reduced G9a/GLP targeting to chromatin, concomitant with a shift towards expression of MERVL TF activators, leading to high level MERVL expression. Overexpression of the G400R hnRNP K mutant that abrogates nucleic acid binding activity induces a dominant-negative effect and leads to de-repression of MERVLs that are repressed by H3K9me2-rich domains and G9a/GLP, and possibly also those in low H3K9me2 regions.

G9a/GLP form a core complex with Wiz in mESCs²⁷ and Wiz and ZNF644 in 293T cells.²⁶

Here I demonstrated the presence of a novel complex between hnRNP K and G9a/GLP in mESCs and found that hnRNP K was required for a proportion of total H3K9me2 and G9a recruitment to its targets, suggesting that it plays a general role in the recruitment and function of G9a/GLP. Unlike the interactions between hnRNP K and SETDB1/KAP1,²⁴⁵ the interactions between hnRNP K and G9a/GLP were sensitive to RNase digestion pointing to a role for intact RNA in assembly of this complex. Consistent with this, nuclear MERVL-derived sense and antisense RNAs are associated with both hnRNP K and G9a/GLP, suggesting a possible role for RNA in chromatin recruitment of this complex. The reason as to why hnRNP K is required for G9a/GLP recruitment to some of its chromatin targets is unclear, but it may serve as an adaptor protein to link these KMTs with RNA at particular genomic loci. It is also probable that hnRNP K and G9a/GLP associate with many other

nuclear RNAs that contribute to G9a/GLP chromatin recruitment and global H3K9me2.

These could be identified by next-generation sequencing of hnRNP K and G9a/GLP-bound RNAs (RIP-seq). It remains to be determined whether Wiz and/or Zfp644 are in any way associated with hnRNP K in this RNA-dependent G9a/GLP complex.

Unexpectedly, G9a was required for hnRNP K binding to all target loci tested, including those marked by H3K9me3/SETDB1, pointing to an integral role for this KMT in hnRNP K binding to chromatin in mESCs. Although hnRNP K is required for SETDB1 recruitment to ERVs in wt mESCs²⁴⁵ in *G9a*^{-/-} mESCs where hnRNP K enrichment at ERVs is dramatically reduced, for reasons that are unclear, SETDB1 binding at ERVs is increased. This may indicate a possible antagonism between G9a and SETDB1 binding at ERV chromatin. Indeed, G9a is also bound at IAP elements⁸⁶ and is required for maintaining their DNA methylation, however loss of G9a does not affect H3K9me3 or IAP silencing⁸⁴ consistent with the results reported here. In addition, deletion of *Setdb1* in mESCs leads to increased levels of H3K9me2 concomitant with loss of H3K9me3 on proviral chromatin,⁸⁹ consistent with a scenario in which G9a binding to ERVs may increase in *Setdb1* null mESCs. Based on these observations it is tempting to speculate that in *G9a* null cells where SETDB1 binding at ERVs is apparently enhanced, the requirement of hnRNP K in recruiting SETDB1 to ERVs may cease. A key test for this hypothesis would be to determine whether deletion of *G9a* suppresses the effect of hnRNP K depletion on SETDB1 recruitment and H3K9me3 at ERVs. If hnRNP K promotes SETDB1 recruitment to ERVs via influence on chromatin-bound protein SUMOylation (as predicted in the model from Chapter 3), then this would also implicate G9a in this pathway.

The molecular basis for the loss of hnRNP K enrichment on its chromatin targets in *G9a* null cells is unclear, however a recent proteomic study demonstrated that hnRNP K is monomethylated at K139 in 293T cells.³⁴² Therefore, hnRNP K could be methylated by G9a, which in turn may regulate its chromatin occupancy or binding stability, either directly or indirectly via crosstalk with other post-translational modifications and methyl-lysine binding proteins. Further experiments to determine whether K139 is indeed methylated by G9a/GLP and whether mutation of K139 affects hnRNP K chromatin binding in cells are clearly warranted.

Surprisingly, despite their physical interactions, the depletion of hnRNP K led to de-repression of MERVL elements enriched and devoid of H3K9me2, suggesting that hnRNP K can influence MERVL repression in the presence or absence of G9a/GLP activity. Similarly, the loss of G9a or GLP is likely to influence expression of MERVLs located in H3K9me2-rich domains and possibly those outside such domains, the latter indirectly due to the induction of 2-cell stage-specific TFs^{85,86} such as *Zscan4*. Taken together these data are consistent with a model in which hnRNP K and different RNA species might recruit the G9a/GLP complex to its chromatin targets, including transcriptionally inactive domains harbouring MERVL elements to deposit H3K9me2 (**Figure 21B**). Indeed, sense and/or antisense MERVL transcripts could promote the broad spreading of hnRNP K and G9a/GLP in *cis* or *trans*, based on complementarity to other elements, to replenish H3K9me2 during DNA replication, which maintains a repressive barrier to transcription (**Figure 21B**). Upon hnRNP K depletion, some aspect of G9a/GLP recruitment to chromatin is compromised

leading to reduction of H3K9me2 and eventual transcriptional de-repression of genes and MERVL LTRs located in these H3K9me2-rich domains (**Figure 21B**). Notably, overexpression of wt hnRNP K in wt or *G9a*^{-/-} mESCs failed to either enhance MERVL repression or suppress the *G9a* null phenotype suggesting that hnRNP K is necessary but not sufficient to promote MERVL silencing and acts upstream of G9a/GLP. The role of hnRNP K in promoting MERVL repression is likely dependent on its binding to RNAs (**Figure 21B**) since hnRNP K can directly bind nuclear MERVL RNA in a KH3 domain dependent manner and overexpression of a KH3 domain DNA/RNA-binding mutant hnRNP K perturbed MERVL repression. However, since hnRNP K binds to single-stranded C-rich DNA in addition to RNA, the possibility that hnRNP K also binds to MERVL and other DNA sequences to direct G9a/GLP recruitment cannot be excluded. Indeed the G400R point mutation abolishes both RNA and DNA binding activity by the KH3 domain of hnRNP K^{336,337,340} and therefore, MERVLs that harbour or lack H3K9me2 enrichment may be de-repressed upon overexpression of this mutant, due to loss of hnRNP K binding to DNA (**Figure 21B**). In parallel or upstream of this complex, the LSD1/CoREST complex may counteract active chromatin modifications at MERVL including H3K4 methylation and histone acetylation,³²¹ which may facilitate subsequent G9a/GLP activity by providing an H3 tail epitope towards which it is highly active.¹⁸

While this work was in progress, Ishiuchi and colleagues³⁴³ showed that the S-phase histone H3/H4 chaperone complex CAF-1 is required for MERVL silencing. This study demonstrated that MERVLs, along with a large proportion of 2-cell stage-specific genes, but not IAP or L1 are dramatically induced upon loss of CAF-1 nucleosome assembly activity in

mESCs, increasing the number of mESCs showing characteristics of a totipotent 2-cell-like state. These data point to an important role for chromatin assembly in suppressing transcription from MERVL LTRs. Notably, this study also reported that 2-cell-stage genes that are induced upon loss of CAF-1 are generally located within relatively close proximity to MERVL LTRs and there was also substantial transcriptional read-through from LTRs into such genes,³⁴³ revealing that activation of MERVL LTRs can be directly linked to expression of 2-cell stage genes. Given the broad role for CAF-1 in H3/H4 deposition into nucleosomes during S-phase and the extremely high upregulation of MERVL detected in this study (~200-fold by qRT-PCR and RNA-seq) it is likely that depletion of CAF-1 affects MERVLs located in H3K9me2-rich domains as well as those outside such regions, as observed in this study. Since nucleosome deposition is necessary for G9a/GLP to deposit H3K9me2, it is probable that hnRNP K and G9a/GLP function downstream of CAF-1 and provide a further repressive barrier to MERVL expression and prevent reacquisition of 2-cell stage transcriptome characteristics in pluripotent mESCs. Therefore, a major unifying theme for the diversity of factors that are required to maintain MERVL silencing may be that these factors exert a globally repressive effect on chromatin, similar to CAF-1-mediated nucleosome assembly, which is necessary for preventing full scale dedifferentiation into a 2-cell-like state, perhaps as a consequence of activation of specific MERVL LTRs, which are “poised” for induction in mESCs.

5. Concluding Remarks and Future Directions

5.1 KMT-dependent silencing of ERVs involves the co-repressive activities of hnRNP K

The transcriptional silencing of retrotransposons during mammalian development provides a biologically relevant model system to characterize the functions of the H3K9 family of KMTs with the goal of increasing our understanding of their roles in human disease. In the present work, by investigating the regulatory mechanisms governing ERV repression by SETDB1/KAP1 and G9a/GLP in mESCs, I identified and characterized hnRNP K as a novel co-repressive factor for both of these KMT complexes and uncovered important mechanistic themes in the function of these KMTs in transcriptional silencing.

The mechanisms by which different ERVs are silenced by H3K9-specific KMTs in mESCs are very different. While this conclusion was generally evident from previous findings, my work now extends and builds upon these data and provides new insights into the molecular basis of this difference. The deposition of H3K9me3 by SETDB1 at class I and II ERVs follows an apparently hierarchical mechanism. This was initially supported by the observation that SETDB1 recruitment at ERVs requires KAP1,^{89,90} which in turn requires KRAB-ZFPs and other zinc finger proteins.^{220,254,301} However, my work shows that H3K9me3 at ERVs ultimately requires not only KRAB-ZFP/KAP1 binding, but also SUMOylation of chromatin-bound proteins (including KAP1), making the regulation of chromatin SUMOylation a critical feature of this pathway. To date, there are only a few studies which have investigated SUMOylation of chromatin-bound proteins genome-wide.^{308,311,344} Consistent with the results presented here, a recent RNAi screen for proviral silencing factors demonstrated essential roles for SUMO2 and Ubc9 and ChIP-seq

confirmed the co-occupancy of KAP1 and SUMO2 on ERV chromatin.³¹¹ Moreover, SUMO1 and/or SUMO2 are also found at the 3' ends of KRAB-ZFP genes³⁰⁸ which are also bound by SETDB1/KAP1 and ZNF274 and enriched in H3K9me3 in human transformed cell lines.^{255,309} Thus a hallmark of SETDB1/H3K9me3 recruitment by KAP1 at both ERVs and genes is chromatin SUMOylation. In this context, hnRNP K may exert a co-repressive effect by promoting or enhancing chromatin protein SUMOylation at KAP1-bound ERVs. It is important to note that SETDB1 and H3K9me3 can be recruited to the genome in other ways including direct binding to TFs,^{283,345} binding to lncRNA³⁰⁶ and binding to short promoter-associated RNAs with AGO2.³⁴⁶ Therefore, recruitment by chromatin protein SUMOylation is a unique mechanism for SETDB1 binding at ERVs. SUMOylation at KAP1-bound loci provides an additional regulatory mechanism whereby SETDB1 recruitment can be decoupled from KRAB-ZFP/KAP1 binding at ERVs. This feature may be important during S-phase, where H3K9me3 is re-established at ERVs by SETDB1 and the presence of chromatin-bound SUMOylated KAP1 and other proteins serving as essential “epigenetic” signalling marks in this process. Another intriguing possibility is that histones occupying ERV sequences may themselves be SUMOylated, since histone H4 is SUMOylated in human cells³⁴⁷ and it is tempting to speculate that histone H4 SUMOylation could be promoted by KAP1 SUMO E3 ligase activity,²⁵⁸ which can act on its binding partners.^{348,349} Interestingly, although SUMOylation is an ancient posttranslational modification, as evidenced by the conservation of SUMOylation and de-SUMOylation machinery from yeast to human, there is currently no decisive evidence linking chromatin protein SUMOylation to H3K9 methylation and retrotransposon silencing in other species. Thus, its involvement in retrotransposon silencing may be a unique feature of the KRAB-

ZFP/KAP1 system in tetrapod vertebrates. Further investigations of protein SUMOylation in retrotransposon silencing in other eukaryotes would address this question.

In contrast with class I and II ERVs, MERVLs harbour low levels of SUMO1-marked proteins,²⁴⁵ indicating that chromatin protein SUMOylation is unlikely to play a central role in their repression. Previous studies suggested that class III MERVL elements may be capable of autonomously recruiting H3K9me2 and G9a/GLP to maintain their silencing.^{85,86} However, my work supports the conclusion that MERVL does not autonomously recruit G9a/GLP/H3K9 dimethylation, but is likely to acquire H3K9me2 as a consequence of genomic location and transcriptional inactivity. Indeed, H3K9me2 occupies broad megabase sized domains in mammalian cells including mESCs^{82,332} and a recent study found that H3K9me1/2-marked nucleosomes stimulate G9a/GLP activity towards neighbouring unmodified nucleosomes, catalyzing the expansion of H3K9me2 marked regions and silencing cohorts of genes during embryonic development.⁸¹ Therefore, unlike the relatively local enrichment of H3K9me3 over specific ERV sequences and limited spreading into the surrounding chromatin,³⁵⁰ H3K9me2 by G9a/GLP is a broadly spreading mark that may only be inhibited by the presence of other chromatin modifications at its boundaries such as H3S10 phosphorylation, H3K4 methylation and histone acetylation. MERVLs and a subset of other genes located in these broad H3K9me2-rich domains apparently require hnRNP K for G9a/GLP recruitment and this may involve an association with lncRNA or other RNA species that contributes targeting information. Alternatively, since after replication only ~50% of the H3K9me2 will need to be re-established, the remaining H3K9me2 may be sufficient to recruit G9a/GLP. In contrast with MERVLs located in H3K9me2-rich domains,

the lack of expression from MERVL LTRs lacking H3K9me2 suggests that the lack of transcriptional activators is important for maintaining their inactivity. Clearly this model is speculative and there are many outstanding questions to be addressed concerning the recruitment and function of hnRNP K and G9a/GLP to MERVL sequences.

5.2 HNRNPK mutations in Mendelian disease links with G9a/GLP

The roles of *HNRNPK* in human development remained unknown until very recently. Au *et al.*²³² reported two different *de novo* mutations in *HNRNPK* identified by whole-exome sequencing in two independent male probands currently ages 17 and 11 with a distinct spectrum of congenital anomalies and intellectual disability, termed Au-Kline syndrome (OMIM:616580). Proband 1 had the mutation c.953+1dup, which is a frameshift in exon 12 and proband 2 had the mutation c.257G>A, which abolishes a splice site in exon 6. Both mutations affect all known *HNRNPK* transcripts and likely lead to loss of function via NMD.²³² These variants were implicated as causative for the phenotype due to their absence from dbSNP, the Exome Server, 1000 genomes project, parents of the probands and additional in-house controls.²³² In addition, western blot analysis of proband fibroblasts demonstrated ~50% reduction in hnRNP K protein levels relative to controls.²³² Clinical features of Au-Kline syndrome include craniofacial anomalies such as rigid metopic sutures, an elongated face, long palpebral fissures, wide nasal ridge, open downturned mouth with high palate and long tongue groove, cardiac anomalies, skeletal anomalies including hip dysplasia and scoliosis, cryptorchidism, hypotonia and mild to moderate intellectual disability with language delay²³² (**Table 3**). Prior to this report, two other studies showed evidence of a novel microdeletion syndrome associated with loss of ~2-2.6 Mb of 9q21.^{231,233} These studies reported two female probands with *de novo* microdeletions

identified by aCGH one encompassing a 2.565 Mb region at 9q21.32-q21.33 including *HNRNPK* and 11 other genes, the second encompassing a 2 Mb region at the same site including *HNRNPK* with six other genes. Both probands showed clinical features overlapping with those identified by Au *et al.*,²³² including similar craniofacial and cardiac anomalies, dysmorphic features, hypotonia, and moderate to severe developmental delay.^{231,233} However, the phenotype presented each of these studies was more severe, and the proband described by Hancarova *et al.*²³³ had very severe developmental delay including intellectual disability, which may be due to the loss of dosage of the additional genes in their probands. Notably, an earlier study had also identified a microdeletion at 9q21 encompassing *HNRNPK* and four other genes in a proband with multiple congenital anomalies and moderate developmental delay,³⁵¹ providing further evidence of a clinically distinct microdeletion syndrome which includes loss one *HNRNPK* allele. Therefore, taken together, these studies collectively demonstrate that haploinsufficiency of *HNRNPK* causes a clinically distinct Mendelian disorder, which involves congenital defects and developmental delay.

My work is consistent with these recent findings that demonstrate an important role for *HNRNPK* in human development and lend important insights into the functions of hnRNP K on a molecular level in embryonic cells in showing that it contributes to epigenetic regulation by G9a/GLP and SETDB1 (**Illustration 6**). I postulate that hnRNP K promotes SETDB1 and G9a/GLP recruitment to gene promoters, thereby ensuring H3K9 methylation and gene repression during embryonic development (**Illustration 6**). As discussed in the introduction, mutations in genes encoding KMTs result in Mendelian disorders due to haploinsufficiency. This paradigm is clearly in accord with the genetic etiology of Au-Kline syndrome and it is

apparent that some of the clinical features of this phenotype could be explained by defects in epigenetic regulatory pathways governed by H3K9 methylation. This is specifically evidenced by a comparison of Au-Kline syndrome with Kleefstra syndrome (KS), which is characterized by intellectual disability and a distinct spectrum of congenital anomalies due to heterozygous mutations in *GLP*.^{104,105} A comparison of these disorders reveals that KS shares some clinical features common with Au-Kline syndrome (**Table 3**). These observations support the notion that hnRNP K is involved in developmental regulation of transcription by G9a/GLP and reduced dosage of either leads to congenital malformation syndromes, where the most prominent unifying feature is intellectual disability (**Table 3**). Consistent with this idea, while intellectual disability is a complex genetic disorder, the disruption of chromatin modifiers is a recurrent theme (reviewed by Benevento *et al.*³⁵²). For example, mutations in genes encoding chromatin regulators such as *MBD5*, *MLL3*, *SMARCB1* and *NR113* have been found to contribute to the locus heterogeneity in the KS spectrum of clinical features, where probands were negative for *EHMT1* mutations.³⁵³ Obviously an important caveat here is that hnRNP K is involved in many different cellular processes¹³⁸ and contributes to hippocampal neuronal function and morphology via post-transcriptional regulation and cytoskeletal pathways.^{354,355} Thus it is likely that hnRNP K deficiency also affects cognitive phenotypes in a G9a/GLP-independent manner.

Previously there were no studies that characterized the phenotypes of *Hnrnpk* null mutations in mice. However, while this thesis was in preparation, Gallardo *et al.* (2015) generated mice carrying a genetrap integration into intron 2 of *Hnrnpk*.³⁵⁶ In contrast with the mESC line generated by Horie *et al.*²⁴¹ carrying a genetrap integration into intron 1, which apparently

had no effect on hnRNP K protein levels this genetrapp efficiently blocked all *Hnrnpk* transcripts from its associated allele as evidenced by the depletion of hnRNP K protein in the heterozygous mice.³⁵⁶ This study demonstrated that hnRNP K is essential for embryonic development, as hnRNP K homozygous mutants (*Hnrnpk*^{-/-}) died prior to E13.5.³⁵⁶ Notably, this study did not determine the developmental time at which the *Hnrnpk* KO embryos showed lethality, nor did it investigate the cause of the lethality. However, based on my work and that of Lin et al.¹⁴¹, which collectively show that hnRNP K depletion abolishes mESC self-renewal and pluripotency, the *Hnrnpk* KO could result in very early lethality, possibly by the expanded blastocyst stage at ~E4.5-5.5 when mESCs can be derived. Intriguingly, *Hnrnpk*^{+/-} mice also showed abnormal phenotypes, including postnatal lethality at ~30% penetrance and those surviving past weaning exhibited a profound growth defect,³⁵⁶ consistent with the idea that hnRNP K is haploinsufficient for embryonic development in mammals. Unfortunately, this study did not characterize the developmental phenotypes of the *Hnrnpk*^{+/-} mice in greater detail, leaving open the question of whether these mice display evidence of similar craniofacial malformations and cognitive impairment found in Au-Kline syndrome patients. Such an investigation is clearly warranted and it would also be of great interest to determine whether *Hnrnpk*^{+/-} blastocysts show evidence of upregulation of different ERV families and whether these embryos and/or adult mice exhibit global H3K9me2 deficiencies, as suggested by my work in mESCs. Moreover, it would also be of interest to determine whether *Hnrnpk*^{+/-}; *Glp*^{+/-} mice show an enhancement of the KS phenotypes observed in *Glp*^{+/-} mice,³⁵⁷ which would indicate that hnRNP K actually promotes G9a/GLP activity during development.

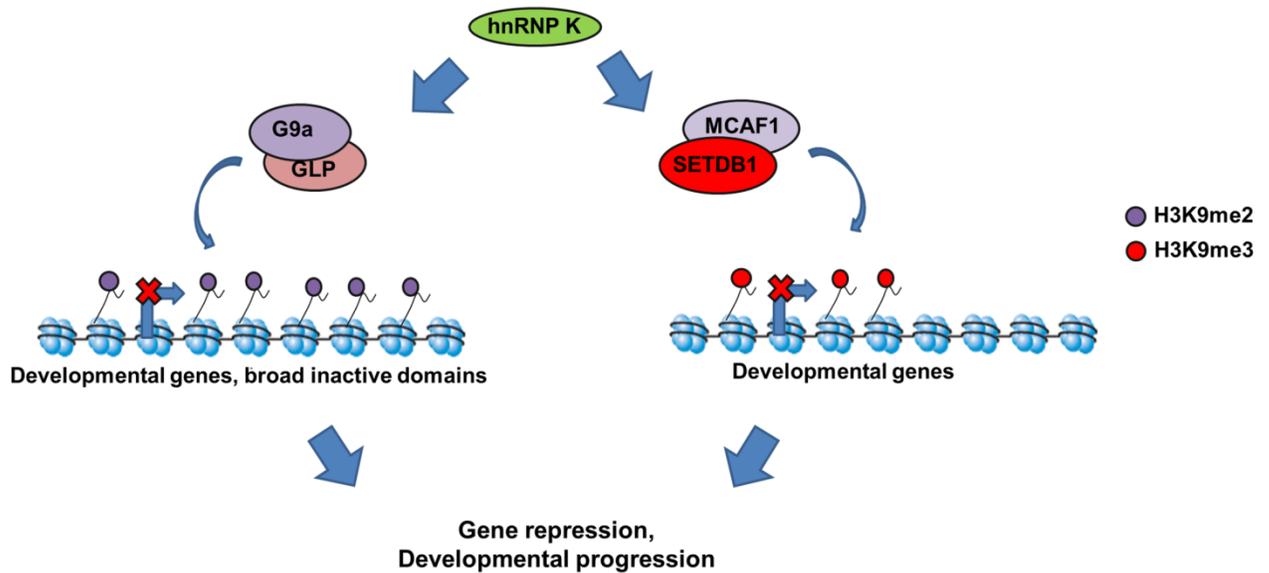


Illustration 6. Proposed role of hnRNP K in human development with the G9a/GLP and SETDB1/MCAF1 KMT complexes. During development, hnRNP K may function in chromatin regulatory pathways with the G9a/GLP and SETDB1/MCAF1 complexes. In association with G9a/GLP, hnRNP K may facilitate H3K9me2 at gene promoters as well as broad transcriptionally inactive domains. HnRNP K may also participate in the recruitment of SETDB1/MCAF1 complexes to deposit H3K9me3 and repress developmental genes. These functions collectively would contribute to appropriate developmental progression and tissue specification.

Table 3. A comparison of major clinical features in Au-Kline syndrome and Kleefstra syndrome caused by mutations in *EHMT1*

Tissue/System	Clinical feature	Au-Kline syndrome ¹	Kleefstra syndrome ²
Craniofacial	rigid metopic sutures,	++	-
	broad and elongated face	++	+
	flat face	-	++
	microcephaly	-	++
	brachycephaly	-	+
	midface hypoplasia	-	+
Eyes	downslanting palpebral fissures	+	-
Mouth	downturned mouth, open bite	+	-
	carp mouth, tented lip	-	++
Nose	wide nasal ridge,	+	-
	short nose,	-	+
Ears	large, low set, malformed	+	+
Heart	Atrial and ventricular septal defects	+	+
Skeletal/ Neuromuscular	hip dysplasia,	++	-
	hypotonia	++	++
	extra lumbar vertebra	++	-
	scoliosis	+	-
Genitourinary	cryptorchidism	++	+
Brain	hypomyelination	+	-
Cognitive	mild to severe intellectual disability	++	++

¹From Au et al. (2015)

²From Kleefstra et al. (2006)

+ = Found in >25% of probands

++ = Found in >50% of probands

Red = Common features

5.3 Role of *HNRNPK* in *KMT*-dependent gene silencing in cancer

Recent evidence suggests that hnRNP K may function as either a tumour-suppressor or oncogene depending on the cell type. A novel role for hnRNP K as a tumour-suppressor was shown in the same study generating the *Hnrnpk* KO. *Hnrnpk*^{+/-} mice were shown to exhibit decreased lifespan as a result of their very high propensity to develop hematological malignancies, particularly of the myeloid lineage.³⁵⁶ Bone marrow samples also revealed very high levels of genomic instability, as evidenced by chromosomal fusions, breakage and polyploidy in *Hnrnpk*^{+/-} mice.³⁵⁶ These phenotypes were linked with the previously reported

ability of hnRNP K to act as a co-activator for p53 in attenuating cell cycle progression by activating *p21*.^{327,356} Similarly, mono-allelic loss of *HNRNPK* is found in ~2% of acute myeloid leukemia (AML) patients,³⁵⁶ which show somatic microdeletion of the 9q21 region encompassing five other genes.³⁵⁸ Notably, G9a was recently shown to function as an oncogenic factor in AML in mice, where loss of G9a or inhibition of its catalytic activity leads to reduced leukemic stem cell frequency.³⁵⁹ Therefore, these studies suggest that hnRNP K and G9a/GLP have different functions in the regulation of hematopoietic cell proliferation and differentiation programs.

In addition to a role for hnRNP K as a tumour-suppressor in hematological malignancy, earlier reports demonstrated overexpression and oncogenic activity of hnRNP K in malignancy (reviewed by Barboro et al.³⁶⁰). Higher expression of *HNRNPK* has been found in melanoma in addition to colorectal, prostate, lung and breast carcinomas and in chronic myeloid leukemia as compared with normal non-neoplastic tissues.^{235,236,361–363} Notably, while hnRNP K expression is stimulated by growth factors³⁶³ and is generally higher in the nucleus of proliferating versus quiescent cells,³⁶⁴ in studies reporting its overexpression in cancer aberrant cytoplasmic localization was also frequently observed and associated with poor prognosis.³⁶⁰ Thus, the erroneous cytoplasmic activity of hnRNP K rather than its nuclear functions may play a dominant role in its oncogenic activity. HnRNP K was shown to activate the *Myc* promoter and gene encoding the translation factor eIF4E during neoplastic transformation, which promotes oncogenesis.^{281,365,366} However, no studies have investigated the contribution of hnRNP K overexpression to the dysfunction of chromatin regulatory pathways in malignancy. As described in the introduction, KMTs including

SETDB1 and G9a are frequently overexpressed in various cancers. Therefore, it is possible that part of the oncogenic mechanism of hnRNP K action involves the promotion of H3K9me_{2/3} by these KMTs at tumour suppressor genes, which promotes malignant tumour growth and metastasis in such cancers.^{35,367} Future studies could investigate this by determining whether depletion of hnRNP K affects H3K9me_{2/3} levels at tumour suppressor genes and influences tumor phenotypes in malignancies where SETDB1 and/or G9a are overexpressed.

In conclusion, my work points to hnRNP K as a novel co-repressor for SETDB1 and G9a/GLP during embryonic development and specifically in embryonic stem cells. Recent evidence from basic and clinical genetic studies is consistent with my work and indicates that hnRNP K is important for human development and somatic cell identity. Further investigation into the role of hnRNP K in transcriptional regulation mediated by these and other KMTs may guide the design of novel therapeutic strategies for the treatment and prevention of intellectual disability and cancer.

Bibliography

1. Bird, A. Perceptions of epigenetics. *Nature* **447**, 396–398 (2007).
2. Berger, S. L., Kouzarides, T., Shiekhatar, R. & Shilatifard, A. An operational definition of epigenetics. *Genes Dev.* **23**, 781–3 (2009).
3. Carone, B. R. *et al.* Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell* **143**, 1084–1096 (2010).
4. Heard, E. & Martienssen, R. a. Transgenerational epigenetic inheritance: Myths and mechanisms. *Cell* **157**, 95–109 (2014).
5. Luger, K., Mäder, a W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
6. Luger, K., Rechsteiner, T. J., Flaus, a J., Wayne, M. M. & Richmond, T. J. Characterization of nucleosome core particles containing histone proteins made in bacteria. *J. Mol. Biol.* **272**, 301–11 (1997).
7. Cutter, A. R. & Hayes, J. J. A brief review of nucleosome structure. *FEBS Lett.* **589**, 2914–22 (2015).
8. Tessarz, P. & Kouzarides, T. Histone core modifications regulating nucleosome structure and dynamics. *Nat. Rev. Mol. Cell Biol.* **15**, 703–708 (2014).
9. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–26 (2006).
10. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–60 (2007).
11. Herz, H. M., Garruss, A. & Shilatifard, A. SET for life: Biochemical activities and biological functions of SET domain-containing proteins. *Trends Biochem. Sci.* **38**, 621–639 (2013).
12. Spellmon, N., Holcomb, J., Trescott, L., Sirinpong, N. & Yang, Z. Structure and Function of SET and MYND Domain-Containing Proteins. *Int. J. Mol. Sci.* **16**, 1406–1428 (2015).
13. Feng, Q. *et al.* Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr. Biol.* **12**, 1052–1058 (2002).
14. Binda, O. On your histone mark, SET, methylate! *Epigenetics* **8**, 457–463 (2013).
15. Chin, H. G. *et al.* Sequence specificity and role of proximal amino acids of the histone H3 tail on catalysis of murine G9a lysine 9 histone H3 methyltransferase. *Biochemistry* **44**, 12998–13006 (2005).
16. Chin, H. G., Patnaik, D., Estève, P. O., Jacobsen, S. E. & Pradhan, S. Catalytic properties and kinetic mechanism of human recombinant Lys-9 histone H3 methyltransferase SUV39H1: Participation of the chromodomain in enzymatic catalysis. *Biochemistry* **45**, 3272–3284 (2006).
17. Sampath, S. C. *et al.* Methylation of a Histone Mimic within the Histone

- Methyltransferase G9a Regulates Protein Complex Assembly. *Mol. Cell* **27**, 596–608 (2007).
18. Binda, O. *et al.* Trimethylation of histone H3 lysine 4 impairs methylation of histone H3 lysine 9: regulation of lysine methyltransferases by physical interaction with their substrates. *Epigenetics* **5**, 767–775 (2010).
 19. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
 20. Wan, M. *et al.* The trithorax group protein Ash2l is essential for pluripotency and maintaining open chromatin in embryonic stem cells. *J. Biol. Chem.* **288**, 5039–5048 (2013).
 21. Ang, Y. S. *et al.* Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145**, 183–187 (2011).
 22. Pasini, D., Bracken, A. P., Hansen, J. B., Capillo, M. & Helin, K. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol. Cell. Biol.* **27**, 3769–79 (2007).
 23. Margueron, R. *et al.* Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* **461**, 762–767 (2009).
 24. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349 (2011).
 25. Tachibana, M. *et al.* Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes Dev.* **19**, 815–826 (2005).
 26. Bian, C., Chen, Q. & Yu, X. The zinc finger proteins ZNF644 and WIZ regulate the G9a/GLP complex for gene repression. *Elife* **4**, 1–17 (2015).
 27. Ueda, J., Tachibana, M., Ikura, T. & Shinkai, Y. Zinc finger protein Wiz links G9a/GLP histone methyltransferases to the co-repressor molecule CtBP. *J. Biol. Chem.* **281**, 20120–20128 (2006).
 28. Souza, P. P. *et al.* The histone methyltransferase SUV420H2 and Heterochromatin Proteins HP1 interact but show different dynamic behaviours. *BMC Cell Biol.* **10**, 41 (2009).
 29. Melcher, M. *et al.* Structure-function analysis of SUV39H1 reveals a dominant role in heterochromatin organization, chromosome segregation, and mitotic progression. *Mol. Cell. Biol.* **20**, 3728–3741 (2000).
 30. Aagaard, L. *et al.* Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31. *EMBO J.* **18**, 1923–1938 (1999).
 31. Lachner, M., O’Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).
 32. Bannister, a J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).

33. Wang, H. *et al.* mAM Facilitates Conversion by ESET of Dimethyl to Trimethyl Lysine 9 of Histone H3 to Cause Transcriptional Repression. *Mol Cell* **12**, 475–487 (2003).
34. Fritsch, L. *et al.* A subset of the histone H3 lysine 9 methyltransferases Suv39h1, G9a, GLP, and SETDB1 participate in a multimeric complex. *Mol. Cell* **37**, 46–56 (2010).
35. Ceol, C. J. *et al.* The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. *Nature* **471**, 513–7 (2011).
36. Grimaud, C., Nègre, N. & Cavalli, G. From genetics to epigenetics: The tale of Polycomb group and trithorax group genes. *Chromosom. Res.* **14**, 363–375 (2006).
37. Elgin, S. C. R. & Reuter, G. Position-effect variegation, heterochromatin formation, and gene silencing in *Drosophila*. *Cold Spring Harb. Perspect. Biol.* **5**, (2013).
38. Denissov, S. *et al.* Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development* **141**, 526–37 (2014).
39. Hu, D. *et al.* The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Mol. Cell. Biol.* **33**, 4745–54 (2013).
40. Bledau, A. S. *et al.* The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation. *Development* **141**, 1022–35 (2014).
41. Andreu-Vieyra, C. V. *et al.* MLL2 is required in oocytes for bulk histone 3 lysine 4 trimethylation and transcriptional silencing. *PLoS Biol.* **8**, 53–54 (2010).
42. Hung, T. *et al.* ING4 Mediates Crosstalk between Histone H3 K4 Trimethylation and H3 Acetylation to Attenuate Cellular Transformation. *Mol. Cell* **33**, 248–256 (2009).
43. Wysocka, J. *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**, 86–90 (2006).
44. Margueron, R. *et al.* Ezh1 and Ezh2 Maintain Repressive Chromatin through Different Mechanisms. *Mol. Cell* **32**, 503–518 (2008).
45. Shen, X. *et al.* EZH1 Mediates Methylation on Histone H3 Lysine 27 and Complements EZH2 in Maintaining Stem Cell Identity and Executing Pluripotency. *Mol. Cell* **32**, 491–502 (2008).
46. Mozzetta, C. *et al.* The Histone H3 Lysine 9 Methyltransferases G9a and GLP Regulate Polycomb Repressive Complex 2-Mediated Gene Silencing. *Mol. Cell* **53**, 277–289 (2014).
47. O’Carroll, D., Erhardt, S., Pagani, M. & Barton, S. C. The Polycomb -Group Gene Ezh2 Is Required for Early Mouse Development. *Mol. Cell. Biol.* **21**, 4330–4336 (2001).
48. Plath, K. *et al.* Role of Histone H3 Lysine 27 Methylation in X Inactivation. *Science* **300**, 131–135 (2003).
49. Silva, J. *et al.* Establishment of histone H3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev. Cell* **4**, 481–495 (2003).

50. Froberg, J. E., Yang, L. & Lee, J. T. Guided by RNAs: X-inactivation as a model for lncRNA function. *J. Mol. Biol.* **425**, 3698–3706 (2013).
51. Boyer, L. a *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349–53 (2006).
52. Cao, R., Wang, L., Wang, H. & Xia, L. Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science* **298**, 1039–1044 (2002).
53. Min, J., Zhang, Y. & Xu, R. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27 Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev.* **17**, 1823–1828 (2003).
54. Tavares, L. *et al.* RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **148**, 664–678 (2012).
55. Yu, M. *et al.* Direct Recruitment of Polycomb Repressive Complex 1 to Chromatin by Core Binding Transcription Factors. *Mol. Cell* **45**, 330–343 (2012).
56. Eskeland, R. *et al.* Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination. *Mol. Cell* **38**, 452–464 (2010).
57. Kaneko, S., Son, J., Shen, S. S., Reinberg, D. & Bonasio, R. PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1258–64 (2013).
58. Kaneko, S., Son, J., Bonasio, R., Shen, S. S. & Reinberg, D. Nascent RNA interaction keeps PRC2 activity poised and in check. *Genes Dev.* **28**, 1983–1988 (2014).
59. Riising, E. M. *et al.* Gene silencing triggers polycomb repressive complex 2 recruitment to CpG Islands genome wide. *Mol. Cell* **55**, 347–360 (2014).
60. Campbell, S., Ismail, I. H., Young, L. C., Poirier, G. G. & Hendzel, M. J. Polycomb repressive complex 2 contributes to DNA double-strand break repair. *Cell Cycle* **12**, 2675–2683 (2013).
61. Edmunds, J. W., Mahadevan, L. C. & Clayton, A. L. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J.* **27**, 406–420 (2008).
62. Kuo, A. J. *et al.* NSD2 Links Dimethylation of Histone H3 at Lysine 36 to Oncogenic Programming. *Mol. Cell* **44**, 609–620 (2011).
63. Tanaka, Y., Katagiri, Z. I., Kawahashi, K., Kioussis, D. & Kitajima, S. Trithorax-group protein ASH1 methylates histone H3 lysine 36. *Gene* **397**, 161–168 (2007).
64. Tanaka, Y. *et al.* Dual function of histone H3 lysine 36 methyltransferase ASH1 in regulation of hox gene expression. *PLoS One* **6**, 1–9 (2011).
65. Fnu, S. *et al.* Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 540–545 (2011).
66. Pfister, S. X. *et al.* SETD2-Dependent Histone H3K36 Trimethylation Is Required for Homologous Recombination Repair and Genome Stability. *Cell Rep.* **7**, 2006–2018 (2014).

67. Lee, J. S. & Shilatifard, A. A site to remember: H3K36 methylation a mark for histone deacetylation. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **618**, 130–134 (2007).
68. Luco, R. *et al.* Regulation of Alternative Splicing by Histone Modifications. *Science* **327**, 996–1000 (2010).
69. Baubec, T. *et al.* Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520**, 243–247 (2015).
70. Hu, M. *et al.* Histone H3 lysine 36 methyltransferase Hypb/Setd2 is required for embryonic vascular remodeling. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2956–2961 (2010).
71. Rayasam, G. V. *et al.* NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *EMBO J.* **22**, 3153–3163 (2003).
72. Nimura, K. *et al.* A histone H3 lysine 36 trimethyltransferase links Nkx2-5 to Wolf-Hirschhorn syndrome. *Nature* **460**, 287–291 (2009).
73. Peters, A. H. F. M. *et al.* Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. *Nat. Genet.* **30**, 77–80 (2002).
74. Santoro, R. & Grummt, I. Epigenetic Mechanism of rRNA Gene Silencing : Temporal Order of NoRC-Mediated Histone Modification, Chromatin Remodeling and DNA methylation. *Mol. Cell. Biol.* **25**, 2539–2546 (2005).
75. Groner, A. C. *et al.* KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. *PLoS Genet.* **6**, (2010).
76. Azzaz, A. M. *et al.* Human heterochromatin protein 1 alpha promotes nucleosome associations that drive chromatin condensation. *J. Biol. Chem.* **289**, 6850–6861 (2014).
77. Lehnertz, B. *et al.* Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. *Curr. Biol.* **13**, 1192–1200 (2003).
78. Peters, A. H. F. M. *et al.* Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**, 323–337 (2001).
79. Bulut-Karslioglu, A. *et al.* Suv39h-Dependent H3K9me3 Marks Intact Retrotransposons and Silences LINE Elements in Mouse Embryonic Stem Cells. *Mol. Cell* **55**, 277–290 (2014).
80. Tachibana, M. *et al.* G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev.* **16**, 1779–1791 (2002).
81. Liu, N. *et al.* Recognition of H3K9 methylation by GLP is required for efficient establishment of H3K9 methylation, rapid target gene repression, and mouse viability. *Genes Dev.* **29**, 379–393 (2015).
82. Wen, B., Wu, H., Shinkai, Y., Irizarry, R. a & Feinberg, A. P. Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nat. Genet.* **41**, 246–250 (2009).
83. Yokochi, T. *et al.* G9a selectively represses a class of late-replicating genes at the nuclear periphery. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19363–19368 (2009).

84. Dong, K. B. *et al.* DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *EMBO J.* **27**, 2691–2701 (2008).
85. Macfarlan, T. S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 (2012).
86. Maksakova, I. A. *et al.* Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. *Epigenetics Chromatin* **6**, 15 (2013).
87. Falandry, C. *et al.* CLLD8/KMT1F is a lysine methyltransferase that is important for chromosome segregation. *J. Biol. Chem.* **285**, 20234–20241 (2010).
88. Dodge, J. E., Kang, Y., Beppu, H. & Lei, H. Histone H3-K9 Methyltransferase ESET Is Essential for Early Development Histone H3-K9 Methyltransferase ESET Is Essential for Early Development. *Mol. Cell. Biol.* **24**, 2478–2486 (2004).
89. Matsui, T. *et al.* Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464**, 927–31 (2010).
90. Rowe, H. M. *et al.* KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237–40 (2010).
91. Quenneville, S. *et al.* In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol. Cell* **44**, 361–372 (2011).
92. Pinheiro, I. *et al.* Prdm3 and Prdm16 are H3K9me1 methyltransferases required for mammalian heterochromatin integrity. *Cell* **150**, 948–960 (2012).
93. Marks, H. *et al.* The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590–604 (2012).
94. Elsässer, S. J., Noh, K. M., Diaz, N., Allis, C. D. & Banaszynski, L. A. Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. *Nature* **522**, 240–4 (2015).
95. Liu, S. *et al.* Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes Dev.* **28**, 2041–55 (2014).
96. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **40**, 897–903 (2008).
97. Allshire, R. C. & Ekwall, K. Epigenetic Regulation of Chromatin States in *Schizosaccharomyces pombe*. *Cold Spring Harb. Perspect. Biol.* **7**, a018770 (2015).
98. Nishioka, K. *et al.* PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent Chromatin. *Mol. Cell* **9**, 1201–1213 (2002).
99. Schotta, G. *et al.* A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. *Genes Dev.* **22**, 2048–2061 (2008).
100. Congdon, L. M., Houston, S. I., Veerappan, C. S., Spektor, T. M. & Rice, J. C. PR-Set7-mediated monomethylation of histone H4 lysine 20 at specific genomic regions

- induces transcriptional repression. *J. Cell. Biochem.* **110**, 609–619 (2010).
101. Karachentsev, D., Sarma, K., Reinberg, D. & Steward, R. PR-Set7-dependent methylation of histone H4 Lys 20 functions in repression of gene expression and is essential for mitosis. *Genes Dev.* **31**, 431–435 (2005).
 102. Sims, J. K., Houston, S. I., Magazinnik, T. & Rice, J. C. A trans-tail histone code defined by monomethylated H4 Lys-20 and H3 Lys-9 demarcates distinct regions of silent chromatin. *J. Biol. Chem.* **281**, 12760–12766 (2006).
 103. Oda, H. *et al.* Monomethylation of histone H4-lysine 20 is involved in chromosome structure and stability and is essential for mouse development. *Mol. Cell. Biol.* **29**, 2278–2295 (2009).
 104. Kleefstra, T. *et al.* Disruption of the gene Euchromatin Histone Methyl Transferase1 (Eu-HMTase1) is associated with the 9q34 subtelomeric deletion syndrome. *J. Med. Genet.* **42**, 299–306 (2005).
 105. Kleefstra, T. *et al.* Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. *Am. J. Hum. Genet.* **79**, 370–377 (2006).
 106. Kurotaki, N. *et al.* Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat. Genet.* **30**, 365–366 (2002).
 107. Stec, I. *et al.* WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a Drosophila dysmorphia gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma. *Hum. Mol. Genet.* **7**, 1071–1082 (1998).
 108. Ng, S. B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat. Genet.* **42**, 790–793 (2010).
 109. Gibson, W. T. *et al.* Mutations in EZH2 cause weaver syndrome. *Am. J. Hum. Genet.* **90**, 110–118 (2012).
 110. Grozeva, D. *et al.* De novo loss-of-function mutations in SETD5, encoding a methyltransferase in a 3p25 microdeletion syndrome critical region, cause intellectual disability. *Am. J. Hum. Genet.* **94**, 618–624 (2014).
 111. Kuechler, A. *et al.* Loss-of-function variants of SETD5 cause intellectual disability and the core phenotype of microdeletion 3p25.3 syndrome. *Eur. J. Hum. Genet.* 753–760 (2014). doi:10.1038/ejhg.2014.165
 112. McGrath, J. & Trojer, P. Targeting histone lysine methylation in cancer. *Pharmacol. Ther.* **150**, 1–22 (2015).
 113. Ford, D. J. & Dingwall, A. K. The Cancer COMPASS: Navigating the functions of MLL complexes in cancer. *Cancer Genet.* **208**, 178–191 (2015).
 114. Rodriguez-Paredes, M. *et al.* Gene amplification of the histone methyltransferase SETDB1 contributes to human lung tumorigenesis. *Oncogene* **33**, 2807–13 (2014).
 115. Sun, Y. *et al.* Histone methyltransferase SETDB1 is required for prostate cancer cell proliferation, migration and invasion. *Asian J. Androl.* **16**, 319–24 (2014).
 116. Jaffe, J. D. *et al.* Global chromatin profiling reveals NSD2 mutations in pediatric acute

- lymphoblastic leukemia. *Nat. Genet.* **45**, 1386–91 (2013).
117. Vougiouklakis, T., Hamamoto, R., Nakamura, Y. & Saloura, V. The NSD family of protein methyltransferases in human cancer. *Epigenomics* **7**, 863–74 (2015).
 118. Morin, R. D. *et al.* Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat. Genet.* **42**, 181–185 (2010).
 119. Yap, D. B. *et al.* Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. *Blood* **117**, 2451–2459 (2011).
 120. Lewis, P. *et al.* Inhibition of PRC2 Activity by a Gain-of-Function H3 Mutation Found in Pediatric Glioblastoma. *Science* **340**, 857–861 (2013).
 121. Zingg, D. *et al.* The epigenetic modifier EZH2 controls melanoma growth and metastasis through silencing of distinct tumour suppressors. *Nat. Commun.* **6**, 6051 (2015).
 122. Hamamoto, R., Saloura, V. & Nakamura, Y. Critical roles of non-histone protein lysine methylation in human tumorigenesis. *Nat. Publ. Gr.* **15**, 110–124 (2015).
 123. McCabe, M. T. *et al.* EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* **492**, 108–112 (2012).
 124. Wang, K. C. & Chang, H. Y. Molecular Mechanisms of Long Noncoding RNAs. *Mol. Cell* **43**, 904–914 (2011).
 125. Tsai, M. *et al.* Long Noncoding RNAs as Molecular Scaffolds for Histone Modification Complexes. *Sci. Technol.* **329**, 689–693 (2010).
 126. Chu, C. *et al.* Systematic Discovery of Xist RNA Binding Proteins. *Cell* **161**, 404–16 (2015).
 127. McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236 (2015).
 128. Batista, P. J. & Chang, H. Y. Long Noncoding RNAs : Cellular Address Codes in Development and Disease. *Cell* **152**, 1298–1307 (2013).
 129. Rinn, J. L. *et al.* Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
 130. Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).
 131. Nagano, T., Mitchell, J. A., Sanz, L. A. & Pauler, F. M. The Air Non-Coding RNA Epigenetically Silences transcription by Targeting G9a to Chromatin. *Science* **322**, 1717–1720 (2008).
 132. Pandey, R. R. *et al.* Kcnq1ot1 Antisense Noncoding RNA Mediates Lineage-Specific Transcriptional Silencing through Chromatin-Level Regulation. *Mol. Cell* **32**, 232–246 (2008).
 133. Sauvageau, M. *et al.* Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *Elife* **2**, e01749 (2013).
 134. Li, L. *et al.* Targeted disruption of Hotair leads to homeotic transformation and gene

- derepression. *Cell Rep.* **5**, 3–12 (2013).
135. Gupta, R. A. *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076 (2010).
 136. Piñol-Roma, S., Choi, Y. D., Matunis, M. J. & Dreyfuss, G. Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. *Genes Dev.* **2**, 215–227 (1988).
 137. Jurica, M. S., Licklider, L. J., Gygi, S. R., Grigorieff, N. & Moore, M. J. Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA* **8**, 426–439 (2002).
 138. Han, S. P., Tang, Y. H. & Smith, R. Functional diversity of the hnRNPs: past, present and perspectives. *Biochem. J.* **430**, 379–92 (2010).
 139. Bomsztyk, K., Denisenko, O. & Ostrowski, J. hnRNP K: one protein multiple processes. *Bioessays* **26**, 629–38 (2004).
 140. Huarte, M. *et al.* A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**, 409–19 (2010).
 141. Lin, N. *et al.* An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. *Mol. Cell* **53**, 1005–1019 (2014).
 142. Cubenas-Potts, C. & Matunis, M. J. SUMO: A Multifaceted Modifier of Chromatin Structure and Function. *Dev. Cell* **24**, 1–12 (2013).
 143. Drag, M. & Salvesen, G. S. Critical Review: DeSUMOylating Enzymes — SENPs. *IUBMB Life* **60**, 734–742 (2008).
 144. Shin, J. A. *et al.* SUMO modification is involved in the maintenance of heterochromatin stability in fission yeast. *Mol. Cell* **19**, 817–828 (2005).
 145. Uchimura, Y. *et al.* Involvement of SUMO modification in MBD1- and MCAF1-mediated heterochromatin formation. *J. Biol. Chem.* **281**, 23180–23190 (2006).
 146. Nacerddine, K. *et al.* The SUMO pathway is essential for nuclear integrity and chromosome segregation in mice. *Dev. Cell* **9**, 769–79 (2005).
 147. Leung, D. C. & Lorincz, M. C. Silencing of endogenous retroviruses: when and why do histone marks predominate? *Trends Biochem. Sci.* **37**, 127–33 (2012).
 148. Cordaux, R. & Batzer, M. a. The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.* **10**, 691–703 (2009).
 149. Ribet, D. *et al.* An infectious progenitor for the murine IAP retrotransposon : Emergence of an intracellular genetic parasite from an ancient retrovirus. *Genome Res.* **18**, 597–609 (2008).
 150. Stocking, C. & Kozak, C. a. Murine endogenous retroviruses. *Cell. Mol. Life Sci.* **65**, 3383–98 (2008).
 151. Friedli, M. & Trono, D. The Developmental Control of Transposable Elements and the Evolution of Higher Species. *Annu. Rev. Cell Dev. Biol.* 1–23 (2015).
doi:10.1146/annurev-cellbio-100814-125514
 152. Katoh, I. & Kurata, S.-I. Association of Endogenous Retroviruses and Long Terminal

- Repeats with Human Disorders. *Front. Oncol.* **3**, 234 (2013).
153. Jern, P. & Coffin, J. M. Effects of retroviruses on host genome function. *Annu. Rev. Genet.* **42**, 709–732 (2008).
 154. Marchi, E., Kanapin, A., Magiorkinis, G. & Belshaw, R. Unfixed endogenous retroviral insertions in the human population. *J. Virol.* **88**, 9529–9537 (2014).
 155. Maksakova, I. a *et al.* Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. *PLoS Genet.* **2**, e2 (2006).
 156. Beck, C. R., Garcia-Perez, J. L., Badge, R. M. & Moran, J. V. LINE-1 Elements in Structural Variation and Disease. *Annu. Rev. Genomics Hum. Genet.* **12**, 187–215 (2011).
 157. Lu, X. *et al.* The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat. Struct. Mol. Biol.* **21**, 423–5 (2014).
 158. Liu, M. & Eiden, M. V. Role of human endogenous retroviral long terminal repeats (LTRs) in maintaining the integrity of the human germ line. *Viruses* **3**, 901–905 (2011).
 159. Cowley, M. & Oakey, R. J. Transposable Elements Re-Wire and Fine-Tune the Transcriptome. *PLoS Genet.* **9**, e1003234 (2013).
 160. Mätlik, K., Redik, K. & Speek, M. L1 antisense promoter drives tissue-specific transcription of human genes. *J. Biomed. Biotechnol.* **2006**, 1–16 (2006).
 161. Stoller, J. Z. *et al.* Ash2l interacts with Tbx1 and is required during early embryogenesis. *Exp. Biol. Med. (Maywood)*. **235**, 569–576 (2010).
 162. Blaise, S., de Parseval, N., Bénil, L. & Heidmann, T. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 13013–13018 (2003).
 163. Mi, S. *et al.* Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**, 785–789 (2000).
 164. Lukic, S., Nicolas, J.-C. & Levine, a J. The diversity of zinc-finger genes on human chromosome 19 provides an evolutionary mechanism for defense against inherited endogenous retroviruses. *Cell Death Differ.* **21**, 381–7 (2014).
 165. Jacobs, F. M. J. *et al.* An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* **516**, 242–5 (2014).
 166. Kazazian, H. H. *et al.* Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* **332**, 164–166 (1988).
 167. Lamprecht, B. *et al.* Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nat. Med.* **16**, 571–579, 1p following 579 (2010).
 168. Miki, Y. *et al.* Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* **52**, 643–645 (1992).

169. Morgan, H. D., Sutherland, H. G., Martin, D. I. & Whitelaw, E. Epigenetic inheritance at the agouti locus in the mouse. *Nat. Genet.* **23**, 314–318 (1999).
170. Kassiotis, G. Endogenous retroviruses and the development of cancer. *J. Immunol.* **192**, 1343–9 (2014).
171. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* **20**, 116–7 (1998).
172. Bourc'his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**, 96–99 (2004).
173. Schübeler, D. Function and information content of DNA methylation. *Nature* **517**, 321–6 (2015).
174. Messerschmidt, D. M., Knowles, B. B. & Solter, D. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes Dev.* **28**, 812–828 (2014).
175. Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502**, 472–9 (2013).
176. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
177. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
178. He, Y.-F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
179. Cortázar, D. *et al.* Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **470**, 419–423 (2011).
180. Hajkova, P. *et al.* Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* **117**, 15–23 (2002).
181. Kurimoto, K. *et al.* Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev.* **22**, 1617–1635 (2008).
182. Seisenberger, S. *et al.* The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. *Mol. Cell* **48**, 849–862 (2012).
183. Inoue, A. & Zhang, Y. Replication-Dependent Loss of 5-Hydroxymethylcytosine in Mouse Preimplantation Embryos. *Science* **334**, 194–194 (2011).
184. Inoue, A., Shen, L., Matoba, S. & Zhang, Y. Haploinsufficiency, but Not Defective Paternal 5mC Oxidation, Accounts for the Developmental Defects of Maternal Tet3 Knockouts. *Cell Rep.* **10**, 463–470 (2015).
185. Iqbal, K., Jin, S.-G., Pfeifer, G. P. & Szabó, P. E. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 3642–3647 (2011).
186. Guo, F. *et al.* Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote. *Cell Stem Cell* **15**, 447–458 (2014).

187. Tomizawa, S. *et al.* Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development* **138**, 811–820 (2011).
188. Ziller, M. J. *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**, 477–81 (2013).
189. Okae, H., Chiba, H., Hiura, H., Hamada, H. & Sato, A. Genome-Wide Analysis of DNA Methylation Dynamics during Early Human Development. *PLoS Genet.* **10**, e1004868 (2014).
190. Wang, L. *et al.* Programming and inheritance of parental DNA methylomes in mammals. *Cell* **157**, 979–991 (2014).
191. Nakamura, T. *et al.* PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nat. Cell Biol.* **9**, 64–71 (2007).
192. Nakamura, T. *et al.* PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* **486**, 415–419 (2012).
193. Popp, C. *et al.* Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**, 1101–1105 (2010).
194. Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M. & Saitou, M. Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *EMBO J.* **32**, 340–53 (2013).
195. Yamaguchi, S. *et al.* Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. *Cell Res.* **23**, 329–39 (2013).
196. Hackett, J. A. *et al.* Germline DNA Demethylation Dynamics and Imprint Erasure Through 5-Hydroxymethylcytosine. *Sci.* **339**, 448–452 (2013).
197. Kobayashi, H. *et al.* High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Res.* **23**, 616–627 (2013).
198. Niwa, O. & Sugahara, T. 5-Azacytidine induction of mouse endogenous type C virus and suppression of DNA methylation. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 6290–6294 (1981).
199. Lasneret, J. *et al.* Activation of intracisternal a particles by 5-azacytidine in mouse Ki-BALB cell line. *Virology* **128**, 485–489 (1983).
200. Niwa, O., Yokota, Y., Ishida, H. & Sugahara, T. Independent mechanisms involved in suppression of the Moloney leukemia virus genome during differentiation of murine teratocarcinoma cells. *Cell* **32**, 1105–13 (1983).
201. Barklis, E., Mulligan, R. C. & Jaenisch, R. Chromosomal position or virus mutation permits retrovirus expression in embryonal carcinoma cells. *Cell* **47**, 391–399 (1986).
202. Sorge, J., Cutting, A. E., Erdman, V. D. & Gautsch, J. W. Integration-specific retrovirus expression in embryonal carcinoma cells. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 6627–6631 (1984).
203. Grez, M., Akgün, E., Hilberg, F. & Ostertag, W. Embryonic stem cell virus, a recombinant murine retrovirus with expression in embryonic stem cells. *Proc. Natl.*

- Acad. Sci. U. S. A.* **87**, 9202–9206 (1990).
204. Pannell, D. *et al.* Retrovirus vector silencing is de novo methylase independent and marked by a repressive histone code. *EMBO J.* **19**, 5884–94 (2000).
 205. Hutnick, L. K., Huang, X., Loo, T.-C., Ma, Z. & Fan, G. Repression of retrotransposal elements in mouse embryonic stem cells is primarily mediated by a DNA methylation-independent mechanism. *J. Biol. Chem.* **285**, 21082–91 (2010).
 206. Wolf, D. & Goff, S. P. TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell* **131**, 46–57 (2007).
 207. Wolf, D., Hug, K. & Goff, S. P. TRIM28 mediates primer binding site-targeted silencing of Lys1,2 tRNA-utilizing retroviruses in embryonic cells. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 12521–6 (2008).
 208. Lupo, A. *et al.* KRAB-Zinc Finger Proteins : A Repressor Family Displaying Multiple Bio- logical Functions. *Curr. Genomics* **14**, 268–278 (2013).
 209. Bellefroid, E. J., Poncelet, D. A., Lecocq, P. J., Revelant, O. & Martial, J. A. The evolutionarily conserved Krüppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 3608–3612 (1991).
 210. Margolin, J. F. *et al.* Krüppel-associated boxes are potent transcriptional repression domains. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 4509–4513 (1994).
 211. Witzgall, R., O’Leary, E., Leaf, a, Onaldi, D. & Bonventre, J. V. The Krüppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 4514–4518 (1994).
 212. Friedman, J. R. *et al.* KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes Dev.* **10**, 2067–2078 (1996).
 213. Peng, H. *et al.* Reconstitution of the KRAB-KAP-1 repressor complex: a model system for defining the molecular anatomy of RING-B box-coiled-coil domain-mediated protein-protein interactions. *J. Mol. Biol.* **295**, 1139–62 (2000).
 214. Schultz, D. C., Friedman, J. R. & Rauscher, F. J. 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 Dev.* **15**, 428–43 (2001).
 215. Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G. & Rauscher, F. J. 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 Dev.* **16**, 919–32 (2002).
 216. Ryan, R. F. *et al.* 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. *Mol. Cell. Biol.* **19**, 4366–78 (1999).
 217. Nielsen, A. L. *et al.* Interaction with members of the heterochromatin protein 1 (HP1) family and histone deacetylation are differentially involved in transcriptional silencing

- by members of the TIF1 family. *EMBO J.* **18**, 6385–6395 (1999).
218. Sripathy, S. P., Stevens, J. & Schultz, D. C. 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. *Mol. Cell. Biol.* **26**, 8623–38 (2006).
 219. Wolf, D., Cammas, F., Losson, R. & Goff, S. P. Primer binding site-dependent restriction of murine leukemia virus requires HP1 binding by TRIM28. *J. Virol.* **82**, 4675–9 (2008).
 220. Wolf, D. & Goff, S. P. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* **458**, 1201–4 (2009).
 221. Robbez-Masson, L. & Rowe, H. M. Retrotransposons shape species-specific embryonic stem cell gene expression. *Retrovirology* **12**, 45 (2015).
 222. Rowe, H. M. *et al.* De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. *Development* **140**, 519–29 (2013).
 223. Leung, D. *et al.* Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 6690–5 (2014).
 224. Tan, S.-L. *et al.* Essential roles of the histone methyltransferase ESET in the epigenetic control of neural progenitor cells during development. *Development* **139**, 3806–3816 (2012).
 225. Fasching, L. *et al.* TRIM28 Represses Transcription of Endogenous Retroviruses in Neural Progenitor Cells. *Cell Rep.* **10**, 20–28 (2015).
 226. Collins, P. L., Kyle, K. E., Egawa, T., Shinkai, Y. & Oltz, E. M. The histone methyltransferase SETDB1 represses endogenous and exogenous retroviruses in B lymphocytes. *Proc. Natl. Acad. Sci.* **112**, 8367–72 (2015).
 227. Castro-Diaz, N. *et al.* Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev.* **28**, 1397–1409 (2014).
 228. Turelli, P. *et al.* Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Res.* (2014). doi:10.1101/gr.172833.114
 229. Leung, D. C. *et al.* Lysine methyltransferase G9a is required for de novo DNA methylation and the establishment, but not the maintenance, of proviral silencing. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 5718–5723 (2011).
 230. Di Giacomo, M., Comazzetto, S., Sampath, S. C., Sampath, S. C. & O’Carroll, D. G9a co-suppresses LINE1 elements in spermatogonia. *Epigenetics Chromatin* **7**, 24 (2014).
 231. Pua, H. H. *et al.* Novel interstitial 2.6 Mb deletion on 9q21 associated with multiple congenital anomalies. *Am. J. Med. Genet. Part A* **164**, 237–242 (2014).
 232. Au, P. Y. B. *et al.* GeneMatcher aids in the identification of a new malformation syndrome with intellectual disability, unique facial dysmorphisms, and skeletal and connective tissue caused by de novo variants in HNRNPK. *Hum. Mutat.* **36**, 1009–14 (2015).
 233. Hancarova, M. *et al.* Deletions of 9q21.3 including *NTRK2* are associated with severe

- phenotype. *Am. J. Med. Genet. Part A* **167**, 264–267 (2015).
234. Tang, F. *et al.* Downregulation of hnRNP K by RNAi inhibits growth of human lung carcinoma cells. *Oncol. Lett.* **7**, 1073–1077 (2014).
 235. Carpenter, B. *et al.* Heterogeneous nuclear ribonucleoprotein K is over expressed, aberrantly localised and is associated with poor prognosis in colorectal cancer. *Br. J. Cancer* **95**, 921–7 (2006).
 236. Barboro, P. *et al.* Heterogeneous nuclear ribonucleoprotein K: altered pattern of expression associated with diagnosis and prognosis of prostate cancer. *Br. J. Cancer* **100**, 1608–16 (2009).
 237. Maksakova, I. a *et al.* H3K9me3-binding proteins are dispensable for SETDB1/H3K9me3-dependent retroviral silencing. *Epigenetics Chromatin* **4**, 12 (2011).
 238. Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**, 805–14 (2006).
 239. Iacovino, M. *et al.* Inducible cassette exchange: A rapid and efficient system enabling conditional gene expression in embryonic stem and primary cells. *Stem Cells* **29**, 1580–1587 (2011).
 240. Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).
 241. Horie, K. *et al.* A homozygous mutant embryonic stem cell bank applicable for phenotype-driven genetic screening. *Nat. Methods* **8**, 1071–1081 (2011).
 242. Pelisch, F., Pozzi, B., Risso, G., Muñoz, M. J. & Srebrow, A. DNA damage-induced heterogeneous nuclear ribonucleoprotein K sumoylation regulates p53 transcriptional activation. *J. Biol. Chem.* **287**, 30789–99 (2012).
 243. Yang, L. *et al.* Molecular cloning of ESET , a novel histone H3-specific methyltransferase that interacts with ERG transcription factor. *Oncogene* **21**, 148–152 (2002).
 244. Ostareck, D. H. *et al.* mRNA silencing in erythroid differentiation: hnRNP K and hnRNP E1 regulate 15-lipoxygenase translation from the 3' end. *Cell* **89**, 597–606 (1997).
 245. Thompson, P. J., Dulberg, V., Moon, K. & Foster, L. J. hnRNP K Coordinates Transcriptional Silencing by SETDB1 in Embryonic Stem Cells. *PLoS Genet.* **11**, e1004933 (2015).
 246. Karimi, M. M. *et al.* DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell Stem Cell* **8**, 676–87 (2011).
 247. Tanese, N. Small-scale density gradient sedimentation to separate and analyze multiprotein complexes. *Methods* **12**, 224–34 (1997).
 248. Han, M.-H., Lin, C., Meng, S. & Wang, X. Proteomics analysis reveals overlapping functions of clustered protocadherins. *Mol. Cell. Proteomics* **9**, 71–83 (2010).

249. Podlaski, F. J. & Stern, a S. Site-specific immobilization of antibodies to protein G-derivatized solid supports. *Methods Mol. Biol.* **147**, 41–48 (2000).
250. Chan, Q. W. T., Howes, C. G. & Foster, L. J. Quantitative comparison of caste differences in honeybee hemolymph. *Mol. Cell. Proteomics* **5**, 2252–62 (2006).
251. Kristensen, A. R., Gsponer, J. & Foster, L. J. Protein synthesis rate is the predominant regulator of protein expression during differentiation. *Mol. Syst. Biol.* **9**, 689 (2013).
252. Sarge, K. D. & Park-Sarge, O.-K. in *Mol. Endocrinol. Methods Mol. Biol.* (Park-Sarge, O.-K. & Curry, T. E.) **590**, 265–277 (Humana Press, 2009).
253. Zhao, J., Sun, B. K., Erwin, J. a, Song, J.-J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756 (2008).
254. Schlesinger, S., Lee, A. H., Wang, G. Z., Green, L. & Goff, S. P. Proviral silencing in embryonic cells is regulated by Yin Yang 1. *Cell Rep.* **4**, 50–8 (2013).
255. Fietze, S., O’Geen, H., Blahnik, K. R., Jin, V. X. & Farnham, P. J. ZNF274 recruits the histone methyltransferase SETDB1 to the 3’ ends of ZNF genes. *PLoS One* **5**, (2010).
256. Nishitsuji, H., Abe, M., Sawada, R. & Takaku, H. ZBRK1 represses HIV-1 LTR-mediated transcription. *FEBS Lett.* **586**, 3562–3568 (2012).
257. Sadic, D. *et al.* Atrx promotes heterochromatin formation at retrotransposons. *EMBO Rep.* **16**, 836–850 (2015).
258. Ivanov, A. V *et al.* PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing. *Mol. Cell* **28**, 823–37 (2007).
259. Lee, Y.-K., Thomas, S. N., Yang, A. J. & Ann, D. K. Doxorubicin down-regulates Kruppel-associated box domain-associated protein 1 sumoylation that relieves its transcription repression on p21WAF1/CIP1 in breast cancer MCF-7 cells. *J. Biol. Chem.* **282**, 1595–606 (2007).
260. Sarraf, S. a & Stancheva, I. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol. Cell* **15**, 595–605 (2004).
261. Li, X. *et al.* Role for KAP1 serine 824 phosphorylation and sumoylation/desumoylation switch in regulating KAP1-mediated transcriptional repression. *J. Biol. Chem.* **282**, 36177–89 (2007).
262. Garvin, A. J. *et al.* The deSUMOylase SENP7 promotes chromatin relaxation for homologous recombination DNA repair. *EMBO Rep.* **14**, 975–83 (2013).
263. Tang, F. *et al.* Deterministic and stochastic allele specific gene expression in single mouse blastomeres. *PLoS One* **6**, e21208 (2011).
264. Bilodeau, S., Kagey, M. H., Frampton, G. M., Rahl, P. B. & Young, R. a. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev.* **23**, 2484–9 (2009).
265. Orlov, S. V. *et al.* Novel repressor of the human FMR1 gene - Identification of p56 human (GCC)_n-binding protein as a Kruppel-like transcription factor ZF5. *FEBS J.*

- 274**, 4848–4862 (2007).
266. Ho, L. *et al.* esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nat. Cell Biol.* **13**, 903–13 (2011).
267. Nicholson, T. B., Chen, T. & Richard, S. The physiological and pathophysiological role of PRMT1-mediated protein arginine methylation. *Pharmacol. Res.* **60**, 466–474 (2009).
268. Li, N. *et al.* Structure of the eukaryotic MCM complex at 3.8 Å. *Nature* **524**, 186–191 (2015).
269. Franz, C. *et al.* Nup155 regulates nuclear envelope and nuclear pore complex formation in nematodes and vertebrates. *EMBO J.* **24**, 3519–3531 (2005).
270. Kajiro, M. *et al.* The E3 Ubiquitin Ligase Activity of Trip12 Is Essential for Mouse Embryogenesis. *PLoS One* **6**, e25871 (2011).
271. Seki, Y., Kurisaki, A., Watanabe-susaki, K., Nakajima, Y. & Nakanishi, M. TIF1 β regulates the pluripotency of embryonic stem cells in a phosphorylation-dependent manner. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 10926–31 (2010).
272. Ho, L. *et al.* An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 5181–6 (2009).
273. Calo, E. *et al.* RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. *Nature* **518**, 249–253 (2015).
274. Watkins, N. J. *et al.* Assembly and Maturation of the U3 snoRNP in the Nucleoplasm in a Large Dynamic Multiprotein Complex. *Mol. Cell* **16**, 789–798 (2004).
275. Grange, T., de Sa, C. M., Oddos, J. & Pictet, R. Human mRNA polyadenylate binding protein: evolutionary conservation of a nucleic acid binding motif. *Nucleic Acids Res.* **15**, 4771–4787 (1987).
276. Durut, N. & Sáez-Vásquez, J. Nucleolin: Dual roles in rDNA chromatin transcription. *Gene* **556**, 7–12 (2015).
277. Reichman, T. W., Mun, L. C. & Mathews, M. B. The RNA Binding Protein Nuclear Factor 90 Functions as Both a Positive and Negative Regulator of Gene Expression in Mammalian Cells. *Society* **22**, 343–356 (2002).
278. Chen, T., Sun, Y., Ji, P., Kopetz, S. & Zhang, W. Topoisomerase II α in chromosome instability and personalized cancer therapy. *Oncogene* 4019–4031 (2014). doi:10.1038/onc.2014.332
279. Nano, N. & Houry, W. a. Chaperone-like activity of the AAA+ proteins Rvb1 and Rvb2 in the assembly of various complexes. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368**, 20110399 (2013).
280. Denisenko, O. N., O'Neill, B., Ostrowski, J., Van Seuning, I. & Bomsztyk, K. Zik1, a transcriptional repressor that interacts with the heterogeneous nuclear ribonucleoprotein particle K protein. *J. Biol. Chem.* **271**, 27701–6 (1996).
281. Tomanaga, T. & Levens, D. Heterogeneous nuclear ribonucleoprotein K is a DNA

- binding transactivator. *J. Biol. Chem.* **270**, 27887–93 (1995).
282. Dejgaard, K. & Leffers, H. Characterisation of the nucleic-acid-binding activity of KH domains. Different properties of different domains. *Eur. J. Biochem.* **241**, 425–31 (1996).
283. Yuan, P. *et al.* Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. *Genes Dev.* **23**, 2507–20 (2009).
284. Matunis, M. J., Michael, W. M. & Dreyfuss, G. Characterization and primary structure of the poly(C)-binding heterogeneous nuclear ribonucleoprotein complex K protein. *Mol. Cell. Biol.* **12**, 164–71 (1992).
285. Germain-Desprez, D., Bazinet, M., Bouvier, M. & Aubry, M. Oligomerization of transcriptional intermediary factor 1 regulators and interaction with ZNF74 nuclear matrix protein revealed by bioluminescence resonance energy transfer in living cells. *J. Biol. Chem.* **278**, 22367–73 (2003).
286. Ostrowski, J. *et al.* Purification , Cloning , and Expression of a Murine Phosphoprotein That Binds the KB Motif in Vitro Identifies It as the Homolog of the K Protein Human Heterogeneous Nuclear Ribonucleoprotein. *J. Biol. Chem.* **269**, 17626–17634 (1994).
287. Lee, S. W. *et al.* SUMOylation of hnRNP-K is required for p53-mediated cell-cycle arrest in response to DNA damage. *EMBO J.* **31**, 4441–52 (2012).
288. Mahajan, R., Gerace, L. & Melchior, F. Molecular characterization of the SUMO-1 modification of RanGAP1 and its role in nuclear envelope association. *J. Cell Biol.* **140**, 259–70 (1998).
289. Rodriguez, M. S. *et al.* SUMO-1 modification activates the transcriptional response of p53. *EMBO J.* **18**, 6455–61 (1999).
290. Yang, B. *et al.* MAGE-A, mMAGE-b, and MAGE-C proteins form complexes with KAP1 and suppress p53-dependent apoptosis in MAGE-positive cell lines. *Cancer Res.* **67**, 9954–62 (2007).
291. Li, X. *et al.* SUMOylation of the transcriptional co-repressor KAP1 is regulated by the serine and threonine phosphatase PP1. *Sci. Signal.* **3**, ra32 (2010).
292. Dinh, P. X., Das, A., Franco, R. & Pattnaik, A. K. Heterogeneous nuclear ribonucleoprotein K supports vesicular stomatitis virus replication by regulating cell survival and cellular gene expression. *J. Virol.* **87**, 10059–69 (2013).
293. Fujikura, J. *et al.* Differentiation of embryonic stem cells is induced by GATA factors. *Genes Dev.* **16**, 784–9 (2002).
294. Zhang, C., Ye, X., Zhang, H., Ding, M. & Deng, H. GATA factors induce mouse embryonic stem cell differentiation toward extraembryonic endoderm. *Stem Cells Dev.* **16**, 605–13 (2007).
295. Cheng, B., Ren, X. & Kerppola, T. K. KAP1 represses differentiation-inducible genes in embryonic stem cells through cooperative binding with PRC1 and derepresses pluripotency-associated genes. *Mol. Cell. Biol.* **34**, 2075–91 (2014).
296. Mikula, M., Bomsztyk, K., Goryca, K., Chojnowski, K. & Ostrowski, J. Heterogeneous nuclear ribonucleoprotein (HnRNP) K genome-wide binding survey

- reveals its role in regulating 3'-end RNA processing and transcription termination at the early growth response 1 (EGR1) gene through XRN2 exonuclease. *J. Biol. Chem.* **288**, 24788–98 (2013).
297. Mikula, M. & Bomsztyk, K. Direct recruitment of ERK cascade components to inducible genes is regulated by heterogeneous nuclear ribonucleoprotein (hnRNP) K. *J. Biol. Chem.* **286**, 9763–75 (2011).
 298. Zeng, L. *et al.* Structural insights into human KAP1 PHD finger-bromodomain and its role in gene silencing. *Nat. Struct. Mol. Biol.* **15**, 626–33 (2008).
 299. Fukuda, I. *et al.* Ginkgolic acid inhibits protein SUMOylation by blocking formation of the E1-SUMO intermediate. *Chem. Biol.* **16**, 133–40 (2009).
 300. Balasubramanyam, K., Swaminathan, V., Ranganathan, A. & Kundu, T. K. Small molecule modulators of histone acetyltransferase p300. *J. Biol. Chem.* **278**, 19134–40 (2003).
 301. Tan, X. *et al.* Zfp819, a novel KRAB-zinc finger protein, interacts with KAP1 and functions in genomic integrity maintenance of mouse embryonic stem cells. *Stem Cell Res.* **11**, 1045–59 (2013).
 302. Charroux, B., Angelats, C., Fasano, L., Kerridge, S. & Vola, C. The levels of the bancal product, a Drosophila homologue of vertebrate hnRNP K protein, affect cell proliferation and apoptosis in imaginal disc cells. *Mol. Cell. Biol.* **19**, 7846–56 (1999).
 303. Denisenko, O. & Bomsztyk, K. Yeast hnRNP K-Like Genes Are Involved in Regulation of the Telomeric Position Effect and Telomere Length Yeast hnRNP K-Like Genes Are Involved in Regulation of the Telomeric Position Effect and Telomere Length. *Mol. Cell. Biol.* **22**, 286–97 (2002).
 304. Denisenko, O. N. & Bomsztyk, K. The product of the murine homolog of the Drosophila extra sex combs gene displays transcriptional repressor activity . The Product of the Murine Homolog of the Drosophila extra sex combs Gene Displays Transcriptional Repressor Activity. *Mol. Cell. Biol.* **17**, 4707–4717 (1997).
 305. Howarth, M. M. *et al.* Long noncoding RNA EWSAT1 -mediated gene repression facilitates Ewing sarcoma oncogenesis. *J Clin Invest* **124**, 5275–5290 (2014).
 306. Bao, X. *et al.* The p53-induced lincRNA-p21 derails somatic cell reprogramming by sustaining H3K9me3 and CpG methylation at pluripotency gene promoters. *Cell Res.* **25**, 80–92 (2014).
 307. Ichimura, T. *et al.* Transcriptional repression and heterochromatin formation by MBD1 and MCAF/AM family proteins. *J. Biol. Chem.* **280**, 13928–35 (2005).
 308. Neyret-Kahn, H. *et al.* Sumoylation at chromatin governs coordinated repression of a transcriptional program essential for cell growth and proliferation. *Genome Res.* **23**, 1563–79 (2013).
 309. Iyengar, S., Ivanov, A. V, Jin, V. X., Rauscher, F. J. & Farnham, P. J. Functional analysis of KAP1 genomic recruitment. *Mol. Cell. Biol.* **31**, 1833–47 (2011).
 310. Kuo, C.-Y. *et al.* An Arginine-rich Motif of Ring Finger Protein 4 (RNF4) Oversees the Recruitment and Degradation of the Phosphorylated and SUMOylated Krüppel-

- associated Box Domain-associated Protein 1 (KAP1)/TRIM28 Protein during Genotoxic Stress. *J. Biol. Chem.* **289**, 20757–72 (2014).
311. Yang, B. X. *et al.* Systematic Identification of Factors for Provirus Silencing in Embryonic Stem Cells. *Cell* 1–16 (2015). doi:10.1016/j.cell.2015.08.037
 312. Koch, C. M., Honemann-Capito, M., Egger-Adam, D. & Wodarz, A. Windei, the Drosophila homolog of mAM/MCAF1, is an essential cofactor of the H3K9 methyl transferase dSETDB1/Eggless in germ line development. *PLoS Genet.* **5**, e1000644 (2009).
 313. Kim, S. *et al.* PRMT5 Protects Genomic Integrity during Global DNA Demethylation in Primordial Germ Cells and Preimplantation Embryos. *Mol. Cell* **5**, 564–579 (2014).
 314. Tchasovnikarova, I. A. *et al.* Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science* **348**, 1481–1485 (2015).
 315. Kokura, K., Sun, L., Bedford, M. T. & Fang, J. Methyl-H3K9-binding protein MPP8 mediates E-cadherin gene silencing and promotes tumour cell motility and invasion. *EMBO J.* **29**, 3673–3687 (2010).
 316. Maeda, I. *et al.* Max is a repressor of germ cell-related gene expression in mouse embryonic stem cells. *Nat. Commun.* **4**, 1754 (2013).
 317. Ribet, D. *et al.* Murine endogenous retrovirus MuERV-L is the progenitor of the ‘orphan’ epsilon viruslike particles of the early mouse embryo. *J. Virol.* **82**, 1622–1625 (2008).
 318. Schoorlemmer, J., Pérez-Palacios, R., Climent, M., Guallar, D. & Muniesa, P. Regulation of Mouse Retroelement MuERV-L/MERVL Expression by REX1 and Epigenetic Control of Stem Cell Potency. *Front. Oncol.* **4**, 14 (2014).
 319. Peaston, A. E. *et al.* Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* **7**, 597–606 (2004).
 320. Kigami, D. MuERV-L Is One of the Earliest Transcribed Genes in Mouse One-Cell Embryos. *Biol. Reprod.* **68**, 651–654 (2002).
 321. Macfarlan, T. S. *et al.* Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev.* **25**, 594–607 (2011).
 322. Kigami, D., Minami, N., Takayama, H. & Imai, H. MuERV-L Is One of the Earliest Transcribed Genes in Mouse One-Cell Embryos. *Biol. Reprod.* **68**, 651–654 (2002).
 323. Gifford, W. D., Pfaff, S. L. & MacFarlan, T. S. Transposable elements as genetic regulatory substrates in early development. *Trends Cell Biol.* **23**, 218–226 (2013).
 324. Guallar, D. *et al.* Expression of endogenous retroviruses is negatively regulated by the pluripotency marker Rex1/Zfp42. *Nucleic Acids Res.* **40**, 8993–9007 (2012).
 325. Foster, C. T. *et al.* Lysine-specific demethylase 1 regulates the embryonic transcriptome and CoREST stability. *Mol. Cell. Biol.* **30**, 4851–4863 (2010).
 326. Hisada, K. *et al.* RYBP Represses Endogenous Retroviruses and Preimplantation- and Germ Line-Specific Genes in Mouse Embryonic Stem Cells. *Mol. Cell. Biol.* **32**, 1139–1149 (2012).

327. Moumen, A., Masterson, P., O'Connor, M. J. & Jackson, S. P. hnRNP K: an HDM2 target and transcriptional coactivator of p53 in response to DNA damage. *Cell* **123**, 1065–78 (2005).
328. Ritchie, S. a. *et al.* Identification of the SRC pyrimidine-binding protein (SPy) as hnRNP K: Implications in the regulation of SRC1A transcription. *Nucleic Acids Res.* **31**, 1502–1513 (2003).
329. Uribe, D. J., Guo, K., Shin, Y. J. & Sun, D. Heterogeneous nuclear ribonucleoprotein K and nucleolin as transcriptional activators of the vascular endothelial growth factor promoter through interaction with secondary DNA structures. *Biochemistry* **50**, 3796–3806 (2011).
330. Da Silva, N., Bharti, A. & Shelley, C. S. hnRNP-K and Pur(alpha) act together to repress the transcriptional activity of the CD43 gene promoter. *Blood* **100**, 3536–44 (2002).
331. Minajigi, A. *et al.* A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349**, aab2276 (2015).
332. Lienert, F. *et al.* Genomic Prevalence of Heterochromatic H3K9me2 and Transcription Do Not Discriminate Pluripotent from Terminally Differentiated Cells. *PLoS Genet.* **7**, e1002090 (2011).
333. Titov, D. V *et al.* XPB, a subunit of TFIIH, is a target of the natural product triptolide. *Nat. Chem. Biol.* **7**, 182–188 (2011).
334. Hu, W., Begum, N. a., Mondal, S., Stanlie, A. & Honjo, T. Identification of DNA cleavage- and recombination-specific hnRNP cofactors for activation-induced cytidine deaminase. *Proc. Natl. Acad. Sci.* **112**, 5791–6 (2015).
335. Svoboda, P. *et al.* RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev. Biol.* **269**, 276–285 (2004).
336. Braddock, D. T., Baber, J. L., Levens, D. & Clore, G. M. Molecular basis of sequence-specific single-stranded DNA recognition by KH domains : solution structure of a complex between hnRNP K KH3 and single-stranded DNA. *EMBO J.* **21**, 3476–85 (2002).
337. Backe, P. H., Messias, A. C., Ravelli, R. B. G., Sattler, M. & Cusack, S. X-ray crystallographic and NMR studies of the third KH domain of hnRNP K in complex with single-stranded nucleic acids. *Structure* **13**, 1055–10607 (2005).
338. Messias, A. C., Harnisch, C., Ostareck-Lederer, A., Sattler, M. & Ostareck, D. H. The DICE-binding Activity of KH Domain 3 of hnRNP K Is Affected by c-Src-mediated Tyrosine Phosphorylation. *J. Mol. Biol.* **361**, 470–481 (2006).
339. Kim, J. H., Hahm, B., Kim, Y. K., Choi, M. & Jang, S. K. Protein-protein interaction among hnRNPs shuttling between nucleus and cytoplasm. *J. Mol. Biol.* **298**, 395–405 (2000).
340. Baber, J. L., Libutti, D., Levens, D. & Tjandra, N. High precision solution structure of the C-terminal KH domain of heterogeneous nuclear ribonucleoprotein K, a c-myc transcription factor. *J. Mol. Biol.* **289**, 949–962 (1999).

341. Akiyama, T. *et al.* Transient bursts of Zscan4 expression are accompanied by the rapid derepression of heterochromatin in mouse embryonic stem cells. *DNA Res.* **22**, 307–318 (2015).
342. Moore, K. E. *et al.* A general molecular affinity strategy for global detection and proteomic analysis of lysine methylation. *Mol. Cell* **50**, 444–456 (2013).
343. Ishiuchi, T. *et al.* Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat. Struct. Mol. Biol.* **22**, 662–71 (2015).
344. Liu, H. W. *et al.* Chromatin modification by SUMO-1 stimulates the promoters of translation machinery genes. *Nucleic Acids Res.* **40**, 10172–10186 (2012).
345. Yeap, L.-S., Hayashi, K. & Surani, M. A. ERG-associated protein with SET domain (ESET)-Oct4 interaction regulates pluripotency and represses the trophectoderm lineage. *Epigenetics Chromatin* **2**, 12 (2009).
346. Cho, S., Park, J. S. & Kang, Y.-K. AGO2 and SETDB1 cooperate in promoter-targeted transcriptional silencing of the androgen receptor gene. *Nucleic Acids Res.* **42**, 13545–13556 (2014).
347. Shiio, Y. & Eisenman, R. N. Histone sumoylation is associated with transcriptional repression. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 13225–30 (2003).
348. Liang, Q. *et al.* Tripartite motif-containing protein 28 is a small ubiquitin-related modifier E3 ligase and negative regulator of IFN regulatory factor 7. *J. Immunol.* **187**, 4754–63 (2011).
349. Neo, S. H. *et al.* TRIM28 Is an E3 Ligase for ARF-Mediated NPM1/B23 SUMOylation That Represses Centrosome Amplification. *Mol. Cell. Biol.* **35**, 2851–2863 (2015).
350. Rebollo, R. *et al.* Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. *PLoS Genet.* **7**, (2011).
351. Cooper, G. M. *et al.* A copy number variation morbidity map of developmental delay. *Nat. Genet.* **43**, 838–846 (2011).
352. Benevento, M., van de Molengraft, M., van Westen, R., van Bokhoven, H. & Nadif Kasri, N. The role of chromatin repressive marks in cognition and disease: A focus on the repressive complex GLP/G9a. *Neurobiol. Learn. Mem.* **124**, 88–96 (2015).
353. Kleefstra, T. *et al.* Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. *Am. J. Hum. Genet.* **91**, 73–82 (2012).
354. Folci, A. *et al.* Loss of hnRNP K Impairs Synaptic Plasticity in Hippocampal Neurons. *J. Neurosci.* **34**, 9088–95 (2014).
355. Proepper, C. *et al.* Heterogeneous nuclear ribonucleoprotein K interacts with Abi-1 at postsynaptic sites and modulates dendritic spine morphology. *PLoS One* **6**, (2011).
356. Gallardo, M. *et al.* hnRNP K Is a Haploinsufficient Tumor Suppressor that Regulates Proliferation and Differentiation Programs in Hematologic Malignancies. *Cancer Cell* **28**, 486–99 (2015).
357. Balemans, M. C. M. *et al.* Hippocampal dysfunction in the Euchromatin histone methyltransferase 1 heterozygous knockout mouse model for Kleefstra syndrome.

- Hum. Mol. Genet.* **22**, 852–866 (2013).
358. Kronke, J. *et al.* Clonal evolution in relapsed NPM1-mutated acute myeloid leukemia. *Blood* **122**, 100–108 (2013).
 359. Lehnertz, B. *et al.* The methyltransferase G9a regulates HoxA9-dependent transcription in AML. *Genes Dev.* **28**, 317–327 (2014).
 360. Barboro, P., Ferrari, N. & Balbi, C. Emerging roles of heterogeneous nuclear ribonucleoprotein K (hnRNP K) in cancer progression. *Cancer Lett.* **352**, 152–159 (2014).
 361. Wen, F. *et al.* Higher expression of the heterogeneous nuclear ribonucleoprotein k in melanoma. *Ann. Surg. Oncol.* **17**, 2619–27 (2010).
 362. Du, Q. *et al.* The role of heterogeneous nuclear ribonucleoprotein K in the progression of chronic myeloid leukemia. *Med. Oncol.* **27**, 673–679 (2010).
 363. Mandal, M. *et al.* Growth factors regulate heterogeneous nuclear ribonucleoprotein K expression and function. *J. Biol. Chem.* **276**, 9699–9704 (2001).
 364. Ostrowski, J. & Bomsztyk, K. Nuclear shift of hnRNP K protein in neoplasms and other states of enhanced cell proliferation. *Br. J. Cancer* **89**, 1493–1501 (2003).
 365. Lynch, M. *et al.* hnRNP K Binds a Core Polypyrimidine Element in the Eukaryotic Translation Initiation Factor 4E (eIF4E) Promoter , and Its Regulation of eIF4E Contributes to Neoplastic Transformation. *Mol. Cell. Biol.* **25**, 6436–6453 (2005).
 366. Takimoto, M. *et al.* Specific Binding of Heterogeneous Ribonucleoprotein Particle Protein K to the Human c-myc Promoter , in Vitro *. *J. Biol. Chem.* **268**, 18249–18258 (1993).
 367. Chen, M. W. *et al.* H3K9 histone methyltransferase G9a promotes lung cancer invasion and metastasis by silencing the cell adhesion molecule Ep-CAM. *Cancer Res.* **70**, 7830–7840 (2010).