Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

presented by

Teresa Pankert, M.Sc.

born in Aachen, Germany

Oral examination: 04.02.2016

A 3’-UTR sequence element modifies chromatin and regulates alternative splicing

Referees: PD Dr. Karsten Rippe

PD Dr. Georg Stöcklin

Declaration

I hereby declare that I have written the submitted dissertation “A 3’-UTR sequence element modifies chromatin and regulates alternative splicing” myself and in this process have used no other sources or materials than those explicitly indicated. I hereby declare that I have not applied to be examined at any other institution, nor have I used the dissertation in this or any other form at any other institution as an examination paper, nor submitted it to any other faculty as a dissertation.

_____ _____

(place, date) Teresa Pankert

This work was carried out under the scientific guidance of PD Dr. Karsten Rippe in the research group “Genome Organization and Function” at the German Cancer Research Center (DKFZ) and the BioQuant Center in Heidelberg, Germany, from January 2012 to November 2015.

During the first three years I was supported by the graduate scholarship program of the Cusanuswerk, Bonn, Germany.

Any living cell carries with it the experiences of a billion years of experimentation by its ancestors. You cannot expect to explain so wise an old bird in a few simple words.

Max Delbrück, “A Physicist Looks at Biology”, 1949

Table of Contents

List of publications

The following publication was generated in the course of this thesis:

Caudron-Herger M, **Pankert T**, Seiler J, Nemeth A, Voit R, Grummt I, Rippe K (2015) Alu element-containing RNAs maintain nucleolar structure and function. *EMBO J*

Summary

The 3'-untranslated regions (3'-UTRs) of protein coding transcripts play a well-established role as *cis* regulators of RNA stability and protein translation. Recent findings also point to functions in the nucleus. This thesis reports on the analysis of a specific 3'-UTR, referred to here as CU-RNA, originating from the *CDV3* gene (carnitine deficiency-associated gene expressed in ventricle 3) that was previously identified as a 3'-UTR with a putative nuclear function. In an initial characterization of CU-RNA it was shown that it has nuclear functions due to the following criteria: (i) the transcript was able to rescue RNase A-induced chromatin aggregation, (ii) the transcript remained within the nucleus after microinjection, (iii) the endogenous transcript was present in the nucleus, (iv) by immobilizing the RNA at a specific genomic locus it was shown that locally tethered CU-RNA induces chromatin compaction, demonstrating its chromatin modifying activity. This chromatin recruitment system was subsequently used to investigate the mechanism of action of CU-RNA. It was found that CU-RNA derived fragments recruit enhancer of zeste homolog 2 (EZH2), the enzyme that sets trimethylation of histone H3 at lysine 27 (H3K27me3). By deletion analysis a CU-RNA sequence element, CU-RNA-T0, was identified that was necessary to also induce H3K27me3. The splicing regulators PTBP1 (polypyrimidine tract binding protein 1) and U2AF2 (U2 small nuclear RNA auxiliary factor 2), and the H3K27me3-promoting factor HNRNPK (heterogeneous nuclear ribonucleoprotein K) were found to specifically interact with CU-RNA-T0. In contrast, an alternative splicing variant of *CDV3* that lacks the T0 element (CU-RNA-ΔT0) and did not induce H3K27me3, bound the splice factor HNRNPC (heterogeneous nuclear ribonucleoprotein C). It was therefore investigated whether the H3K27me3 inducing activity of CU-RNA-T0 affects splicing. Reducing the H3K27me3 level at the *CDV3* locus by inhibition of EZH2 enzymatic activity indeed changed the ratio of the *CDV3* transcript variant levels. Based on these results a model is proposed, in which the CU-RNA-T0 element regulates alternative splicing of its own transcript, the *CDV3*, via a feedback loop that establishes a specific chromatin signature. Thereby stable steady-state levels between the splicing variants are regulated that might be distinct with respect to their 3'-UTR sequence dependent cytoplasmic activities. This contribution to the understanding of the mechanism governing alternative splicing is of great importance to be able to develop novel drugs targeting deregulated alternative splicing in cancer.

Zusammenfassung

Die 3'-untranslatierten Regionen (3'-UTRs) von proteinkodierenden Transkripten spielen eine wichtige Rolle als *cis*-regulatorische Elemente von RNA-Stabilität und Proteintranslation. Neuere Erkenntnisse weisen darauf hin, dass sie darüber hinaus auch Funktionen im Zellkern haben könnten. In dieser Arbeit wurde eine bestimmte 3'-UTR, genannt CU-RNA, analysiert, die vom *CDV3* Gen (carnitine deficiency-associated gene expressed in ventricle 3) stammt, und in vorhergehenden Publikationen als 3'-UTR mit vermeintlich nukleärer Funktion identifiziert wurde. In einer anfänglichen Charakterisierung wurde anhand der folgenden Kriterien festgestellt, dass die CU-RNA eine Funktion im Zellkern hat: Erstens konnte das Transkript RNase A-induzierter Chromatinverdichtung entgegenwirken, zweitens lokalisierte es nach Mikroinjektion im Zellkern, drittens konnte das endogene Transkript im Zellkern nachgewiesen werden und viertens führte die Rekrutierung von CU-RNA an einen bestimmten genomischen Lokus zu Chromatinverdichtung, was seine Fähigkeit Chromatin zu modifizieren demonstrierte. Dieses Rekrutierungssystem wurde im Weiteren dazu verwendet den Wirkmechanismus der CU-RNA zu untersuchen. Dabei wurde herausgefunden, dass Fragmente der CU-RNA das Enzym EZH2 (enhancer of zeste homolog 2) rekrutieren, welches für die Trimethylierung des Lysins an Position 27 von Histon 3 (H3K27me3) verantwortlich ist. Deletionsstudien zeigten, dass ein spezifisches Element der CU-RNA, CU-RNA-T0, notwendig ist um auch H3K27me3 am Chromatin zu induzierten. Die Spleißfaktoren PTBP1 (polypyrimidine tract binding protein 1) und U2AF2 (U2 small nuclear RNA auxiliary factor 2) und der H3K27me3-fördernde Faktor HNRNPK (heterogeneous nuclear ribonucleoprotein K) wurden zudem als CU-RNA-T0 interagierende Proteine identifiziert. Eine alternative Spleißvariante des *CDV3* Gens, die das T0 Element nicht enthält (CU-RNA-ΔT0) und nicht H3K27me3 induzierte, interagierte hingegen mit dem Spleißfaktor HNRNPC (heterogeneous nuclear ribonucleoprotein C). Folglich wurde untersucht, ob die H3K27me3-fördernde Aktivität von CU-RNA-T0 alternatives Spleißen beeinflusst. Tatsächlich verschob sich das Verhältnis der CU-RNA Transkriptvarianten, wenn die H3K27me3 Level mittels EZH2-Inhibitierung verringert wurden. Basierend auf diesen Ergebnissen wird ein Model aufgestellt, in welchem das CU-RNA-T0 Element das alternative Spleißen des eigenen Transkripts, *CDV3*, reguliert, indem es eine spezifische Chromatinsignatur etabliert. Dadurch wird das „steady-state" Verhältnis der Transkriptvarianten reguliert, die sich zum Beispiel im Bezug auf ihre 3'-UTR

sequenzabhängigen zytoplasmatischen Aktivitäten unterscheiden können. Dieser Beitrag zum Verständnis des Mechanismus zur Regulation von alternativem Spleißen ist von großer Bedeutung, um gezielt Medikamente gegen dereguliertes alternatives Spleißen in Krebszellen zu entwickeln.

Abbreviations

AB	antibody
Ac	acetylation
ACTB	β-actin
AEBP2	adipocyte enhancer-binding protein 2
AGO1	argonaute RISC catalytic component 1
aluRNA	RNA originating from intronic Alu elements
AM	Active Motif
AO3	CHO cell line with a single, stable *lac*O array integration
AP	alkaline phosphatase
ARE	AU-rich element
as	anti-sense
ATP	adenosine triphosphate
ATXR5	Arabidopsis trithorax-related proteins 5
ATXR6	Arabidopsis trithorax-related proteins 6
BCA	bicinchoninic acid assay
BIV	bovine immunodeficiency virus
BSA	bovine serum albumine
Cas9	CRISPR associated protein 9
cDNA	complementary DNA
CDV3	carnitine deficiency-associated gene expressed in ventricle 3
ChIP	chromatin immunoprecipitation
ChIP-seq	chromatin immunoprecipitation and sequencing
CHO	Chinese hamster ovary
CI	confidence interval
CLSM	confocal laser scanning fluorescence microscopy
CORO1C	coronin, actin binding protein, 1C
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSK	cytoskeleton
CTP	cytidine triphosphate
CU-RNA	CDV3-3'-UTR
CU-RNA-T0	CDV3-3'-UTR, T0 fragment
CU-RNA-ΔT0	CDV3-3'-UTR, ΔT0 fragment
DAPI	4',6-diamidino-2-phenylindole
DDX17	DEAD (Asp-Glu-Ala-Asp) box helicase 17
DDX21	DEAD (Asp-Glu-Ala-Asp) box helicase 21
DH5α	chemocompetent *E. Coli* strain
DIG	digoxigenin
DMEM	Dulbecco's modified eagle medium
DMEM/F12	Dulbecco's modified eagle medium nutrient mixture F-12
DMSO	dimethylsulfoxid
DNA	deoxyribonucleic acid
Dscam	Down syndrome cell adhesion molecule 1
dsDNA	double stranded deoxyribonucleic acid
E. Coli	*Escherichia Coli*
EBER2	Epstein-Barr virus encoded RNA 2
EC4	U2OS cell line with a single, stable, telomeric *lac*O array integration
EDTA	ethylenediaminetetraacetic acid
EED	embryonic ectoderm development
EGTA	ethylene glycol tetraacetic acid
EIF2S3	eukaryotic translation initiation factor 2, subunit 3 gamma
EIF4B	eukaryotic translation initiation factor 4B
ESE	exonic splicing enhancers
EtOH	ethanol
EU	ethynyl uridine

EZH2	enhancer of zeste 2
F4IIB8	U2OS cell line with a single, stable, centromeric *lac*O array integration
FACS	fluorescence-activated cell sorting
FADS1	fatty acid desaturase 1
FAM83A	family with sequence similarity 83, member A
FBS	fetal bovine serum
fg	femto grams
FGFR2	fibroblast growth factor receptor 2
FISH	fluorescence *in situ* hybridiation
FL	full length
fl	femto liters
G9a	lysine N-Methyltransferase
GFP	green fluorescent protein
GO	gene ontology
GSK343	GlaxoSmithKline343
GTP	guanosine triphosphate
h	hours
H2AK119	histone H2A lysine 119
H2BK5me1	histone H2B lysine 5 monomethylation
H3	histone H3
H3.3K27M	H3.3 with lysine 27-to-methionine mutation
H3F3A	H3 histone, family 3A
H3K27me1	histone H3 lysine 27 monomethylation
H3K27me2	histone H3 lysine 27 dimethylation
H3K27me3	histone H3 lysine 27 trimethylation
H3K36me3	histone H3 lysine 36 trimethylation
H3K4me1	histone H3 lysine 4 monomethylation
H3K79me1	histone H3 lysine 79 monomethylation
H3K9me3	histone H3 lysine 9 trimethlyation
HBiX1	heterochromatin protein 1-binding protein
HCl	hydrochloric acid
HDAC1	histone deacetylase 1
HeLa	Henrietta Lacks
Her2/neu	human epidermal growth factor receptor 2
HNRNPC	heterogeneous nuclear ribonucleoprotein C
HNRNPK	heterogeneous nuclear ribonucleoprotein K
HOTAIR	HOX (homeobox protein) transcript antisense RNA
HP1	heterochromatin protein 1
HRP	horseradish peroxidase
ICeChIP	internal standard calibrated ChIP
IDT	Integrated DNA technologies
IF	immunofluorescence
IgG	immunoglobulin G
IGV	integrative genome viewer
IP	immunoprecipitation
ISE	intronic splicing enhancers
ISS	intronic splicing silencers
iv txn	*in vitro* transcription
JARID2	Jumonji, AT rich interactive domain 2
JMJD	Jumonji domain containing
K	lysine
KAT	lysine acetyltransferase
kb	kilobase pairs
Kcnq1ot1	KCNQ1 (potassium channel, voltage gated KQT-like subfamily Q, member 1) overlapping transcript 1
KDM2A	lysine (K)-specific demethylase 2A
KDM6	lysine (K)-specific demethylase 6
KMT	lysine methyltransferase

L1	long interspersed repetitive sequence 1
LacI	lactose operon repressor
*lac*O	lactose operon
LB	Luria Bertani
lncRNA	long non-coding ribonucleic acid
M	molar
M	micro molar
MCORE	multi-scale correlation evaluation
Me	methlyation
min	minutes
miRNA	microRNA
mM	milli molar
MOPS	morpholinepropanesulfonic acid
MRG15	mortality factor 4 like 1
mRNA	messenger ribonucleid acid
MS2	bacteriophage MS2
NaP	sodium phosphate
NB	northern blot
NCL	nucleolin
ncRNA	non-coding ribonucleic acid
NEB	New England Biolabs
ng	nano gram
NLS	nuclear localization signal
nm	nano meter
NPM	nucleophosmin
nt	nucleotides
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PI	propidium iodide
PlaB	pladienolide B
PRC2	polycomb repressive complex 2
pre-mRNA	precursor ribonucleid acid
PTBP1	polypyrimidine tract binding protein 1
PTTG1IP	pituitary tumor-transforming 1 interacting protein
qRT-PCR	quantitative real time polymerase chain reaction
qRT-PCR	quantitative real time polymerase chain reaction
RBAP48	retinoblastoma binding protein 4
RBP	RNA-binding protein
rDNA	ribosomal DNA
RepA	repeat A
RIPA	Radioimmunoprecipitation assay buffer
RISC	RNA induced silencing complex
RNA	ribonucleic acid
RNA Pol II	RNA polymerase II
RPMI	Roswell Park Memorial Institute
rRNA	ribosomal ribonucleic acid
RT-PCR	reverse transcription polymerase chain reaction
SAH	S-adenosyl-homocysteine
SAM	S-adenosylmethionin
sc	Stanta Cruz
SDS	sodium dodecyl sulfate
sec	seconds
SILAC	stable isotope labeling by amino acids in cell culture
siRNA	small interfering RNA
SMCHD1	structural maintenance of chromosomes hinge domain–containing protein 1
SNF2L	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 1
SNP	small nucleotide polymorphism

SNP	small nucleotide polymorphism
snRNA	small nuclear RNA
snRNP	small nuclear ribonulceoprotein
SOC	super optimal broth with catabolite repression
SSC	sodium chloride-sodium citrate
SSR3	signal sequence seceptor, gamma
ssRT-PCR	strand specific reverse transcriptase polymerase chain reaction
STARD7	star-related lipid transfer (START) domain containing 7
StrepII	strep-tactin II
SURE 2	chemocompetent *E. Coli* strain
SURF4	surface 4 integral membrane protein
SUZ12	suppressor of zeste 12 protein homolog
TAE	tris-acetate/EDTA buffer
TALE	Transcription Activator-like Effector Nuclease
TAR	trans-activation response
Tat	trans-activator
TBME	tert-butyl methyl ether
TBS	tris-buffered saline
TCA	trichloroacetic acid
TE	tris/EDTA buffer
TIA	T-cell-restricted intracellular antigen-1
TIAL/TIAR	T-cell-restricted intracellular antigen-1 like
tRNA	transfer RNA
U1	small nuclear ribonulceoprotein 1
U2	small nuclear ribonulceoprotein 2
U2AF1/U2AF35	U2 small nuclear RNA auxiliary factor 1
U2AF2/U2AF65	U2 small nuclear RNA auxiliary factor 2
U2OS	U-2 osteosarcoma cell line
U4	small nuclear ribonulceoprotein 4
U5	small nuclear ribonulceoprotein 5
U6	small nuclear ribonulceoprotein 6
USA	United States of America
UTP	uridine triphosphat
UTR	untranslated region
UV	ultraviolet
VRC	ribonucleoside vanadyl complex
WB	western blot
XIST	X-inactive specific transcript
ZMBH	Zentrum für Molekulare Biologie der Universität Heidelberg
µl	micro liters
µm	micro grams

Introduction

Gene expression can be regulated at many steps throughout the pathway leading from the information encoded in the DNA to the functional protein. At each of these steps, the process can be modulated to influence which particular genes are expressed and how much is expressed. Thus, cells can control transcription, RNA processing, RNA localization, RNA stability, translation, protein activity, and protein degradation. Besides its function as the messenger conveying genetic information from DNA to the ribosome, RNA is also an essential player in controlling gene expression at all levels *in cis* or *in trans* (Figure 1).

Extensive studies of the diverse roles of RNA in all of these steps have led to the identification of various RNA species that are classified according to their functions in the cell. Examples of how different types of RNAs regulate gene expression include the following: On the level of chromatin modulation, a wide range of models of how RNAs exert their function has been suggested (Rinn & Chang, 2012): RNAs have been proposed i) to be the scaffold molecules of higher order chromatin organizing ribonucleoprotein complexes (Guttman et al, 2011; Hung et al, 2011; Khalil et al, 2009; Kino et al, 2010; Kotake et al, 2011; Pandey et al, 2008; Yap et al, 2010), ii) to be at the center of chromatin modifying complexes by targeting them to chromatin and regulating their enzymatic activities (Gupta et al, 2010; Huarte et al, 2010; Rinn et al, 2007), iii) to act as decoys that titrate away DNA-binding proteins (Spitale et al, 2011; Zappulla & Cech, 2006), or iv) to foster chromosome looping (Yao et al, 2010). One famous example of a chromatin modifying RNA is the X inactivation specific transcript (XIST). It is involved in the processes of X chromosome inactivation in female mammalian cells where it plays a crucial role in the recruitment and regulation of the polycomb repressive complex 2 (PRC2) (for more details see introduction, section 2.1) (Brockdorff, 2013; Brockdorff et al, 1992). As a mode of controlling transcription RNAs have recently been proposed to assist transcription factors in finding their binding sites, as described for example for the Epstein-Barr virus encoded RNA 2 (EBER2) (Lee et al, 2015). Splicing of primary messenger RNA

(pre-mRNA) has long been known to be controlled by small nuclear RNAs (snRNAs) (Maniatis & Reed, 1987). The nuclear export of RNAs is directly linked to the process of mRNA maturation and polyadenylation thus, it also involves *cis* and *trans* acting signals encoded in RNA (Chekanova & Belostotsky, 2003; Luo & Reed, 1999). Likewise, the complex control of RNA stability and decay is influenced by *cis*-regulatory RNA elements, such as AU-rich elements (AREs) (Stoecklin & Muhlemann, 2013), and by microRNAs regulating the abundance of specific mRNAs in the cell (Valencia-Sanchez et al, 2006). In addition, translation regulation involves *cis*-regulatory RNA elements such as the 5' cap structure and the 3' poly(A) tail. Lastly, the processes of translation requires transfer RNAs (tRNAs) that are loaded with the amino acids (Gebauer & Hentze, 2004), and the ribosomal RNAs (rRNAs) as building blocks of the ribosomes (Yusupov et al, 2001).

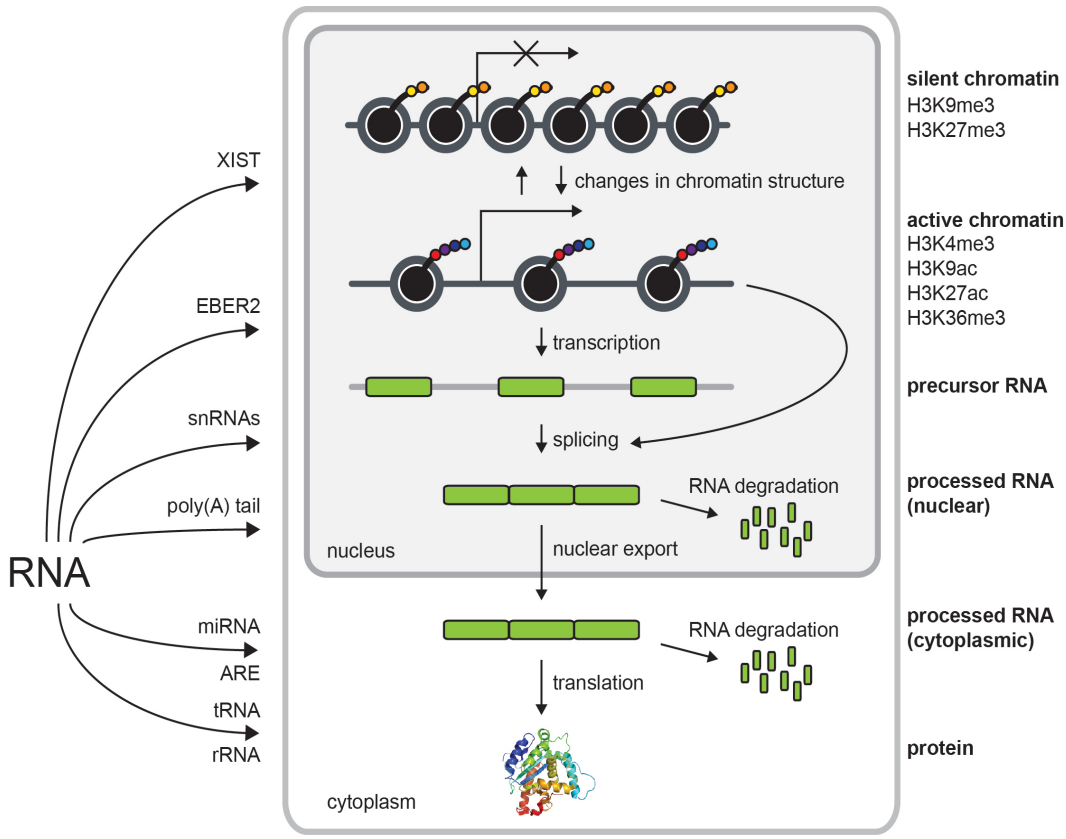


Figure 1: Role of RNA within the flow of information from DNA to protein in a eukaryotic cell. Examples of RNAs and elements within the RNAs that play a role in controlling each of these steps either *in cis* or *in trans* are indicated with the arrows on the left. XIST: X inactivation specific transcript, EBER: Epstein-Barr virus encoded RNA 2, snRNAs: small nuclear RNAs, poly(A) tail: poly-adenylated tail, miRNA: microRNAs, ARE: AU-rich elements, tRNA: transfer RNA, rRNA: ribosomal RNA.

While more and more RNA species with regulatory roles in gene expression particularly at the level of chromatin organization are identified, little is known about the chromatin-regulatory functions of the mRNA itself. This thesis focuses on the role of the 3' untranslated regions (3'-UTRs) of mRNAs in regulating chromatin organization. It is examined how 3'-UTRs can influence chromatin, and thereby how they can regulate RNA splicing and determine expression of different splice variants of the same transcript. The regulation of gene expression by chromatin modifications, the role of RNA as a regulatory factor of this process and the regulation of RNA splicing will therefore be introduced in more detail in the following.

1 Covalent modifications of histones regulate gene expression

The genomic DNA of eukaryotic cells is highly packaged into a dynamic nucleoprotein complex called chromatin (van Holde, 1989). The structure of chromatin plays an important regulatory role in defining cellular identity and function (Fisher, 2002; Hsieh & Gage, 2004; Jaenisch & Bird, 2003; Valk-Lingbeek et al, 2004). Although all cells of an organism have the same genomic information, various transcriptional programs, resulting from the up- and down-regulation of specific genes, participate in defining the nature and function of each cell. The transcriptional program can be selected by regulating chromatin structure in response to internal or external stimuli.

1.1 Nucleosomes as the molecular building blocks of chromatin

The basic building block of chromatin is the nucleosome, consisting of an octamer of two of each of the core histones H2A, H2B, H3, and H4 (Baltimore, 2001), wrapped by 146 base pairs (bp) of DNA. Histones are small basic proteins consisting of a globular domain and a flexible, charged N-terminus, called the histone tail that protrudes out from the nucleosome (Luger et al, 1997; Marino-Ramirez et al, 2011). These tails interact with the DNA outside the nucleosome particles or with the linker DNA between the histones (Zheng & Hayes, 2003).

1.2 The role of histone posttranslational modifications in gene expression

The N-terminal tails of histones are subject to a vast number of posttranslational modifications (Kouzarides, 2007). An ever-increasing number of modifications at more than 128 different residues have been identified so far (Dawson & Kouzarides, 2012; Huang et al, 2014; Khare et al, 2012). Acetylation of lysines, methylation of lysines and arginines, and phosphorylation of serines and theronines and in particular modifications of the tail of histone H3 are amongst the best-studied histone modifications. A map of well-known modifications of the histone H3 tail and the enzymes setting and removing these modifications is depicted in Figure 2.

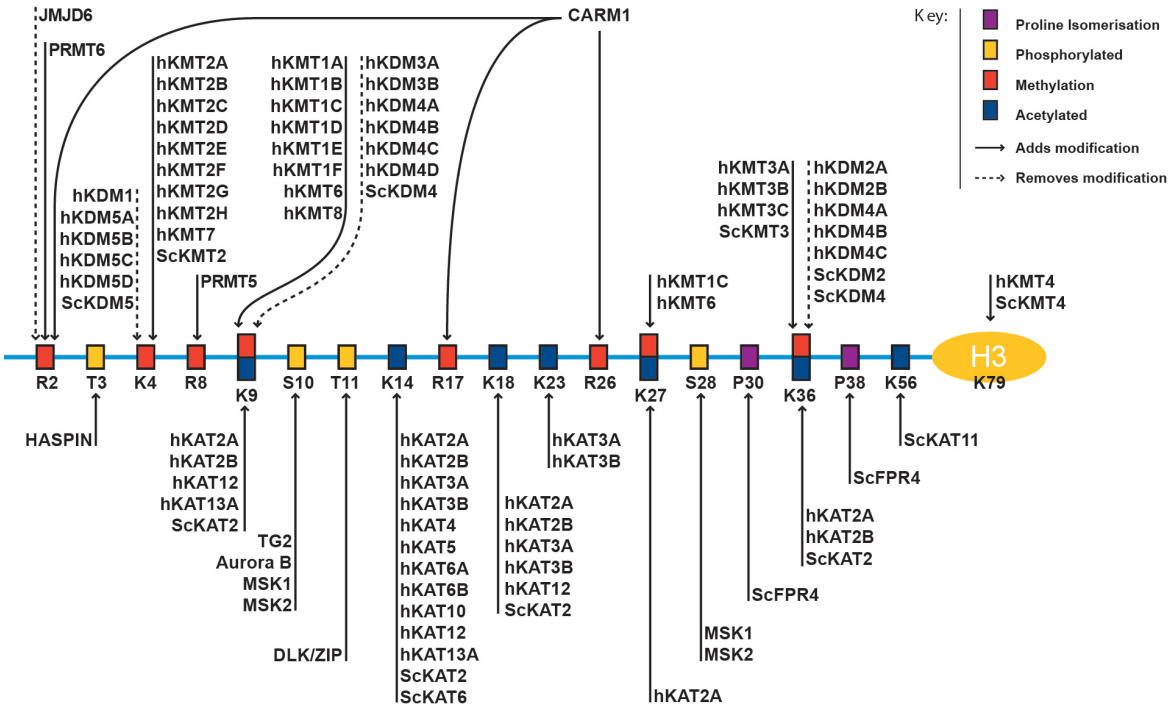


Figure 2: Histone H3 modification maps. The most common methylation, phosphorylation, acetylation, and isomerization sites are marked, and the corresponding enzymes setting or removing these modifications are indicated. Numbers denote amino acid positions. Adapted from Abcam.

Histone modifications influence chromatin structure by different mechanisms: They can affect interactions between adjacent nucleosomes and between nucleosomes and DNA (Rippe, 2012; Rippe et al, 2008). Lysine acetylation, for

example, leads to higher chromatin accessibility since it neutralizes the positive charge of lysines and weakens the electrostatic interactions with the negatively charged DNA (Görisch et al, 2005; Shogren-Knaak et al, 2006; Wang & Hayes, 2006). Furthermore, histone modifications can facilitate the binding of certain proteins with specific reader domains, while other modifications prevent their binding. For example, the heterochromatin protein 1 (HP1) is involved in the formation and maintenance of transcriptionally inactive chromatin and binds trimethylated histone H3 at lysine 9 (H3K9me3) (Grewal & Jia, 2007; Jacobs & Khorasanizadeh, 2002; Kouzarides, 2007). However, it is ejected from its binding site by phosphorylation of serine 10, which is a mark of mitotic chromatin (Fischle et al, 2005; Mateescu et al, 2004). Thus, by steering the recruitment of proteins that bind to histone tails, posttranslational modifications of histones regulate genomic functions (Jenuwein & Allis, 2001). Thereby, they define different domains of chromatin, e.g. the more open and transcriptionally active euchromatin in contrast to the more densely packed and transcriptionally inactive heterochromatin. Typical euchromatic marks include trimethylation of histone H3 at lysines 4 and 36 (H3K4me3 and H3K36me3), and acetylation of histone H3 lysines 9 and 27 (H3K9ac and H3K27ac) (Dawson & Kouzarides, 2012). Trimethylation of lysines 9 and 27 of histone H3 (H3K9me3 and H3K27me3) are classically associated with heterochromatin although more diverse functions have also been described (Gonzalez et al, 2015; Mikkelsen et al, 2007; Vakoc et al, 2005).

1.3 H3K27me3 and the polycomb repressive complex 2

H3K27me3 is one of the most extensively studied histone modifications. It was originally identified as a histone mark that is associated with gene silencing (Cao et al, 2002). However, additional functions became apparent more recently, such as its involvement in regulating alternative splicing (see introduction, section 2.2) (Gonzalez et al, 2015). Monomethylation of H3K27 has been suggested to be set by the Arabidopsis trithorax-related proteins 5 and 6 (ATXR5 and ATXR6) in

Arabidopsis thaliana (Jacob et al, 2009). Furthermore, the histone methyltransferase G9a that controls methylation of H3K9 (Tachibana et al, 2001) was shown to also contribute to monomethylation of H3K27 *in vitro* and *in vivo* in embryonic stem cells (Wu et al, 2011). However, trimethylation of H3K27 is distinct in that it has only one known methyltransferase, the enhancer of zeste 2 (EZH2) (Kuzmichev et al, 2004). EZH2 is part of the PRC2 complex, which is known for its functions in development and cell differentiation (Boyer et al, 2006; Bracken et al, 2006). The PRC2 complex is build up by four core components: EZH1/2, suppressor of zeste 12 homolog (SUZ12), embryonic ectoderm development (EED), and retinoblastoma binding protein 4 (RBAP46/48) (Figure 3). EED binds to H3K27me3 and enhances the enzymatic activity of EZH2 (Margueron et al, 2009). SUZ12 is crucial for the stability of the complex and has regulatory functions (Cao & Zhang, 2004). Both EZH2 and SUZ12 contain RNA-binding domains (Cifuentes-Rojas et al, 2014). The histone chaperones RBAP46 and RBAP48 interact with histones H3 and H4 (Song et al, 2008).

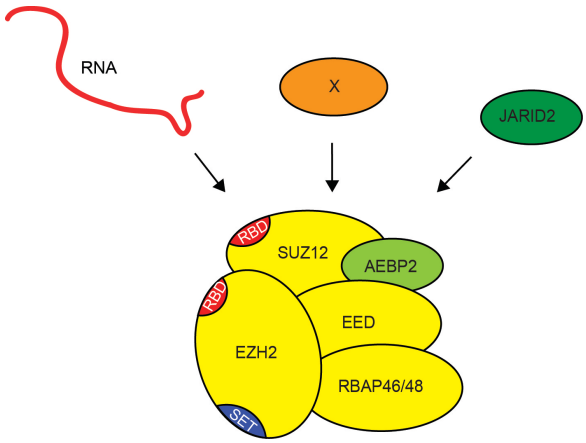


Figure 3: Scheme of polycomb repressive complex 2 (PRC2). The core components, EZH2, EED, SUZ12, and RBAP46/48, with putative RNA-binding domains (RBD) and the catalytic histone methyltransferase domain (SET) are illustrated. Accessory factors such as JARID2 or RNA are found together with the complex in a context-dependent manner. Adapted from Brockdorff, 2013.

The core units interact with various other components depending on the cellular state and environmental cues. Examples of PRC2 accessory subunits are the adipocyte enhancer-binding protein 2 (AEBP2) and Jumonji AT rich interactive domain 2 (JARID2). AEBP2 is a zinc finger protein that is known for its PRC2 targeting activity (Kim et al, 2009). JARID2 is a member of the Jumonji C family of lysine demethylases, but lacks the catalytic activity (Nayak et al, 2011).

Different studies demonstrated the importance of JARID2 for recruiting PRC2 to specific chromatin sites, and its essential role in differentiation and self-renewal of embryonic stem cells (Pasini et al, 2010; Peng et al, 2009; Shen et al, 2009). The probably best studied function of H3K27me3 and the PRC2 complex is the inactivation of one of the X chromosomes in female mammalian cells (Rougeulle et al, 2004). This process will be introduced in more detail in section 2.1 of the introduction.

Besides its function as the protein complex that sets H3K27me3, the PRC2 complex also modulates chromatin independently of its methyltransferase activity. Several studies have reported on the chromatin compacting activity of the PCR2 complex (Naughton et al, 2010; Rego et al, 2008). In agreement with these findings, experiments by Margueron and colleagues revealed that the PCR2 complex with EZH2 as the catalytically active component is mainly responsible for setting of the H3K27me3 mark. However, when in complex with EZH1, PCR2 it is mainly responsible for repressing transcription by chromatin compaction (Margueron et al, 2008). Thus, PCR2 is a multifunctional protein complex whose activity is highly dependent on its composition and accessory binding factors.

2 RNA as a regulatory factor of gene expression and RNA processing

Over the past decades, RNA transcripts not primarily dedicated to protein synthesis have emerged as key factors in diverse nuclear activities and as regulators of nuclear function and organization (Batista & Chang, 2013; Bergmann & Spector, 2014; Caudron-Herger & Rippe, 2012; Mercer & Mattick, 2013). As such, they play important roles in the organization of sub-nuclear structures (Mao et al, 2011; Shevtsov & Dundr, 2011), in altering chromatin states (Caudron-Herger et al, 2011; Caudron-Herger et al, 2015; Deng et al, 2009; Wong et al, 2007; Yang et al, 2011) and in regulating gene expression (Bergmann & Spector, 2014; Mercer & Mattick, 2013; Yang et al, 2013; Yao et al, 2010). They mainly exert this function by interacting with effector proteins and

modulating their activity. The involvement of XIST and PRC2 in X chromosome inactivation will be introduced as an example of RNA-regulated chromatin organization. Furthermore, the connection between chromatin and alternatives splicing will be established. In this process the local chromatin signature, partly resulting from its histone modification composition and influence by RNA, directly affects expression of alternative transcripts (Haque & Oberdoerffer, 2014).

2.1 RNA-mediated chromatin modification

XIST and the PRC2 complex in X chromosome inactivation

The role of the PRC2 complex has most extensively been described in the process of X chromosome inactivation (Xi) in female mammalian cells. Besides the PRC2 complex and its accessory protein factors, this process involves a number of *cis*-acting RNAs: the 17 kilobases long XIST (Brockdorff et al, 1992), its antisense transcript TSIX (Jeon & Lee, 2011) and the RepA transcript (Zhao et al, 2008). The latter originates from the first exon of XIST and contains the so called A repeats, which form a very specific stem loop structure and are essential for XIST-mediated chromosome silencing (Brockdorff, 2013; Margueron & Reinberg, 2011). The coordinated expression of these transcripts regulates the recruitment of chromatin modifying factors that eventually render one of the X chromosomes in female mammalian cells inactive. The current model involves the following steps (Figure 4): Before X inactivation, the antisense transcript of XIST, TSIX, is transcribed from the X chromosome and titrates away RepA and PRC2, blocking RepA-PRC2 interactions. Upon the onset of X chromosome inactivation, TSIX is down regulated. PRC2 then contacts RepA via the EZH2 and SUZ12 subunits. This recruitment takes place co-transcriptionally and leads to PRC2 loading onto chromatin. However, as long as the RNA is bound to the PRC2 complex its enzymatic activity is blocked. When in complex with JARID2, the RNA-PCR2 interaction is loosened and the methyltransferase activity is unblocked. EZH2 then sets the H3K27me3 leading to a cascade of events including further histone modifications such as ubiquitylation of histone H2AK119,

which eventually results in X chromosome inactivation (Cifuentes-Rojas et al, 2014). The involvement of JARID2 in this process has been reported by several studies independently (da Rocha et al, 2014; Kaneko et al, 2014a). Another publication demonstrates the necessity of the presence of the heterogeneous nuclear ribonucleoprotein K (HNRNPK) for the PCR2 complex to be able to set H3K27me3 (Chu et al, 2015).

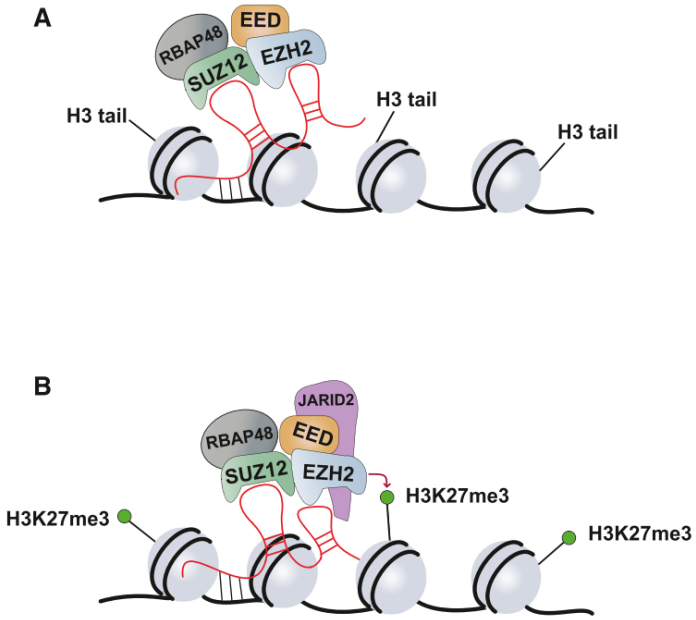


Figure 4: Model for the interaction of PRC2 with RNA and JARID2. (A) RNA binding recruits the complex to a chromatin target site, but the methyltransferase activity is blocked. **(B)** JARID2 interacts with the complex leading to a weakening of the RNA binding and unblocks the enzymatic activity. This results in H3K27me3 deposition. Histones are shown in grey, the DNA in black, and the RNA in red. (Cifuentes-Rojas et al, 2014; Kretz & Meister, 2014)

The PRC2 complex and its regulation by RNA

As illustrated in Figure 4, PRC2 recruitment and regulation is controlled by RNA. In agreement with this, PRC2 has been shown to bind RNA of human origin as well as irrelevant transcripts from ciliates and bacteria *in vitro* in a size dependent manner, with lower affinity for shorter transcripts (Davidovich et al, 2013). A quantitative binding study showed that while EZH2 alone binds RNA promiscuously, the whole PRC2 complex binds the RepA transcript in a highly specific manner compared to other transcripts (Cifuentes-Rojas et al, 2014). In a more detailed analysis conducted in the same study, the PRC2 components SUZ12, JARID2 and EZH2 have specifically been identified as RNA-binding

components. EED is suggested to positively influence the affinity of EZH2 for RNA, while JARID2 is proposed to weaken PRC2 binding to RNA. Moreover the RNA-binding capacity of EZH2 is also regulated by its cell-cycle dependent phosphorylation state (Kaneko et al, 2010).

Many studies suggest that also *in vivo* the binding of PRC2 to RNA is rather promiscuous. While candidate-based approaches have identified for example XIST (Brockdorff et al, 1992), RepA (Zhao et al, 2008), HOX (homeobox protein) transcript antisense RNA (HOTAIR) (Rinn et al, 2007; Tsai et al, 2010), KCNQ1 (potassium channel, voltage gated KQT-like subfamily Q, member 1) overlapping transcript 1 (Kcnq1ot1) (Pandey et al, 2008) or Braveheart (Klattenhoff et al, 2013) as RNAs that interact with the PRC2 complex *in vivo*, several unbiased approaches advocate the promiscuous role of RNAs in PRC2 binding and regulation. Immunoprecipitations of PRC2 have identified interacting RNAs such as antisense, intergenic and promoter-associated RNAs (Zhao et al, 2010), lincRNAs (Khalil et al, 2009), the 5' ends of nascent transcripts (Kaneko et al, 2013; Kanhere et al, 2010), full nascent transcripts (Kaneko et al, 2014b), intronic RNAs (Guil et al, 2012) and all mRNAs in general (Karapetyan et al, 2013). Finally, a recent study discusses whether indeed the PRC2 interactions are specific, promiscuous or non-specific (Davidovich et al, 2015). They clearly distinguish promiscuity from unspecificity and find that PRC2 exhibits both specificity and promiscuity in RNA-binding *in vitro*. It is concluded that these two features are not mutually exclusive. Models to explain the biological relevance of the promiscuous RNA-binding capacity of PRC2 have been suggested by several groups: Davidovich and colleagues suggested a "junk-mail" model in which PCR2 binding to RNA allows the complex to scan for target genes that have escaped repression and thereby leading to the maintenance of the repressed state (Davidovich et al, 2013). Kaneko and colleagues proposed a mechanism, by which PRC2 interacts with RNA to sense the transcriptional state of a given locus to then translate it into epigenetic information (Kaneko et al, 2014b; Kaneko et al, 2013). Having all controversies and different models in mind, the binding

specificity of RNA to PRC2 remains to be demonstrated in more detail and the function of the promiscuous RNA-binding capacity awaits further analyses.

2.2 Alternative splicing and its regulation by the local chromatin surrounding

RNA splicing and alternative splicing

Typical human genes consist of exons with an average lengths of 145 nucleotides (nt) in length that are divided by introns reaching up to ten times this length (International Human Genome Sequencing et al, 2001). In order to obtain a functional mRNA transcript, the introns have to be removed from the primary transcripts in a co-transcriptional process called RNA splicing (Black, 2003). To allow for proper identification of exons and introns, each exon is defined by specific sequences that mark the splice site at the boundaries between the intron and exon (5' splice site, 3' splice site, and branch site, Figure 5A). The components of the splicing machinery recognize these splice sites, which leads to the assembly of a multiprotein complex, named the spliceosome. Besides its function in recognizing the intron/exon boundaries, the spliceosome's major role is to catalyze the "cut-and-paste" reaction, cutting out the introns and re-joining the exons. Depending on the context, the spliceosome can consist of more than one hundred proteins with regulatory functions (Faustino & Cooper, 2003). The core components consist of the five small nuclear ribonulceoproteins (snRNPs) U1, U2, U4, U5, and U6. Each of these snRNPs consists of a single uridine-rich small nuclear RNA (snRNA) and several proteins. The U1 snRNP binds the 5' splice site, whereas the U2 snRNP contacts the branch site via RNA-RNA interactions between the snRNA and the pre-mRNA (Figure 5B). Through rearrangements of RNA-RNA, RNA-protein, and protein-protein interactions within the spliceosome (Du & Rosbash, 2002; Lallena et al, 2002; Liu et al, 2002; Liu, 2002) the pre-mRNA is ultimately spliced by the catalytic activity of the U4/U6/U5 complex (Faustino & Cooper, 2003; Villa et al, 2002).

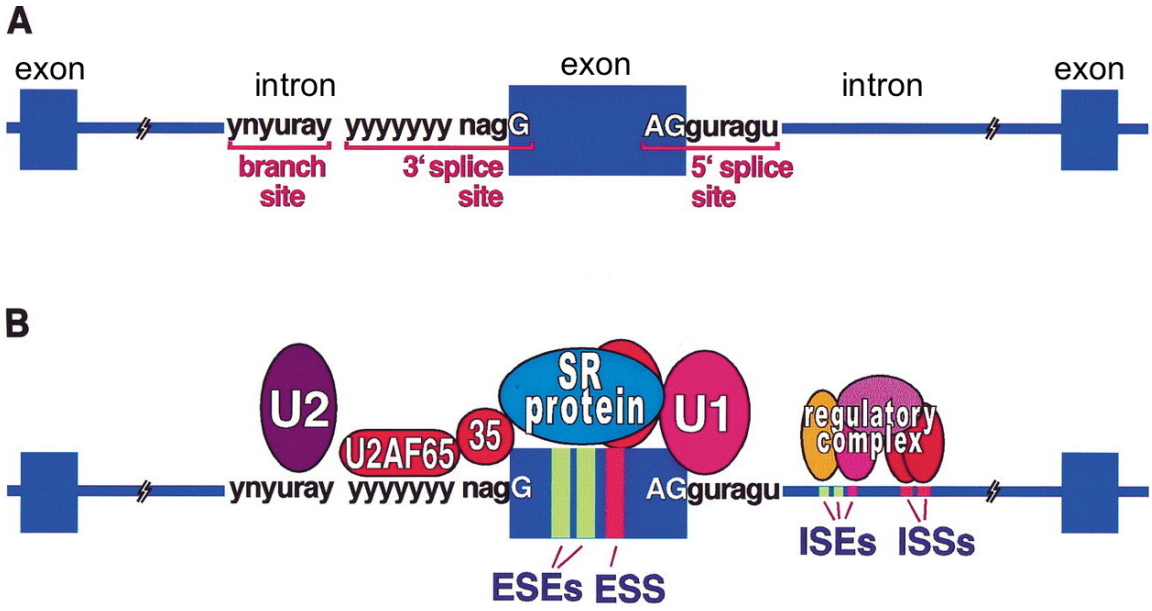


Figure 5: Schematic representation of classic and auxiliary splicing signals (n = G, A, U, or C; y = pyrimidine; r = purine). (A) The classical splicing signals, which in most human introns are required for recognition of exons (Tarn & Steitz, 1997). **(B)** Factors that bind both classical and auxiliary splicing elements. Auxiliary elements within exons (ESEs) and introns (ISEs and ISSs) are often required for the efficient splicing of both constitutive and alternative exons. The intronic elements have a function in modulating the cell-specific use of alternative exons by binding multicomponent regulatory complexes. (Faustino & Cooper, 2003)

Via a process called alternative splicing mRNA isoforms with different stability and coding potential can be produced. Particular exons of a gene may be included within or excluded from the final mRNA. Thereby, several transcript isoforms can be produced from the same locus significantly increasing the diversity of transcripts in the cell. For example, the Drosophila melanogaster gene Dscam has over 38,000 possible alternative splice forms (Black, 2000; Celotto & Graveley, 2001; Graveley, 2001; Schmucker et al, 2000). Furthermore, alternative splicing can also affect protein production by regulation of the stability of the resulting mRNA (Hughes, 2006). Finally, related proteins with qualitatively different functions can be synthesized. Alternative splicing of the nucleosome-remodeling factor subunit SNF2L, for example, removes the nuclear localization signal (NLS) from the protein and thereby alters its subcellular distribution (Lazzaro et al, 2008). Interestingly, this gene also contains an alternative exon, which, when expressed, renders the enzyme catalytically inactive (Barak et al, 2004). Thus, alternative splicing has far reaching implications for protein diversity, stability, localization and function.

To regulate alternative splicing, there is a large number of splicing regulatory proteins that recognize distinct RNA sequences across the transcriptome in addition to the basic components of the spliceosome (Fu & Ares, 2014). The cell type-specific expression levels of these RNA-binding proteins (RBPs) control each alternative splicing event. Well-studied examples are various heterogeneous nuclear ribonucleoproteins (hnRNPs), such as HNRNPC (Konig et al, 2010), the polypyrimidine tract binding protein 1 (PTBP1) (Licatalosi et al, 2012; Xue et al, 2009), the U2 small nuclear RNA auxiliary factors U2AF1 (Shao et al, 2014) and U2AF2 (Wahl et al, 2009) (also referred to as U2AF35 and U2AF65), the T-cell-restricted intracellular antigen-1 (TIA1) (Wang et al, 2010), and T-cell-restricted intracellular antigen-1 like (TIAL/TIAR) (Le Guiner et al, 2001). The ensemble of these proteins has been suggested to act as the so called “splicing code” (Fu, 2004) although attempts to “read” this code and predict alternative splicing events have so far been limited.

Regulation of alternative splicing by the local histone modification signature

Since chromatin is the template for transcription, changes in its structure and in the reading and writing of local histone modifications can modulate splicing decisions (Haque & Oberdoerffer, 2014; Kornblihtt et al, 2013; Luco & Misteli, 2011; Luco et al, 2010; Schwartz & Ast, 2010). Both nucleosome positioning and histone modifications were reported to influence splicing decisions. Different models have been described to explain the influence of chromatin on alternative splicing: It has been suggested that the elongation rate of RNA polymerase II (RNA Pol II) affects recruitment of the splicing machinery, with fast elongation favoring inclusion of a downstream exon. A change in chromatin conformation such as localized modifications of specific histones could slow down RNA Pol II. This would favor recruitment of specific splicing factors to weaker exons and induce exon inclusion (Kornblihtt et al, 2013). Moreover, histone modifications could directly recruit splicing factors via a chromatin-adaptor system that consists of a chromatin-binding protein to read the histone mark and a splicing factor to

modulate splicing of the pre-mRNA (Luco & Misteli, 2011). In agreement with this, DNA methylation has been shown to affect RNA splicing by a mechanism involving HP1 (Yearim et al, 2015).

Histone modifications such as H3K36me3, H3K79me1, H2BK5me1 (Schwartz et al., 2009) and H3K4me1 (Spies et al, 2009) but also mono-, di-, and trimethylation of H3K27 (Andersson et al, 2009) have been suggested to mark exons. In particular, the balance between H3K27me3 and H3K36me3 at specific loci has been suggested to play a role in the establishment and maintenance of specific splicing patterns (Gonzalez et al, 2015; Luco & Misteli, 2011). A mechanism has been proposed, by which the expression of an anti-sense transcript of the fibroblast growth factor receptor 2 (*FGFR2*) induces recruitment of PRC2 *in cis* to lead to enrichment of H3K27me3 and the H3K36me3 demethylase KDM2a. The combination of these two events in turn inhibits recruitment of the chromatin-adaptor complex of mortality factor 4 like 1 (MRG15) and PTBP1 (MRG15-PTBP1) thereby favoring specific exon inclusion. Given that the PCR2 complex binds to RNA promiscuously and is vastly regulated by RNA (see introduction, section 2.1), its recruitment to loci where RNA is being produced and its involvement in RNA splicing is of particular interest for this study.

3 Cellular functions of 3'-UTRs

In eukaryotes mature mRNAs have a tripartite structure consisting of a 5'-UTR, a coding region, and a 3'-UTR (Mignone et al, 2002). The UTRs are known to play crucial roles in the post-transcriptional regulation of gene expression. Since this thesis focuses on 3'-UTRs, a more detailed introduction on their cytoplasmatic and nuclear functions is given in the following.

The 3'-UTR begins at the stop codon and its end is defined by the poly(A) tail (Zhao et al, 2011). The *cis*-regulatory elements within 3'-UTRs can influence polyadenylation, translation efficiency (van der Velden & Thomas, 1999), localization (Jansen, 2001) and stability of the transcript (Bashirullah et al, 2001). For some genes, alternative splicing results in the use of an alternative final exon

that produces variation in both the 3'-UTR length and sequence, leading to variation in the expression of the gene (Hughes, 2006).

3.1 Cytoplasmic functions of 3'-UTRs

Classically, 3'-UTRs are known for their roles that take place in the cytoplasm (Matoulkova et al, 2012). These include the following: First, 3'-UTRs contain binding sites for microRNAs (miRNAs). MiRNAs are short trans-acting RNAs that regulate the expression of the majority of protein-coding genes post-transcriptionally. Their binding to the 3'-UTR promotes destabilization and degradation of mRNA or inhibits translation (Farazi et al, 2011; Friedman et al, 2009). Second, many 3'-UTR contain AU-rich elements (AREs) which influence mRNA decay, stability, or translation (Xu et al, 1997). These functions are mediated by ARE-binding proteins that recognize the AREs in a cell type-, timing-, and cellular localization-dependent manner. Third, the poly(A) tail attached to 3'-UTRs is bound by poly(A) tail-binding proteins which also contribute to regulation of mRNA translation, stability and export (Mandel et al, 2008). For example, proteins that bind the poly(A) tail interact with factors involved in translation initiation, causing circularization of the transcript and promoting translation. The absence of the poly(A) tail in many cases leads to exonuclease-mediated degradation of the mRNA (Pichon et al, 2012). Finally, secondary structure motives found in 3'-UTRs aid binding of specific RBPs. Attracting these proteins can be of importance for example to relocalize the transcript to specific subcellular compartments and again to regulate translation (Barrett et al, 2012). In summary, the 3'-UTR of protein-coding transcripts is highly important for posttranscriptional regulation of gene expression. Varying length and sequence of the 3'-UTR is therefore one of many mechanisms to alter gene expression.

3.2 Nuclear functions of 3'-UTRs

Besides their functions in regulating mRNA stability, localization and translation, 3'-UTRs have been suggested to have alternative roles. Mercer and colleagues

found that a large number of 3'-UTRs in human, mouse and fly are also expressed independently from their associated protein-coding sequences, to which they are normally linked. It is proposed that they are produced by post-transcriptional cleavage from the rest of the mRNA transcript and that they have nuclear localizations (Furuno et al, 2006; Mercer et al, 2011). In another study, Caudron-Herger and colleagues suggested that a nuclear RNA fraction enriched in long 3'-UTR sequences maintains higher order chromatin in an open configuration (Caudron-Herger et al, 2011). Both findings point to a nuclear function of 3'-UTR sequences in chromatin organization. Interestingly, a systematic analysis of known long non-coding RNAs (lncRNAs) revealed similarities between 3'-UTRs and lncRNAs in terms of structural features and sequence composition (Niazi & Valadkhan, 2012). So far a vast number of lncRNAs have been linked to chromatin organization whereas very little is known about the nuclear activities of 3'-UTRs. Their direct link to chromatin organization and function and whether and how this is linked to the translation of the associated protein-coding sequence remains to be investigated.

4 Methods to detect and functionally characterize nuclear RNAs in single cells

A multitude of biochemical and deep sequencing-based methods are nowadays available to detect and functionally characterize nuclear RNAs. While these methods offer great advantages in terms of high-throughput analyses, they suffer from their limitations in studying individual cells. Microscopy-based readouts, on the other hand, have the advantage that samples can be studied at the single-cell level and individual observations do not get lost due to population averaging. The first method described here is microinjection used as a tool to monitor the subcellular localization of RNAs and to link this to their function. The second technique termed the "MS2/LacI-mediated recruitment system" is used to recruit a transcript of interest to a specific chromatin locus and to examine its effect on the local chromatin environment. In both approaches it is an important issue to be

able to detect the transcripts because endogenous or ectopically expressed RNAs, unlike proteins, so far cannot be directly tagged with fluorescent proteins to observe their behavior in living cells.

4.1 Microinjection of RNA into the nuclei of living cells

To examine the effect of overrepresentation of a specific RNA in cells and to introduce a visible label to an RNA of interest one can either transfect a plasmid encoding for the RNA, transfect the RNA itself or use microinjection. The method introduced here is microinjection of RNAs directly into the nuclei of living cells (Figure 6A).

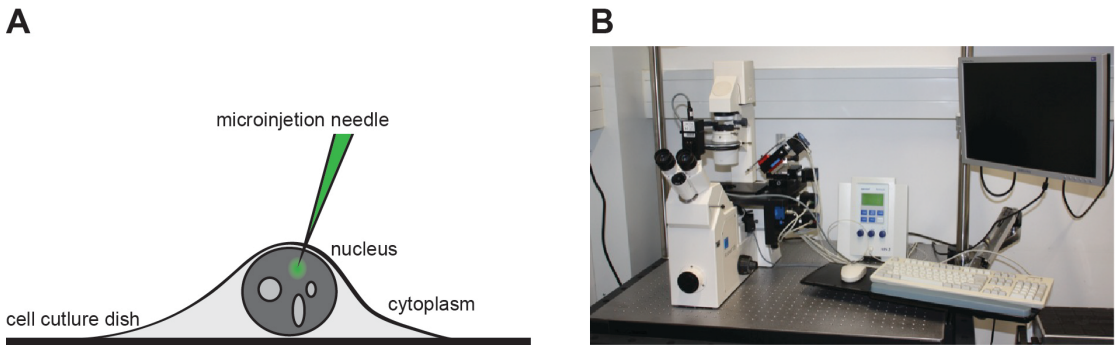


Figure 6: The principle of computer assisted microinjection. (A) Schematic illustration of the basic principle of microinjection: A needle with a tip diameter is loaded with the desired RNA. Using the computer assisted microinjection system, the RNA is injected directly into the nucleus. **(B)** Photograph of the computer assisted microinjection system that was used to microinject RNAs directly into the nuclei of living cells.

This has a number of advantages over parallel methods that are available to deliver high amounts of RNA transcripts into living cells: First, the RNAs can be used at any desired concentration, a parameter that can be very closely controlled by microinjections. Second, morphology and physiology can be directly correlated on the minutes-time scale after microinjection. Third, a whole pool of RNAs of known or even unknown identity can be examined. Lastly, RNAs can also be microinjected at high concentrations to cells that are difficult to transfect. The main disadvantage is, however, that microinjection is an extremely laborious

technique and only a restricted number of cells can be treated. Computer assisted microinjection systems such as the one used in this study (Figure 6B) are a vast improvement to this point. Although the number of microinjected cells is still too small to be able to perform classical biochemical studies, it is large enough to obtain statistically significant results from single cell imaging data.

4.2 Recruitment of RNA to a specific genomic locus

To investigate the influence of proteins on chromatin a chromatin-tethering systems has been developed and employed in several studies. It involves a stable integration of the bacterial *lac*O sequence into mammalian genomes (Belmont & Straight, 1998; Robinett et al, 1996) and makes use of the fact that the interaction of the LacI protein with the *lac*O sequence is highly specific and affine (Figure 7A). Transfecting these cells with the LacI protein fused to a fluorescent protein has facilitated studying the dynamics of chromatin in living cells (Chuang & Belmont, 2007; Jegou et al, 2009; Straight et al, 1996). Furthermore, the system has been exploited to investigate protein-protein interactions in the nuclei of living mammalian cells and to recruit proteins of interest to chromatin (Chung et al, 2011). One main advantage of the system is that while the *lac*O sequence is naturally part of the E.coli lac operon, it offers a unique, well-defined binding site once it is artificially incorporated into the genome of mammalian cells.

Instead of recruiting proteins, another well-described nucleic acid-protein interaction system has been introduced to recruit RNA transcripts to the chromatin. It employs the MS2 bacteriophage protein, which binds to an RNA hairpin, the MS2 stem loops, with high affinity and specificity (Johansson et al, 1998). In this recruitment system, the MS2 coat protein is fused to both a fluorescent protein and the LacI (Figure 7B). Furthermore the RNA of interest is tagged with several MS2 stem loop sequences. This allows for enrichment of the RNA at the chromatin and studying its influence on the local chromatin surrounding. A previous study made use of a similar system to immobilize RNAs

at the chromatin. It was observed that RNA can function as a structural element and a nucleator of nuclear bodies (Shevtsov & Dundr, 2011), suggesting that the MS2/LacI-mediated recruitment system can indeed be used to study the mechanism of action of nuclear RNAs.

This model system can have far reaching applications, two of which are of particular interest to this study: investigation of the effect of RNA on the state of the chromatin in terms of its compaction and posttranslational modifications of histones and examination of the RNA's protein interaction partners. In all cases the visualization of RNA at the *lac*O array can be accomplished by several methods as described in section 4.3 of the introduction.

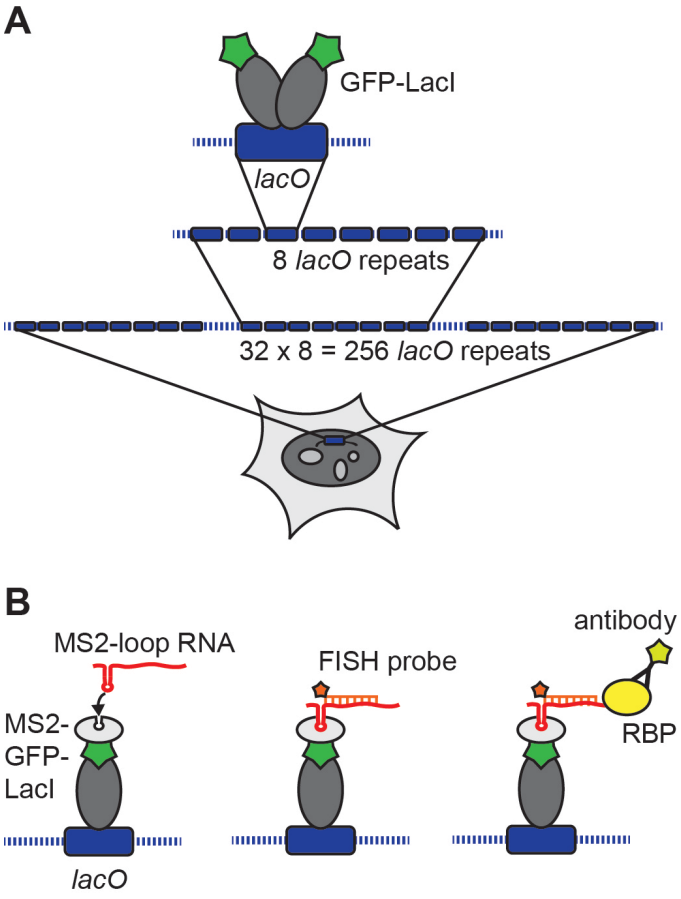


Figure 7: Scheme of the MS2/LacI-mediated recruitment system. (A) Arrays of the *lac*O repeats are stably integrated into the genome of mammalian cells. The LacI fused to GFP or a similar fluorescent protein binds with high affinity and specificity to these arrays. Transfection of LacI-GFP can therefore be used to visualize these genomic loci. **(B)** Experimental approach to tether MS2 stem loop-tagged RNAs via MS2-GPF-LacI to stably integrated *lac*O arrays. For simplicity only one of the MS2 stem loops sequences is depicted here. In the experiments described in this thesis plasmids encoding for 18 MS2 loops were used. The RNA can for example be visualized by RNA-FISH as depicted schematically. Its putative protein interaction partners can be assessed by immunofluorescence. GFP: green fluorescent protein. Adapted from Jegou et al., 2009 and Shevtsov and Dundr, 2011.

4.3 Labeling and visualization of RNAs

Labeling and visualization of microinjected RNA

In order to observe RNAs after microinjection or recruitment to the *lac*O arrays they need to be labeled. Unlike proteins, RNAs cannot be directly labeled in living cells. Nonetheless, for visualization of RNA after microinjection, a number of methods have been developed in the recent years that have significantly expanded the possibilities to visualize RNAs in living cells. For example, RNAs can be tagged with specific sequence motives, which are recognized by RBPs fused to GFP. For example, the same RNA-protein interactions as described above involving the MS2 coat protein (see introduction, section 4.2) could also be exploited for RNA labeling (Bertrand et al, 1998; Daigle & Ellenberg, 2007; Ozawa et al, 2007; Valencia-Burton & Broude, 2007; Valencia-Burton et al, 2007). Alternatives with similar characteristics as the MS2 coat protein would include another bacteriophage protein, the lambda N protein (Daigle & Ellenberg, 2007; De Gregorio et al, 1999), the Tat (trans-activator) peptide from the bovine immunodeficiency virus (Wakiyama et al, 2012), or the MS2 coat protein-related PP7 coat protein, which binds a stem loop in the PP7 RNA (Lim et al, 2001). All of these might, however, change the natural behavior of the RNA due to the relatively large size of the GFP fusion protein in comparison to the RNA itself.

A number of small-molecule-based RNA imaging methods have been developed to circumvent this: These methods have in common that they use an aptamer sequence introduced into the transcript. For example, RNAs can be tagged with sequences, which bind molecules that only fluoresce when in complex with the RNA (Arora et al, 2015; Paige et al, 2011). Similarly to using aptamers that are specifically bound by RNA-binding proteins, this method can be a disadvantage for several reasons: First, the aptamer sequence, even though it is usually rather short, presents a foreign sequence and thus potentially introduces changes to the RNA in terms of structure, binding affinities to proteins, mobility and function. Second, the RNAs of interest have to be cloned together with the aptamer sequence, which can be laborious when investigating large numbers of transcripts. Third, only RNAs of known sequence can be investigated, so that

these techniques make it impossible to study for example pools of unknown RNAs. To avoid cloning and transfecting the RNA of interest with an aptamer sequence, stem-loop shaped oligonulceotides, called molecular beacon probes, can be used. They have an internally quenched fluorophore whose fluorescence is restored upon hybridization of the probe with the RNA. This methods and its derivatives, however, suffer from challenges in delivery and stability of the probes (Tyagi, 2009).

To circumvent also these limitations alternative RNA labeling methods make use of direct labeling of the RNA of interest. This can be done for example by producing the RNA of interest *in vitro* in the presence of a fluorescently labeled nucleotide. Fluorescein, for example, is a small, auto-fluorescent molecule that should not hinder the RNA to adopt is physiological secondary structure and is readily incorporated into the RNA by T7 polymerase *in vitro* (Proudnikov & Mirzabekov, 1996). When labeling purified RNA pools of unknown sequences, propidium iodide (PI) can be used. It preferentially stains RNA but also DNA in case of low RNA concentrations (Suzuki et al, 1997). On the one hand, it is thus a well-suited agent to stain and observe the microinjected RNA. On the other hand, it is also a suitable localization marker to identify the microinjected cells. The main advantage is that it can be used to simultaneously label an entire pool of RNAs.

Visualization of RNA recruited to chromatin

To label RNAs that have been recruited to the *lac*O array, one possibility involves the incorporation of ethynyl-uridine (EU) into RNA. This compound can be added to the cell culture medium and is readily incorporated into all freshly produced RNA by RNA Pol II. After fixation of the cells, the transcripts with incorporated EU are made visible by attaching a fluorophore to the EU using an azide-alkyne cycloaddition reaction, also called “click” reaction (Liang & Astruc, 2011). Thus, detection of RNA at the arrays is based on the assumption that the enrichment at the arrays is strong enough to be distinguishable it from the nuclear background where nascent RNAs are being produced constantly.

The second, more specific, possibility to label *lac*O-immobilized RNA is to perform RNA-FISH directed against either the MS2 stem loops (Goodier et al, 2010) or directly against the RNA of interest (Figure 7B). The LacI/MS2-mediated recruitment system theoretically leads to a clear enrichment of the RNA-FISH signal at the *lac*O arrays because of the high affinity of the MS2 stem loops to the MS2 coat protein. It is expected that most of the MS2 stem loop-tagged RNA is found at the array and only very little is distributed over the nuclear background. Both methods are tested and compared in the course of the present thesis.

5 Scope of the thesis

Alternative splicing of pre-mRNA is a crucial step in the expression of many genes and provides an important control layer affecting transcript and proteome diversity in a specific cell type (Chen & Manley, 2009). Numerous studies showed that cancer cells take advantage of alternative splicing to produce aberrant proteins harboring altered, added, or deleted functional domains, which drives tumorigenesis (David & Manley, 2010; Oltean & Bates, 2014; Zhang & Manley, 2013). A detailed understanding of the principles that govern alternative splicing and the mechanisms that cells use to stably maintain their splicing pattern is thus essential to discover novel drug targets for therapeutic interventions.

While it has been described in much detail how the outcome of alternative splicing is controlled by the combination of multiple *cis*-acting RNA elements within the pre-mRNA (Barash et al, 2010), the recruitment of regulatory splicing factors can also be influenced by chromatin modifications (Alló et al, 2009; Haque & Oberdoerffer, 2014; Kornblihtt et al, 2013; Luco & Misteli, 2011; Luco et al, 2010; Schwartz & Ast, 2010). These chromatin modifications are well known to be influenced by nuclear lncRNAs that play roles in guiding and regulating the activity of chromatin modifying enzymes (Rinn & Chang, 2012). As a novel player for RNA-dependent chromatin modifications specific 3'-UTRs have been suggested to exert nuclear functions besides their well-described role in regulating mRNA stability and protein translation in the cytoplasm (Caudron-

Herger et al, 2011; Mercer et al, 2011). Moreover, 3'-UTRs resemble lncRNAs in terms of structural features and sequence composition (Niazi & Valadkhan, 2012). Given that 3'-UTRs are part of the pre-mRNA, as well as the spliced transcript, it seems conceivable that they could be involved in modifying the chromatin and thereby regulating alternative splicing. However, to date, it remains unknown how 3'-UTRs can affect chromatin modifications and whether this activity is linked to alternative splicing decisions.

To address these open questions, I investigated whether 3'-UTRs are linked to chromatin modifications and nuclear RNA processing as an additional circuit by which RNA feeds back on its own production. Towards this goal, this thesis focused on the analysis of a particular 3'-UTR transcript, CU-RNA, originating from the 3000 nucleotide (nt) long 3'-UTR of the *CDV3* gene. A number of methodological advances had to be performed in order to be able to detect and characterize nuclear RNAs. Using these techniques, CU-RNA was found to compact chromatin and could be linked to the recruitment and regulation of EZH2 and thus the setting of H3K27me3, the latter being specific for one of the two transcript isoforms of *CDV3*. In a second set of experiments, the H3K27me3-promoting activity of CU-RNA could be linked to its own alternative splicing. A model is proposed, in which CU-RNA regulates its own alternative splicing via a feedback loop that establishes a splicing specific chromatin signature. According to this model, stable steady-state levels between CU-RNA splicing variants are established that might be distinct with respect to their 3'-UTR sequence-dependent cytoplasmic activities.

Thus, this thesis aims at unraveling a novel mechanism, which extends the previously known scope of functions of 3'-UTRs in the cell. A mechanism is suggested, by which a 3'-UTR modulates chromatin modifications and regulates alternative splicing. Furthermore, it is concluded from this work that a chromatin dependent splicing regulatory system could control the abundance of certain transcript isoforms by an epigenetic-based mechanism. Through the involvement of EZH2 in this process, a direct link to deregulation of H3K27me3 and alternative splicing in cancer could be made, since both H3K27me3 and EZH2 have been

shown to be deregulated in many cancer types (Chase & Cross, 2011; Plass et al, 2013).

Materials and Methods

1 Materials

1.1 Biological materials

Bacterial strains

The *E. coli* chemocompetent strains DH5α (Thermo Fisher Scientific, USA) and SURE 2 (Stratagene, USA) were used to amplify and prepare plasmid DNA.

Cell lines

The HeLa, HeLa S3 and U2OS osteosarcoma cell lines were used (DSMZ, Braunschweig, Germany). For establishment of LacI-mediated RNA recruitment studies the U2OS cells line with a single telomeric *lac*O integration was used (Jegou et al, 2009). The CHO hamster cell line AO3, also with a stable integration of a *lac*O array, was employed for recruitment studies (Luijsterburg et al, 2012; Tumbar et al, 1999).

Primers

Synthetic DNA oligonucleotides were used as primers for creating DNA fragments for molecular cloning, quantitative real-time PCR analysis, and for producing templates for RNA *in vitro* transcription. All primers were purchased from Eurofins MWG Operon (Germany) and are listed in Table 1.

Table 1: Primers used in this study. Restriction sites and T7 promoter sequences are highlighted. Ctrl: control, *iv* txn: *in vitro* transcription, ssRT-PCR: strand specific RT-PCR.

Name	Sequence (5' → 3')	Application
NotI-CDV3a-fwd	AAAGCGGCCGCTGCTACCTGGAATTAAATG	cloning
NotI-CDV3a-rev	AAAGCGGCCGCACAGTCCAAAATCCATATATTTAC	cloning
NotI-CDV3-H1-fwd	AAAGCGGCCGCGGGTGGCTTTTAGAATAAATCCC	cloning
NotI-CDV3-H1-ref	AAAGCGGCCGCACAAATCACAGCCCTACACAG	cloning

Name	Sequence (5' → 3')	Application
NotI-CDV3-H2-fwd	AAAGCGGCCGCGGAAGGGGAAGGGAGAGAAT	cloning
NotI-CDV3-H2-rev	AAAGCGGCCGCGCCAAACCATGCCTGAGAAT	cloning
NotI-CORO1C-fwd	AAAGCGGCCGCGCAAGAACTTGTGCTTG	cloning
NotI-CORO1C-rev	AAAGCGGCCGCCTCTGAAAATGGAAGCAAG	cloning
NotI-STARD7-fwd	AAAGCGGCCGCCAGGCTTTGGGATAAGAAG	cloning
NotI-STARD7-rev	AAAGCGGCCGCTACCAATCTTTATGTATTTATTC	cloning
NotI-SSR3-fwd	AAAGCGGCCGCACCATGTCAGCTTCACC	cloning
NotI-SSR3-rev	AAAGCGGCCGCCAGATGAATCTTTTATTTC	cloning
NotI-Ctrl-fwd	AAAGCGGCCGCTGCGGGACTACCTCTCAAAC	cloning
NotI-Ctrl-rev	AAAGCGGCCGCGCAAACGGGAACACGATATT	cloning
NotI-RepA-fwd	AAAGCGGCCGCCACTCTCTTTTCTATATTTT	cloning
NotI-RepA-rev	AAAGCGGCCGCTTCATCCATAAAAAGCACCGA	cloning
T7-CORO1C-fwd	TAATACGACTCACTATAGGGCAAGAACTTGTGCTTG	*iv* txn
T7-CORO1C-rev	GACAGAACACTATGGCCAC	*iv* txn
T7-CDV3a-fwd	TAATACGACTCACTATAGGGGCTACCTGGAATTAAATGAC	*iv* txn
T7-CDV3a-rev	CTTGCACTGCCAAACCATGCC	*iv* txn
T7-STARD7-fwd	TAATACGACTCACTATAGGGCAGGCTTTGGGATAAGAAGGG	*iv* txn
T7-STARD7-fwd	GTAACTGTACAGACAGACCCCAAG	*iv* txn
T7-AS-STARD7-fwd	CAGGCTTTGGGATAAGAAGGG	*iv* txn
T7-AS-STARD7-fwd	TAATACGACTCACTATAGGGTAACTGTACAGACAGACCCCAAG	*iv* txn
T7-SSR3-fwd	TAATACGACTCACTATAGGGCTGGCTTTGTGTCTATGG	*iv* txn
T7-SSR3-rev	GAGTAGGTCGCATCTCTGTG	*iv* txn
T7-SURF4-fwd	TAATACGACTCACTATAGGGCAGTCACAGATCCCTACCTG	*iv* txn
T7-SURF4-rev	GCC AGA ATC AAT TGG TAG ATC C	*iv* txn
T7-EIF4B-fwd	TAATACGACTCACTATAGGGCCTCTACATCCTGTGCTT	*iv* txn
T7-EIF4B-rev	CATCAAGACATCTCCCAGG	*iv* txn
T7-FADS-fwd	TAATACGACTCACTATAGGGCAACAGCCACCCTGCCCAGTC	*iv* txn
T7-FADS-rev	TGC CTC TGG CCA GGT TGT AAA	*iv* txn
T7-DDXD21-fwd	TAATACGACTCACTATAGGGAGCACATTGTGCCTCCTTTTG	*iv* txn
T7-DDXD21-rev	ACCCATGGCTTGATGGGAAGC	*iv* txn
T7-EIF2S3-fwd	TAATACGACTCACTATAGGGCGGATGGATTTGGAAGTTGGA	*iv* txn
T7-EIF2S3-ref	CCTCATGCTACAACACAACC	*iv* txn
T7-PTTG1IP-fwd	TAATACGACTCACTATAGGGAGCGCTCCAGCACATCAGT	*iv* txn
T7-PTTG1IP-rev	CGAAGACCGACCTTACAGAC	*iv* txn
T7-FAM38A-fwd	TAATACGACTCACTATAGGGCTGTCTACAGCAACCTG	*iv* txn
T7-FAM38A-ref	GTTCCTGCCTGCAGGTTCTCC	*iv* txn
T7-Ctrl-RNA-1-fwd	TAATACGACTCACTATAGGGGATCTCAGCTCACTGGAACC	*iv* txn
T7- Ctrl -RNA-1-ref	CCAGATCCTTATACTCAACAG	*iv* txn

Name	Sequence (5' → 3')	Application
T7- Ctrl -RNA-2-fwd	TAATACGACTCACTATAGGGGAGTACGGGCACCATCCATA	*iv* txn
T7- Ctrl -RNA-2-ref	TGTGGCACTGCAGAAAGTG	*iv* txn
T7- Ctrl -RNA-3-fwd	TAATACGACTCACTATAGGGTCCGGTTTGGCCGATTCTG	*iv* txn
T7- Ctrl -RNA-3-ref	CAAGAGTGAAACTCCGTCC	*iv* txn
T7-RepA-fwd	TAATACGACTCACTATAGGGCACTCTCTTTTCTATATTTT	*iv* txn
T7-RepA-rev	TTCATCCATAAAAAGCACCGA	*iv* txn
CDV3a/CDV3b-fwd	CCAAGCATGTAGAAAGCCGG	qRT-PCR
CDV3a-rev	AGTGGTGAAAAGGCTGGGAT	qRT-PCR
CDV3b-rev	TGTGAGTGGCTGCTTTTCTG	qRT-PCR
ACTB-fwd	TCCCTGGAGAAGAGCTACGA (Arnoult et al, 2012)	qRT-PCR
ACTB-rev	AGCACTGTGTTGGCGTACAG (Arnoult et al, 2012)	qRT-PCR
CU-RNA -fwd	TAATACGACTCACTATAGGGTGTTAATGTTACAGAATTCTTGCACTG	ssRT-PCR
CU-RNA-rev	TGCTACCTGGAATTAAATGACTACA	ssRT-PCR

Expression vectors

The following expression plasmid vectors were used for transfection of cells (Table 2).

Table 2: Expression vectors used in this study

Plasmid	Reference
pcDNA3.1-MS2loops	this study
pcDNA3.1-MS2loops-CORO1C-3'-UTR	this study
pcDNA3.1-MS2loops-CDV3a-3'-UTR/CU-RNA	this study
pcDNA3.1-MS2loops-STARD7-3'-UTR	this study
pcDNA3.1-MS2loops-SSR3-3'-UTR	this study
pcDNA3.1-MS2loops-CU-RNA-H1	this study
pcDNA3.1-MS2loops-CU-RNA-H2	this study
pcDNA3.1-MS2loops-CU-RNA-Q1	this study
pcDNA3.1-MS2loops-CU-RNA-Q2	this study
pcDNA3.1-MS2loops-CU-RNA-T0	this study
pcDNA3.1-MS2loops-CU-RNA-T1	this study
pcDNA3.1-MS2loops-CU-RNA-T2	this study
pcDNA3.1-MS2loops-CU-RNA-ΔT0	this study
pcDNA3.1-MS2loops-RepA	this study
pSV2-MS2-GFP-LacI	(Caudron-Herger et al, 2015)

G-blocks

The following G-blocks were used as PCR-templates for molecular cloning, for *in vitro* transcription and as templates in quantitative real-time PCR to validate qRT-PCR primers. All were purchased from Integrated DNA Technologies (IDT, USA) and are listed in Table 3**.**

Table 3: G-blocks used as PCR-templates for molecular cloning.

Name	Sequence (5' → 3')	Application
CDV3-Q1	CACACCTACAGCTACGTTGAGCGGCCGCGGGTGGCTTTTAGAATAA ATCCCAGCCTTTTCACCACTTTGAGAGCACAGCTGGCTTGAATTTAA ACCTCCACGTGGACTTGCTCCTCTGCTCTGTCTTAAAAAATGTGTAA GGGACCTTGGGGTGGCTTTTAGAATAAATCCCGGAGGGCAGTGATT CACCACACTCGAGTTTCTTTACCTGATCTAATCAGATCCACTCCCACA AAATCGATGTGAGGAGAAATCATTGACTAATGAGTAGAATTTACTTGA ATGAGAAAATTGTTACCCTTGTGCATAAAGAGAGAGATGTATTATCTA TTACTTGCTAAAAGTAAGAGTCTTAGGAGGAATGTCATTACACAGCTT TTAACAGTTTTCTTCAAGTTTGTCCTGAAATAGGGTCACAGTTAATAC TGTTCAGTAGAAGATAGGACCCTAGAGCTTCAATACAGCGTTCTGTG TTTTCTGTTTGAATTGAAGCCTTAAAATGGTTTGATAAGGATTTTCTAT ATGCCTCCACTCCTACCCCTCCCCAGGATTCTTCTAAGGGGTGGCTT TTTTTTAAATTCAAGGACGAATATTTTCAAAAACCCTAGTGAAGAAGA AATATTTACTGATTACATTTCTTTTCCCTTAGGGATAAAGAAATGGAG AAGAGCTTTGAAGTAGTAAGACACGCGGCCGCGTGGCTGCATCTGC TCGAAG	Cloning
CDV3-Q2	CACACCTACAGCTACGTTGAGCGGCCGCGATGAGGTTTCAAAAAAC CAGGCCCTTAAACTTCAGCTAGACAACCAATATGCTGTGCTTGAAAA TCAGAAAAGCAGCCACTCACAATACAATTAAGGAATGGGCTTTGCTA ACCCTTCTGAGGTAACTAGACTGCAGCTAACCACCACCAACAGCCAT TCATCATCTGATCTCTGCTGGATCTACAGACACCGATGCAGACCACT CGATTTCATGACCGGCCCTATTGCACTATGGAAGTTAAAGTGTCACG ACTGCTCTATGCATATTGGATTTAGGGGAATTTTCATTGTTACATAAA TGTGTGAACTAGTTTCAACAGTGTTCTTTCATATTTACTCTGCAAATA CAAAAAACCAAAACCTGCAGCCAGTGGTCATTTCAAAATCTTTTTATG TTCAGATACTGAGCCTTCATAAGGGTTGACTACCTCAGATTTGCTGC ACTCATTGTGGACTTCATGTGGATCACAACTTCTGGATAAGAAGATTA CAACTATTAAGTGTCGATGTGAACCTTGCAACCAGCTCTACTGGATT CTTATCAGAAATCCTGCATAAAAAGTCAGCCATCTGGGTTCTGATCT GCTGTAAAAGATGAAGATTTAAGTGACCTTAATTAACCTGTCCTGTGC CCTACCCTTAAGGAATACTCTCTGTAGTAGGCTGTTGTTATATTAGAC TTCCTGGAACACACTGCTGAAAAGAACTGATGTGTTCAGATCATCTG TGTAGGGCTGTGATTTGTAATTTAAACTAATTGTATTCTGAGGTAGCG GCCGCGTGGCTGCATCTGCTCGAAG	Cloning
CDV3-T0 (CU-RNA-T0)	CACACCTACAGCTACGTTGAGCGGCCGCAAATTCAAGGACGAATATT TTCAAAAACCCTAGTGAAGAAGAAATATTTACTGATTACATTTCTTTTC CCTTAGGGATAAAGAAATGGAGAAGAGCTTTGAAGTAGTAAGACACA AAAATAGAGGTAGGGATGAGGTTTCAAAAAACCAGGCCCTTAAACTT CAGCTAGACAACCAATATGCTGTGCTTGAAAATCAGAAAAGCAGCCA CTCACAATACAATTAAGGAATGGGCTTTGCTAACCCTTCTGAGGTAA CTAGCGGCCGCGTGGCTGCATCTGCTCGAAG	Cloning
CDV3-T1	CACACCTACAGCTACGTTGAGCGGCCGCGATGAGGTTTCAAAAAAC CAGGCCCTTAAACTTCAGCTAGACAACCAATATGCTGTGCTTGAAAA TCAGAAAAGCAGCCACTCACAATACAATTAAGGAATGGGCTTTGCTA ACCCTTCTGAGGTAACTAGACTGCAGCTAACCACCACCAACAGCCAT TCATCATCTGATCTCTGCTGGATCTACAGACACCGATGCAGACCACT CGATTTCATGACCGGCCCTATTGCACTATGGAAGTTAAAGTGTCACG ACTGCGCGGCCGCGTGGCTGCATCTGCTCGAAG	Cloning
CDV3-T2	CACACCTACAGCTACGTTGAGCGGCCGCTCTATGCATATTGGATTTA GGGGAATTTTCATTGTTACATAAATGTGTGAACTAGTTTCAACAGTGT TCTTTCATATTTACTCTGCAAATACAAAAAACCAAAACCTGCAGCCAG TGGTCATTTCAAAATCTTTTTATGTTCAGATACTGAGCCTTCATAAGG GTTGACTACCTCAGATTTGCTGCACTCATTGTGGACTTCATGTGGAT CACAACTTCTGGATAAGAAGATTACAACTATTAAGTGTCGATGTGAAC GCGGCCGCGTGGCTGCATCTGCTCGAAG	Cloning

Name	Sequence (5' → 3')	Application
CDV3-T3	CACACCTACAGCTACGTTGAGCGGCCGCCTTGCAACCAGCTCTACT GGATTCTTATCAGAAATCCTGCATAAAAAGTCAGCCATCTGGGTTCT GATCTGCTGTAAAAGATGAAGATTTAAGTGACCTTAATTAACCTGTCC TGTGCCCTACCCTTAAGGAATACTCTCTGTAGTAGGCTGTTGTTATAT TAGACTTCCTGGAACACACTGCTGAAAAGAACTGATGTGTTCAGATC ATCTGTGTAGGGCTGTGATTTGTAATTTAAACTAATTGTATTCTGAGG TAGCGGCCGCGTGGCTGCATCTGCTCGAAG	Cloning
CDV3-ΔT0 (CU-RNA-ΔT0)	CACACCTACAGCTACGTTGAGCGGCCGCGACCACCAGAAATCTACA GTGATACACAGTTCCCATCCCTGCAGTCAACTGCCAAGCATGTAGAA AGCCGGAAGGATAAAGAAATGGAGAAGAGCTTTGAAGTAGTAAGAC ACAAAAATAGAGGTAGGGATGAGGTTTCAAAAAACCAGGCCCTTAAA CTTCAGCTAGACAACCAATATGCTGTGCTTGAAAATCAGAAAAGCAG CCACTCACAATACAATTAAGGAATGGGCTTTGCTAACCCTTCTGAGG TAACTAGCGGCCGCGTGGCTGCATCTGCTCGAAG	Cloning
CDV3-S1	CACACCTACAGCTACGTTGAGCGGCCGCAAATTCAAGGACGAATATT TTCAAAAACCCTAGTGAAGAAGAAATATTTACTGATTACATTTCTTTTC CCTTAGGGATAAAGAAATGCGGCCGCGTGGCTGCATCTGCTCGAAG	Cloning
CDV3-S2	CACACCTACAGCTACGTTGAGCGGCCGCGGAGAAGAGCTTTGAAGT AGTAAGACACAAAAATAGAGGTAGGGATGAGGTTTCAAAAAACCAGG CCCTTAAACTTCAGCTAGACAGCGGCCGCGTGGCTGCATCTGCTCG AAG	Cloning
CDV3-S3	CACACCTACAGCTACGTTGAGCGGCCGCACCAATATGCTGTGCTTG AAAATCAGAAAAGCAGCCACTCACAATACAATTAAGGAATGGGCTTT GCTAACCCTTCTGAGGTAACTAGCGGCCGCGTGGCTGCATCTGCTC GAAG	Cloning
CDV3a	AACTGGGAAGAAGGTGGAGGTGGTGGTGGAGGTATGGAAAAATCTT CAGGTCCCTGGAATAAAACAGCTCCAGTACAAGCACCTCCTGCTCCA GTAATTGTTACAGAAACCCCAGAACCAGCGATGACTAGTGGTGTGTA TAGGCCTCCTGGGGCCAGGTTAACCACAACAAGGAAAACACCACAA GGACCACCAGAAATCTACAGTGATACACAGTTCCCATCCCTGCAGTC AACTGCCAAGCATGTAGAAAGCCGGAAGTACTTAAAATGAATGCTAC CTGGAATTAAATGACTACATATGGCACATCTACATCTTACTTTAGGGA CATGAATTTTTTAAAATAAATAAAATGGGAAGAAATGTAAATGACATC AGATGGAAATGTGTTTTTTGGGGTGGCTTTTAGAATAAATCCCAGCC TTTTCACCACTTTGAGAGCACAGCTGG	qRT-PCR validation
CDV3b	GGTGATAACTGGGAAGAAGGTGGAGGTGGTGGTGGAGGTATGGAAA AATCTTCAGGTCCCTGGAATAAAACAGCTCCAGTACAAGCACCTCCT GCTCCAGTAATTGTTACAGAAACCCCAGAACCAGCGATGACTAGTGG TGTGTATAGGCCTCCTGGGGCCAGGTTAACCACAACAAGGAAAACA CCACAAGGACCACCAGAAATCTACAGTGATACACAGTTCCCATCCCT GCAGTCAACTGCCAAGCATGTAGAAAGCCGGAAGGATAAAGAAATG GAGAAGAGCTTTGAAGTAGTAAGACACAAAAATAGAGGTAGGGATGA GGTTTCAAAAAACCAGGCCCTTAAACTTCAGCTAGACAACCAATATG CTGTGCTTGAAAATCAGAAAAGCAGCCACTCACAATACAATTAAGGA ATGGGCTTTGCTAACCCTTCTGAGGTAACTAGACTGCAGCTAACCAC CACCAACAGCCATTCATCATCTGA	qRT-PCR validation

siRNAs

Silencer select siRNAs were purchased from Ambion (Thermo Fisher Scientific, USA) and used for silencing of the respective genes (Table 4).

Table 4: siRNAs used in this study.

siRNA	siRNA ID #
HNRNPK	s6739
MRG15	s21501
Negative control	Negative control #1 (catalog #: 4390844)

Antibodies

Antibodies were used for immunofluorescent stainings, RNA-FISH, western and northern blotting. All primary antibodies (Table 5) were specific to the protein of interest. The secondary antibodies (Table 6) were directed against the primary antibodies and coupled to a fluorescent dye or horseradish peroxidase.

Table 5: Primary antibodies used in this study. IF: immunofluorescence, FACS: fluorescence activated cell sorting, WB: western blot, NB: northern blot.

Antibody	Host	Source	IF/FACS	WB/NB
DIG	Mouse	Roche, clone 1.71.256	1:100	1:10000
EZH2	Rabbit	Active Motif, AM39933	-	1:100
EZH2	Mouse	Active Motif, AM39875	1:50	-
H3	Rabbit	Abcam, ab1791	-	1:5000
H3K27me3	Rabbit	Active Motif, AM39535	1:600	1:600
HNRNPC	Mouse	Santa Cruz, sc-32308	-	1:200
HNRNPK	Mouse	Santa Cruz, sc-32307	-	1:200
JARID2	Rabbit	Abcam, ab192252	1:100	1:1000
JARID2	Rabbit	NovusBio, nb100-2214	1:1000	1:1000
MRG15	Rabbit	Active Motif, AM39362	-	1:500
PTBP1	Rabbit	Cell Signaling, 8776	-	1:1000
U2AF2	Mouse	Sigma-Aldrich, U4758	-	1:5000

Table 6: Secondary antibodies used in this study. IF: immunofluorescence, FACS: fluorescence activated cell sorting, WB: western blot, NB: northern blot.

Antibody	Host	Source	IF/FACS	WB/NB
Alexa 568-anti-mouse IgG	Goat	Thermo Fisher Scientific, USA	1:500	-
Alexa 568-anti-rabbit IgG	Goat	Thermo Fisher Scientific, USA	1:500	-
Alexa 633-anti-mouse IgG	Goat	Thermo Fisher Scientific, USA	1:500	-
Alexa 633-anti-rabbit IgG	Goat	Thermo Fisher Scientific, USA	1:500	-
HRP-linked-anti-mouse IgG	Goat	Cell Signaling, USA	-	1:2000
HRP-linked-anti-rabbit IgG	Goat	Cell Signaling, USA	-	1:2000

Enzymes

The following enzymes were used for cloning, PCR, *in vitro* transcription, qRT-PCR, protein, RNA and DNA extracts (Table 7).

Table 7: Enzymes used in this study.

Enzyme	Source
FastAP	Thermo Fisher Scientific, USA
Proteinase K	Genaxxon, Germany
Polymerase (qRT)	Roche, Germany
Q5 polymerase	New England Biolabs, Germany
Restriction enzymes (FastDigest)	Thermo Fisher Scientific, USA
RNase A	Thermo Fisher Scientific, USA
RNase H	New Enlgand, Biolabs, Germany
RQ1 DNase I	Promega, Germany
T4 DNA ligase	Thermo Fisher Scientific, USA

1.2 Media and solutions

Bacterial culture media

Bacterial culture media used in this study are depicted in Table 8.

Table 8: Bacterial culture media used in this study.

Medium	Composition
Luria Bertani (LB) medium	1 % (w/v) tryptone 0.5 % (w/v) yeast extract 86 mM NaCl pH = 7.0
LB medium with antibiotics	LB + 100 µg/ml ampicillin LB + 50 µg/ml kanamycin
LB agar plates	1000 ml LB 15 g agar antibiotics
SOC medium	New England Biolabs, Germany

Cell culture media and reagents

Cell culture media and reagents used in the study are listed in Table 9. The use of the different media is explained in section 2.2 of the materials and methods.

Table 9: Cell culture media and reagents used in this study.

Medium/Reagents	Source
DMEM (without phenol red)	Thermo Fisher Scientific, USA

Medium/Reagents	Source
DMEM high glucose (without phenol red)	Thermo Fisher Scientific, USA
DMEM/F-12 (without phenol red)	Thermo Fisher Scientific, USA
DMSO	Sigma-Aldrich, Germany
Fetal Bovine Serum (FBS)	PAN Biotech, Germany
L-glutamine	PAA, Germany
Leibovitz's L-15 medium (without phenol red)	Thermo Fisher Scientific, USA
Opti-MEM I reduced serum medium	Thermo Fisher Scientific, USA
Penicillin/Streptomycin	PAN Biotech, Germany
RPMI (without phenol red)	GE Healthcare, Germany
Trypsin	PAN Biotech, Germany

Transfections reagents

All cells were transfected with Lipofectamine 2000 (Thermo Fisher Scientific, USA).

Nucleic acid staining dyes and mounting reagents

4',6-Diamidin-2-phenylindol (DAPI) and propidium iodide (PI) were used in this study to stain DNA and RNA, respectively. Mowiol was used to mount microscopy slides (Table 10).

Table 10: DNA staining dyes and mounting reagents

Dye / mounting reagent	Source / composition
DAPI	Sigma-Aldrich, Germany
PI	Thermo Fisher Scientific, USA
Mowiol	2.4 g of Mowiol 4-88 6 g of glycerol 6 mL of H_2O 12 mL of 0.2 M Tris-Cl (pH 8.5)

Small molecule inhibitors

To inhibit histone H3 lysine 27 methlyation the small molecule inhibitor GSK343 (BioVision, USA) was dissolved in DMSO at a stock concentration of 10 mM and used at a final concentration of 5 µM or 1 µM. Pre-mRNA splicing was inhibited

with 100 nM pladienolide B (PlaB, BioAustralis, USA) dissolved in ethanol with a stock concentration of 1 mM.

Buffers

The composition of all non-commercial buffers used in this study is listed in Table 11.

Table 11: Buffers used in this study.

Buffer	Composition
Buffer A (nucleoli preparation)	10 nM Hepes, pH 7.9 10 mM KCl 1.5 mM $MgCl_2$ 0.5 mM DTT
Cell lysis buffer (nuclear extract preparation)	85 mM KCl 0.5 % NP-40 5 mM HEPES pH 7.4
Covaris sonication buffer	10mM Tris pH8.0 200mM NaCl 1mM EDTA 0.1% Na-Deoxycholate 0.5% n-Lauroylsarcosine + Protease Inhibitor Complete
CSK buffer	100 mM NaCl 300 mM sucrose 3 mM $MgCl_2$ 10 mM PIPES pH 6.8 0.5 % Triton X 100
Elution buffer (ChIP)	50mM Tris pH 8.0 1mM EDTA 1% SDS 50mM $NaHCO_3$
High salt buffer (ChIP)	50mM Hepes pH 7.9 500mM NaCl 1mM EDTA 1% Triton X-100 0.1% Na-deoxycholate 0.1% SDS
Laemmli buffer	0.1% (v/v) 2-Mercaptoethanol 0.0005% (w/v) Bromophenol blue 10% (v/v) Glycerol 2% (w/v) SDS 63 mM Tris-HCl pH 6.8

Buffer	Composition
Li buffer (ChIP)	20mM Tris pH 8.0 1mM EDTA 250mM LiCl 0.5% NP-40 0.5% Na-deoxycholate
methylation buffer (4X)	50 mM Tris-HCl pH 8.5 5 mM $MgCl_2$ 4 mM DTT
NaP buffer	12.2 mM Na_2HPO_4-$7H_2O$ 7.8 mM NaH_2PO_4
Nuclei lysis buffer (nuclear extract preparation)	300 mM NaCl 20 mM Tris-HCl pH 8.0 0.2 % Tween-20 1mM EDTA 1 mM EGTA
PBS (pH 7.2) 1x	1.7 mM KH_2PO_4 137 mM NaCl 10 mM Na_2HPO_4 2.7 mM KCl
RNA-FISH hybridization buffer	30% dextran sulfate 2 mg/mL BSA 4x SSC
RIPA buffer (1X)	150 mM NaCl 1 % NP40 0.5 % Na-Desoxy-cholate, 0.1 % SDS 50 mM Tris pH 8.0 1X Complete Protease Inhibitor Cocktail (Roche, Switzerland)
SSC buffer, 1x	150 mM NaCl 15 nM Na_3Citrate
SDS running buffer (pH 8.3) 1x	25 mM Tris 192 mM glycine 0.1 % (w/v) SDS
TAE	40 mM Tris 13.8 mM $C_2H_4O_2$ 1 mM EDTA, pH 8.0
TE buffer (pH 8.0)	10 mM Tris 1 mM EDTA
TBS (pH 7.5) 1x	50 mM Tris-CL 150 mM NaCl
Transfer buffer 1x	25 mM Tris 192 mM glycine 0.5 % (w/v) SDS 20 % MeOH (v/v) (freshly added before use)
washing buffer (RNA affinity purification)	5 mM Tris-HCl pH 7.5 0.5 mM EDTA 0.5 M NaCl 0.05 % Tween-20

1.3 Commercial kits

The following commercial kits were used in this study (Table 12).

Table 12: Commercial kits used in this study.

Kit	Source
EZH2 Assay Kit	BPS Bioscience, USA
Click-it RNA imaging Kit	Thermo Fisher Scientific, USA
Light Cycler 480 SybrGreen I Master	Roche, Germany
NEBNext Ultra DNA Library Prep Kit	Illumina, USA
NucleoSpin plasmid purification kit	Macherey Nagel, Germany
Pierce BCA Protein Assay Kit	Thermo Fisher Scientific, USA
RNeasy Mini kit	Qiagen, Germany
SuperScript III Reverse Transcriptase	Thermo Fisher Scientific, USA
T7 High Yield RNA Synthesis Kit	New England Biolabs, Germany
*T*elo*TAGGG* Telomere Length Assay	Roche, Germany
Wizard SV Gel and PCR Clean-Up System	Promega, Germany

1.4 Instruments

All non-standard instruments used in this study are listed in Table 13.

Table 13: Instruments used in this study.

Instrument	Source
Chemi Doc Gel documentation system	BioRad, Germany
Covaris sonicator	Covaris, USA
FACSCanto II	BD Bioscience, USA
FemtoJet microinjection module	Eppendorf, Germany
Leica TCS SP5 Confocal Laser Scanning Microscope	Leica Microsystems, Germany
LUNA automated cell counter	Logos Biosystems, USA
Mastercycler	Eppendorf, Germany
Qubit fluorometer	Thermo Fisher Scientific, USA
StepOne Plus RT-qPCR	Applied Biosystems, Germany
Spectrophotometer NanoDrop	Thermo Fisher Scientific, USA
SemiDry Blotter	VWR, Germany
TECAN Infinite M200 plate reader	Tecan, Switzerland
2200 Tape Station Instrument	Agilent Technologies, USA

1.5 Software

All non-standard software applications used in this study are listed in Table 14.

Table 14: Software used in this study.

Software	Source
Adobe Illustrator	Adobe Systems, USA
Adobe Photoshop	Adobe Systems, USA
BEDtools	(Quinlan & Hall, 2010)
bowtie	(Langmead et al, 2009)
Excel	Microsoft, USA
Fastqc	(Andrews, 2010)
ImageJ	National Institutes of Health, USA
Integrative Genome Viewer	(Robinson et al, 2011)
KaleidaGraph	Synergy Software, USA
LAS AF	Leica Microsystems, Germany
MCORE	(Molitor et al, 2015)
R	www.r-project.org
Scaffold	Proteome Software, USA
Serial Cloner	serialbasics.free.fr
Sicer	(Zang et al, 2009)

2 Methods

2.1 Molecular cloning

All RNAs sequences expressed from the pcDNA3.1 plasmid containing the MS2 stem loop sequences (Schmidt et al, 2011) were amplified by PCR using the Q5 polymerase and primers containing the NotI restriction site. Total HeLa cDNA or g-blocks were used as templates for PCR. PCR-products were purified using the Wizard SV Gel and PCR Clean-Up System. Restriction digestion of both insert and target vector were carried out using FastDigest enzymes. Phosphate residues of the vector were removed by FastAP phosphatase for 30 min at 37°C. Both inserts and vector fragments were purified by agarose gel extraction and ligated using the T4 DNA ligase. To amplify the ligated products, they were transformed into *E.coli* SURE2 chemocompetent cells, which were grown in

LB-ampicillin medium at 30°C over night. Plasmids were then isolated using the NuceolSpin plasmid purification kit. All kits and enzymes were used according to the manufacturer's instructions. All cloned expression plasmids were checked by DNA sequencing using the over-night-express GATC services (Germany). Plasmid amplification of pSV2-MS2-GFP-LacI were carried out in DH5α chemocompetent bacteria.

2.2 Cell culture techniques

Cultivation of mammalian cell lines

HeLa cells were cultured in RPMI medium supplemented with 2 mM L-Glutamine, 10% FCS and 100 µg/ml penicillin/streptomycin. HeLa S3 cells were grown in DMEM medium supplemented with 2 mM L-Glutamine, 10 % FCS and 1% penicillin/streptomycin (v/v). AO3 cells were grown in DMEM/F-12 medium supplemented with 2 mM fresh L-Glutamine, 20% FCS and 1% penicillin/streptomycin (v/v). EC4 and F4BII8 cells were cultured in DMEM medium supplemented with 2 mM L-Glutamine, 10 % FCS and 1% penicillin/streptomycin (v/v). For microinjection experiments cells were kept in Leibovitz's L15-medium supplemented with 2 mM L-Glutamine, 10 % FBS and 1% penicillin/streptomycin (v/v). All cells were grown at 37°C in 5% CO_2.

Transient transfections

Transient transfection of plasmids into cells was conducted with Lipofectamine 2000 (Thermo Fisher Scientific, USA) according to the manufacturer's instructions. Briefly, cells were seeded 24 h prior to transfection to reach a confluence of 80-90% at the time point of transfection. The penicillin/streptomycin-containing medium was replaced by antibiotic-free medium. Plasmid DNA was diluted in Opti-MEM, mixed with lipofectamine, also diluted in Opti-MEM, stirred rigorously and incubated at room-temperature for 20 min. Lipofectamine-DNA complexes were then added to the cells. The

transfection mixes were removed from the cells 6-8 h after transfection and cells were fixed 24 h after transfection.

siRNA transfection

For siRNA-mediated knockdown of HNRNPK and MRG15 1.5 x 10^5 cells were reverse-transfected with Lipofectamin 2000 (Thermo Fisher Scientific, USA) and siRNAs against HNRNPK or MRG15 (see Table 4). Silencer select negative control #1 siRNA (Ambion, #3490844) was used as a non-target control. 72 h after transfection, cells were processed for western blot and qRT-PCR analysis.

2.3 Inhibition of EZH2, pre-mRNA splicing and RNA Pol II

To inhibit methylation of histone H3 at lysine 27, HeLa cells were treated with 5-µM or 1 µM GSK343, which was added to the cell culture medium for 3 or 6 days. Cells treated with equal volumes of DMSO were included as negative controls. To inhibit pre-mRNA splicing, cells were treated with 100 nM PlaB in the cell culture medium for 5 h. Cells treated with an equal volume of ethanol were included as a negative control. RNA Pol II was inhibited by culturing cells for 5 h in medium supplemented with 50 µg/ml α-amanitin.

2.4 RNA purification

RNA was prepared from HeLa cells using the RNeasy Mini Kit (Qiagen, Germany) according to the manufacturers instructions. For DNA digestion, 2 µL of RQ1 DNAse, and 10 µL RQ1 10x buffer were added to each sample, and H_2O was added to a total volume of 100 µL. The samples were kept at a reaction temperature of 37°C for 30 min. Subsequently the RNA was purified with RNeasy Mini Kit again according to the manufacturers instructions. The RNA was eluted in 30 µl RNase free H_2O supplemented with 1 µl RiboLock. The concentration of each sample was determined with the NanoDrop. RNA integrity was tested on a 2% E-gel (Thermo Fisher Scientific, USA).

2.5 Northern blotting

15 µg total or nuclear RNA were resolved by 1.1 % agarose, 2% formaldehyde-MOPS gel electrophoresis and blotted onto a Hybond-N+ nylon membrane over night with 20X SSC. The membrane was washed in 2X SSC, baked at 120°C for 20 min and hybridized with DIG-labeled probes at 42°C in DIG Easy Hyb solution (*T*elo*TAGGG* Telomere Length Assay, Roche, Switzerland) over night according to the manufacturer's instructions. The membrane was then washed twice in 2X SSC-0.1% SDS for 5 min at room temperature and twice with 0.5X SSC-0.1% SDS for 20 min at 65°C. The *T*elo*TAGGG* Telomere Length Assay (Roche) was used for detection according to the manufacturer's instructions: the membrane was rinsed in 1X washing buffer and incubated in blocking solution for 30 min at room temperature and another 30 min in anti-DIG-AP working solution. The membrane was then washed twice with washing buffer for 15 min at room temperature. Before detection the membrane was briefly incubated with detection buffer. Approximately 1.5 ml substrate solution were then added to the membrane and the signal was allowed to develop for at least 5 min. Blots were developed on Chemi Doc Gel documentation system.

2.6 Quantitative real-time PCR

First-strand cDNA was transcribed from 1 µg RNA using 50 ng random hexamer primers (Thermo Fisher Scientific, USA), 1 mM dNTPs and 10 U/µl SuperScript III reverse transcriptase according to the manufacturers instructions. Finally, 0.5 U RNase H were added to digest the residual RNA fragments for 20 min. Controls without reverse transcriptase (-RT) were included and analyzed in parallel to exclude contamination with genomic DNA. qRT-PCR was conducted with a Light Cycler 480 SybrGreen I Master mix and 0.5 µM of transcript specific primers in a StepOne Plus RT-PCR system. For quantification, a standard curve was measured that ranged from 0.6 ng/µl to 180 ng/µl cDNA. At least three replicates of each of three biologically independent samples were used. Normalization was done with *ACTB* as an internal control. The following amplification programs were

applied: 95°C for 10 min followed by 40 cycles at 95 °C (10 sec), 60 °C (10 sec) and 72 °C (10 sec). Acquisition of the melt curves and 2% agarose gel runs were conducted for detection of contaminations.

2.7 Ethynyl uridine labeling of RNA

To label RNAs with ethynyl uridine in, cells were incubated with 1 mM 5-ethynyl uridine (EU) over night, fixed with 4% paraformaldehyde/PBS and EU-labeled transcripts were detected using Alexa Fluor 488 azide using the Click-it RNA imaging Kit (Thermo Fisher Scientific, USA) according to the manufacturers instructions.

2.8 RNA fluorescence *in situ* hybridization

For RNA-FISH cells were grown on coverslips and permeabilized in CSK buffer containing 10 mM vanadyl ribonucleoside complex (VRC, New England Biolabs, Germany) or 50 µg/ml RNase A. Cells were subsequently fixed with 4% paraformaldehyde/PBS, dehydrated by sequential washes with ethanol (70%, 85%, and 100%) and air-dried. For MS2 stem loop RNA detection 50ng 5'-Atto-565 labeled antisense probe per slide were mixed with 10 µg salmon sperm DNA and 5 µl formamide. The mixture was heated to 37°C for 10 min and 74°C for 7 min before 5 µl hybridization buffer and 10mM VRC and 50 µg/ml RNase A were added. For detection of endogenous sequences 5 µl of digoxigenin (DIG)-labeled probe (final concentration 50 ng/µl in 50% Formamide, 2x SSC + 10% Dextrane) per slide were used. In both cases the RNA was hybridized overnight in hybridization buffer at 37°C. The next day slides were washed twice with 2X SSC, 50% formamide at room-temperature for 15 min, once with 0.2X SSC, 0.1% Tween at 40°C for 10 min and once with 2X SSC at room temperature for 5 min. When using ATTO labeled probes the slides were washed again in PBS, incubated with 1x DAPI solution and mounted on microscope slides (Menzel-Gläser, Germany) with Mowiol. When using DIG

labeled probes the slides were subjected to immunofluorescence directed against DIG as described in section 2.15 of the materials and methods.

2.9 RNA *in vitro* transcription

DNA templates for *in vitro* transcription were amplified using cDNA from HeLa cells or g-blocks, and specific primers with T7 promoter sequences. *In vitro* transcription was performed using the RNA polymerase T7 High Yield RNA Synthesis Kit according to the manufacturer's instructions. Briefly, 1X reaction buffer, 2 µl T7 Polymerase, ATP, GTP, CTP and UTP at a final concentration of 10 mM and 1 µg template DNA were mixed and incubated at 37°C for 2 h. The RNA was purified using the RNeasy Mini kit (Qiagen) and its integrity was tested on a 2% E-gel (Thermo Fisher Scientific, USA). For the production of fluorescently labeled or biotinylated transcripts, UTP-fluorescein (fluorescein-12-UTP, Thermo Fisher Scientific, USA) and UTP-biotin (biotin-16-UTP, Epicentre, Great Britain), respectively, were used and mixed with unlabeled UTP at a ratio of 1:1.

2.10 Rescue experiment

HeLa cells grown on 18 mm cover slips were washed in PBS and permeabilized for 30 sec in 0.035% Triton X-100/PBS (v/v), washed in PBS and incubated for 5 min in 3 mg/ml RNase A/PBS. Cells were then washed in PBS and incubated with cell culture medium containing 30 µl RiboLock to inhibit residual RNase A. For testing various RNAs, cells were incubated with the RNA of interest for 15 min after addition of 16.7 ng/µl RNA. Cells were then fixed with 4% PFA, the nuclei were counterstained with DAPI and the slides mounted with Mowiol.

2.11 Nucleoli preparation

Nucleoli were prepared as described by (Busch et al, 1963; Sullivan et al, 2001). HeLa cells were harvested by trypsinization, washed twice with ice-cold PBS and

resuspended in buffer A. After 5 min incubation on ice the cell suspension was dounced 10 to 15 times in a pre-cooled dounce tissue homogenizer. The homogenized cells were centrifuged at 218 g for 5 min, taken up in 3ml 0.35 M sucrose/0.5 mM $MgCl_2$ and sonicated until there were no intact cells or nuclei in the suspension as observed by checking using a phase contrast microscope. The sonicated sample was layered over 3 ml of a 0.88mM sucrose/0.5 mM $MgCl_2$ cushion and centrifuged at 3000g for 10 min at 4°C. The nucleoli were then contained in the pellet and further processed for RNA extraction.

2.12 Nuclear extract preparation

Cells were washed with PBS once, harvested by scraping in PBS and pelletized. 2-3 volumes cell lysis buffer were added and the cells were kept on ice for 10 min. To obtain the nuclear pellets the cells were centrifuged at 4 °C and 5000 rpm for 5 min. The nuclei pellet was resuspended in 3-4 volumes of nuclei lysis buffer, and incubated for 30 min at 4 °C on a rotor. The supernatant containing the nuclear protein extract, was transferred in a fresh tube after centrifugation at 13000 rpm for 10 min at 4 °C. The NaCl concentration of the nuclear lysate was adjusted to 200 mM with the nuclei lysis buffer without NaCl. The protein concentration of the extract was determined with the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, USA) according to the manufacturers instructions and using the NanoDrop.

2.13 *In Vitro* methylation assay

To investigate the activity of EZH2 *in vitro* the EZH2 Assay Kit (BPS Bioscience, USA) was used according to the manufacturer's instructions. Briefly, the microwells were rehydrated by adding TBST buffer to every well and incubating at room temperature for 15 min. The EZH2 complex was diluted in 1X HMT assay buffer at 10 ng/µl. The EZH2 complex, 400 µM S-adenosylmethionine and the RNA or protein lysate of interest were added to the wells in duplicates in a total volume of 50 µl and incubated at room temperature for 1 h. Instead of the

reaction buffer supplied in the kit, the reaction was carried out in a custom-made 4X methylation buffer (Cifuentes-Rojas et al, 2014). The wells were then washed three times with TBST buffer. Next, 100 µl of blocking buffer was added to every well and incubated on a rotating platform for 10 min. The wells were incubated with the primary antibody, α-H3K27me3, diluted 1:100 in blocking buffer, for 1 h at room temperature with slow shaking. Before addition of the secondary antibody the wells were washed again three times in TBST buffer and again incubated in blocking buffer as described above. The secondary, HRP-coupled antibody was diluted 1:1000 in blocking buffer and incubated on the wells at room temperature with slow shaking for 30 min. Finally the wells were washed again with TBST and 50 µl HRP chemiluminescent substrate A and 50 µl chemiluminescent substrate B were added to the wells. The luminescence was read out immediately with the plate reader Infinite® M200 (Tecan Austria GmbH, AT). The integration time was 1000 ms. The EZH2 activity in the samples (%) was calculated via the ratio of the sample luminescence (L) to the value of the positive control (posCtrl), after subtraction of the blank values.

$$\text{EZH2 activity [\%]} = \frac{L_{sample} - L_{blank}}{L_{posCtrl.} - L_{blank}} \times 100\%$$

2.14 RNA affinity purification

RNA affinity purification was performed as described previously (Caudron-Herger et al, 2015). Briefly, cell nuclei were isolated from HeLa S3 to prepare a nuclear extract using nuclear lysis buffer (see section 2.12, materials and methods). 5 µg of biotinylated RNA transcripts were incubated with 50 µl of pre-washed streptavidin-coated magnetic beads (Thermo Fisher Scientific, USA) at room temperature in 100 mM NaCl solution for 30 min. Beads were washed in washing buffer and incubated for 3 h in 70 µg nuclear extract in a total volume of 100 µl. After 3 steps of washing in nuclear lysis buffer, the beads were resuspended in 50 µl of this buffer and bound proteins were eluted by adding 30 µg of RNase A

at 4°C for 30 min. Beads were retrieved on a magnet and the protein solution was further proceeded for western blot analysis or mass spectrometry.

2.15 Immunofluorescence

For immunofluorescent stainings, cells were grown on glass coverslips and fixed with 4 % paraformaldehyde/PBS. They were permeabilized for 5 min in ice-cold 0.5% Triton X-100/PBS (v/v), washed three times in PBS and blocked with 10 % goat serum/PBS at room temperature for 30 min. Primary antibodies specific to the protein of interest were diluted in the blocking solution as indicated in Table 5 and incubated on the slides for at least 1 h at room temperature. To remove excessive primary antibody, cells were washed three times with 0.002%NP40/PBS (v/v). Appropriate secondary antibodies conjugated with fluorescent dyes (Table 6) were also diluted in the blocking solution and incubated on the slides at room temperature for 30 min. The slides were again washed three times in PBS, incubated with 1x DAPI solution in PBS and mounted on microscope slides (Menzel-Gläser, Germany) with Mowiol.

2.16 Western blotting

Protein lysates were prepared by lysing cells in RIPA buffer, unless indicated otherwise. Protein concentrations of lysates were measure with the Pierce BCA Protein assay kit (Thermo Fisher Scientific, USA) according to the manufacturers instructions and using the NanoDrop. Lysates were taken up in 1X Laemmli buffer and incubated at 95°C for 5 min for denaturation and loaded onto precast-stain-free gels (Mini-PROTEAN® TGX Stain-Free™ Precast Gels, 4-20%). After electrophoretic protein separation, the proteins were transferred to a nitrocellulose (Whatman, UK) or PVDC (BioRad, USA) membrane using semi-dry blotting. The membranes were blocked with 5% milk powder (Roth, Germany) in TBS at room temperature for 1 h and were incubated in primary and secondary antibodies either at 4°C overnight or at room temperature for 1 h (Table 5, Table 6). Antibody incubation was followed by three washing steps with TBS

supplemented with 0.05% Tween 20 (v/v). HRP-linked secondary antibodies were detected using a chemiluminescent ECL reagent (BioRad, Germany) and the Chemi Doc Gel documentation system (BioRad, Germany) was used for detection.

2.17 Acid extraction of histones

To extract soluble proteins approximately 10^6 cells were pelleted and washed once with PBS. The pellet was then dissolved in 0.25 M HCl (100 µl/10^6 cells) and homogenized. The suspension was rotated over night at 4°C. After centrifugation for 5 min at 208000xg and 4°C the supernatant was collected. The remaining pellet was suspended again in half the volume of 0.25 M HCl (50 µl/10^6 cells) to ensure complete extraction of the soluble proteins. After a second centrifugation step for 5 min at 208000xg and 4°C the supernatants were combined. To precipitate histones TCA was added to a final concentration of 33% (v/v) and the mixture was incubated on ice for 30 min. After centrifugation for 20 min at 4°C the pellet was washed twice with TBME/EtOH (1:1). The pellet was then air-dried, dissolved in water and processed for western blotting. (Shechter et al, 2007; Villar-Garea et al, 2008)

2.18 Mass spectrometry

Mass spectrometry was carried out at the Mass Spectrometry Core Facility of the ZMBH (Zentrum für Molekulare Biologie der Universität Heidelberg). Proteins eluted from the RNA affinity purification were shortly run into a Mini-PROTEAN® TGX™ Precast Gels (4-20%) and stained by colloidal Coomassie blue. For each sample, the entire lane was cut into two gel slices. After in-gel digestion with trypsin the samples were subjected to mass spectrometry and analyzed with the Scaffold Proteome Software. The gene ontology (GO) term analysis proteins identified by mass spectrometry was carried out using the DAVID tool (Database for Annotation, Visualization, and Integrated Discovery) (Dennis et al, 2003).

2.19 ChIP-seq

Chromatin immunoprecipitation

$1x10^6$ to $1x10^7$ Cells were collected by trypsinization, washed once with PBS and fixed with 8 ml PBS/formaldehyde (1%, (v/v)) at room temperature for 10 min. The reaction was stopped with 1 ml 1M Glycin for 5 min. The pellet was then washed three times with cold PBS and stored until further processing at -80°C. Cells were then resuspended in 0,9 ml MNase buffer including protease inhibitor and 80U MNase were added for 15 min at 37°C while shaking the cells. Then, 100 µl 10x covaris sonication buffer were added, the mixture was incubated on ice for 5 min and all samples were sonicated with Covaris with the following settings: Burst 200, Cycle 20%, Intensity 8, 20 min. The chromatin was then pelletized at 14800 rpm, 4°C for 10 min and the supernatant was taken for further processing. The chromatin concentration was determined with the Qubit fluorometer and for each immunoprecipitation 4.4 µg HeLa chromatin was used and supplemented with 10% (440 ng) chromatin derived from mouse embryonic stem cells. The supernatants were the pre-cleared by incubation with Protein G beads for 30 min at 4°C on a rotating platform. The beads were pelleted, 10% of the supernatant taken as input and the rest distributed for immunoprecipitation with antibodies against H3K27me3, H3 and normal rabbit IgG. Antibody-chromatin complexes were allowed to form for 2 h at 4°C on a rotating platform. 25 µl Protein G beads (NEB CHIP-Grade, binding capacity 0.4 µg/µl) were pre-equilibrated with Covaris sonication buffer twice for 10 min and incubated with the antibody-chromatin complexes at 4°C over night. The beads were then washed with Covaris buffer, high-salt buffer, Li-buffer and twice with TE buffer, each for 5 min. Final elution was done with 250 µl elution buffer twice for 15 min at 37°C. To reverse-crosslink samples 0.5 µl RNase A (10 mg/ml) were added to the elution buffer and incubated for 1 h at 37°C. Then 20 µl 5M NaCl and 2 µl Proteinase K (20 mg/ml) were added and incubated over night at 65°C. The next day 1 volume of isopropanol (500 µl), 1/10 ammoniumacetat (50 µl) and 1,5 µl glycoblue (Ambion, USA) were added and incubated for 60 min at 20°C. The precipitation

was centrifuged at 14800 rpm, 4°C for 30min. The pellet was washed once with 75% ethanol, dried and dissolved in 20 µl H_2O.

Library preparation for sequencing

Immunoprecipitated DNA from incubations with antibodies against H3K27me3, H3 and normal rabbit IgG were cloned into a multiplexed library using the NEBNext Ultra DNA Library Prep Kit (Illumina, USA) exactly according to the manufacturers instructions. All GSK343 treated samples and DMSO treated samples from two replicates were pooled together, respectively. Peak size and quality of the libraries were checked on a Tape station. Each pool was sequenced on 3 lanes as 50-bp reads on the Illumina HiSeq 2000 platform sequencer (Genomics & Proteomics Core Facility, DKFZ).

Analysis of ChIP-seq data

The quality of the raw sequence data from H3K27me3 ChIPs and the corresponding controls was monitored using FastQC (Andrews, 2010). Sequences were mapped to the human and mouse genomes using bowtie (Langmead et al, 2009) allowing for 2 mismatches with the following options: human: bowtie -t --chunkmbs 256 --best --strata -v 2 -m 1 -e 70 --seed 12345678 hs37d5; mouse: bowtie -t --chunkmbs 256 --best --strata -v 2 -m 1 -e 70 --seed 12345678 mm9. The ambiguity of reads mapping to both the human and the mouse genome was checked by applying BEDtools (intersectBed, options: -f 1 -r -v) (Quinlan & Hall, 2010). All reads in the H3K27me3 sample were normalized by subtraction of a weighted control IP signal obtained from the IgG control and enrichment over the background signal (input) (Molitor et al, 2015). To check for the quality of the replicates pseudo-autocorrelations were calculated using multi-scale correlation evaluation (MCORE) (Molitor et al, 2015). Peaks of H3K27me3 enrichment were called using Sicer (Zang et al, 2009) with the following options: 1 200 75 0.7 600 1e-03. Overlapping peaks in the DMSO and GSK343 treated samples were identified using the intersectBed tool (options: -f 5e-8). The

position of Sicer-identified peaks relative to annotated genes, introns, exons, 5'-UTRs and 3'-UTRs was classified using intersecBed (options: -f 5e-8) with RefSeq annotations as the reference (Pruitt et al, 2014). Traces of the control and ChIP samples were visualized with the Integrative Genomics Viewer (IGV) (Robinson et al, 2011).

2.20 Fluorescence-activated cell sorting

Cells were washed once with PBS, trypsinized and counted. The same number of cells were pelletized and resuspended in 1 ml cold PBS. 3 ml cold ethanol was added dropwise to each cell suspension under constant mixing. To pelletize, the suspensions were centrifuged for 5 min, 400 x g. The pellets were resuspended in 200 µL cold PBS, 0.25 % Triton X-100 (v/v), and incubated on ice for 5 min. After washing the pellets three times with 1 % BSA/PBS solution 100 µL of α-H3K27me3-antibody (Table 5) were added, and incubated over night at 4 °C on a rotor. After washing the cells again three times with 1 % BSA in PBS, they were incubated with the secondary antibody (α-rabbit-Alexa633, Table 6) for 2 h at 4 °C on a rotor in the dark. The cells were washed again three times and resuspended in 200 µL of 1 % BSA/PBS. To singularize the cells, they were filtered before FACS analysis. The FACS analysis was performed with the FACSCanto II, measuring at least 10000 cells of each sample.

2.21 Microinjection

For microinjection of living cells, HeLa were grown in Nunc Lab-Tek chamber slides. Microinjection was performed with a computer-assisted system (AIS2, CellBiology Trading). Injections of approximately 50 fl of volumes were performed using a needle with a tip-diameter of ca. 300 nm, 150 hPA pressure and 0.5 sec injection time. For 10 µl injection mix, 2 µl propidium iodide (PI, 1 mg/ml) were mixed with 1 µg of RNA in PBS or 1 µg of fluorescein labeled RNA in PBS was used directly. Following microinjection, cells were cultured under standard

conditions for 15 min and prepared for confocal microscopy image acquisition by fixation and staining as indicated.

2.22 Fluorescence microscopy

Confocal imaging was done with a Leica TCS SP5 confocal laser scanning microscope (CLSM) equipped with a HCX PL APO lambda blue 63x/1.4 NA oil immersion objective (Leica Microsystems CMS GmbH, Mannheim, Germany). A near UV diode, diode-pumped solid-state, argon and helium-neon lasers were used for DAPI (λ = 405 nm), Alexa 488 or GFP (λ = 488 nm), Alexa 568 (λ = 561 nm) or Atto-565 and Alexa 633 (λ = 633 nm) excitation. For multi-color analysis sequential image acquisition was applied. The emission detection ranges were adjusted to minimize crosstalk between the different channels. The detection pinhole had a diameter corresponding to one airy disk.

2.23 Image analysis

ImageJ (Schneider et al, 2012) was used for the analysis of microscopy pictures. For analysis of the size of particles images were segmented via thresholding and the function “Analyze Particles” was used to automatically measure the size of *lac*O arrays. To measure the enrichment of protein factors at the *lac*O arrays the fluorescent signal of the MS2-GFP-LacI protein was used to identify the position of the arrays. The mean intensity values of the immunofluorescent staining against protein of interest were then measured and the enrichment over the background signal in the nucleus of the same cell was calculated. In both cases only cells that showed a positive signal for the MS2-stem loops RNA-FISH at the arrays were taken into consideration.

Protein levels (and H3K27me3 levels) on western blots were analyzed by selecting the respective lanes on the western blot and displaying the corresponding intensity histograms of the lanes. The area under the peak was calculated and the ratio to the loading control determined.

Results

1 Methods to detect and characterize nuclear RNAs were established

Over the last years, accumulating evidence showed that RNA is a crucial factor in shaping the genome and its nuclear environment. In a report by Caudron-Herger and colleagues, it was demonstrated that digestion of single-stranded RNAs by RNase A induces a distinct micrometer scale chromatin aggregation of decondensed chromatin regions (Caudron-Herger et al, 2011). After purification of an RNA fraction that allows rescuing this aggregation, the associated transcripts were found to be enriched in long 3'-UTR sequences. It is proposed that they play a role in organizing transcriptionally active chromatin. Similarly, it has been observed that depletion of RNA Pol II transcripts from the nucleus results in the dissolution of nucleoli into so-called "necklace" structures (Granick, 1975; Haaf & Ward, 1996). Motivated by these findings, the goal of this thesis was to shed light on the role of long single-stranded RNAs in chromatin organization.

In order to do so, it was essential to establish methods to visualize the RNAs of interest in the cell and to investigate their functions. Building on previously described techniques, two methods were advanced here to achieve these goals: First, microinjection of fluorescently labeled RNA was applied to monitor subcellular localization of RNAs and to explore their functions on the time-scale of minutes after microinjection. Second, a system involving the MS2 coat protein and MS2 stem loop-tagged transcripts was established. It was used to investigate the influence of specific RNAs on the chromatin when they are immobilized at a specific locus. Using these methods, particular focus was laid on the chromatin organizing function of 3'-UTR sequences with respect to regulating the expression of transcripts variants at their locus. Furthermore, also intronic, RNA Pol II derived transcripts that were suggested to play a role in nuclear organization were examined in more detail. The latter were primarily addressed in

the first part of this work, where the main focus was on establishing the methods to detect and characterized nuclear RNAs.

1.1 Microinjection of RNA allows the investigation of their cellular localization and their chromatin organizing function

First, microinjection of RNAs directly into the nuclei of living cells was chosen as a method to observe transcripts independently of a foreign sequence attached to it. For all microinjection experiments shown here a computer assisted microinjection system was used resulting in the efficient treatment of a high number of cells as needed to perform statistical analyses of the observed phenotype (Figure 6).

In vitro transcribed, microinjected RNAs show different subcellular localization

It was tested whether microinjection can be used to investigate the subcellular localization of various RNAs and how different RNAs behave when microinjected directly into the nuclei of living cells. To this end, specific transcripts derived from selected 3'-UTRs were produced *in vitro* in the presence of fluorescein-labeled UTP and microinjected directly into the nuclei of HeLa cells. Assuming an injection volume of approximately 50 fl, every cell received a total amount of approximately 50 fg RNA. The cells were then incubated for 15 minutes before fixation to allow the microinjected RNAs to relocalize from the point of injection to their final position according to their physiological properties. Figure 8 shows that different RNAs behaved differently: Figure 8A depicts RNAs that were retained within the nucleus. In the bottom panel, it is also noted that the microinjected RNA did not leave the cell of interest to enter a neighboring cell. Only the cell that was injected showed the fluorescent signal emitted from the fluorescein-labeled transcript. Figure 8B displays examples of RNAs that were also found outside the nucleus. In the top panel, a cell is depicted where the transcript distributed evenly over the whole cell including nucleus and cytoplasm. The bottom panel illustrates

an RNA that left the nucleus to completely redistribute to the cytoplasm even when microinjected into the nucleus. In conclusion, RNAs of various nature with putatively different physiological roles also behaved differently when microinjected into the nuclei of living cells. It is observed that they redistributed distinctly on the minutes time-scale. Microinjection is therefore a suitable tool to investigate whether an RNA transcript preferentially resides in the nucleus or distributes to the cytoplasm in the living cell.

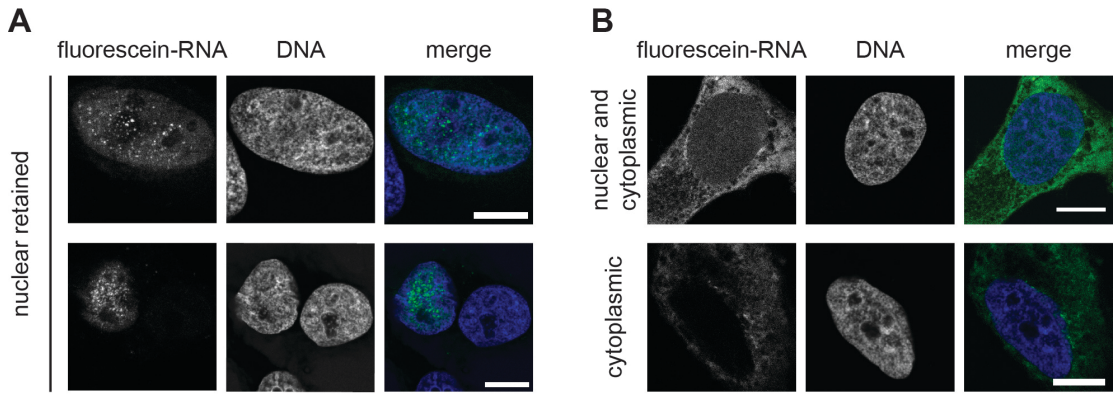


Figure 8: Microinjection as a tool to investigate subcellular localization of RNAs. RNA transcripts were produced *in vitro* in the presence of fluorescein-labeled UTP and microinjected into the nuclei of living cells. The RNAs were allowed to redistribute over the subcellular compartments for 15 minutes. Cells were fixed for image acquisition and counterstained with DAPI. **(A)** Confocal laser scanning microscopy (CLSM) images of microinjected, nuclear retained 3'-UTRs of *CDV3* (top panel) and *EIF2S3* (bottom panel) are shown. **(B)** CLSM images of the microinjected anti-sense 3'-UTRs of *STARD7* (top panel) and *FADS* (bottom panel) are shown as examples of RNAs that left the nucleus to partly or fully redistribute to the cytoplasm. Green: fluorescein-labeled RNA, blue: DAPI, scale bars: 10 µm.

A pool of purified RNAs can perform its physiological function after microinjection

The next step was to test whether microinjected RNAs can still perform their physiological function. HeLa cells were treated with the RNA Pol II inhibitor α-amanitin for five hours. As demonstrated previously by Haaf and colleagues, this caused the loss of nucleolar structure, an increase in the number of nucleolar domains and a decrease in their size (Haaf & Ward, 1996) (Figure 9A). To test whether purified nucleolar RNA can rescue the disrupted phenotype upon microinjection, nucleolar RNA was extracted from HeLa cells. Propidium iodide

(PI) was added to the microinjection mix in order to be able to identify the microinjected cells. PI binds nucleic acids (Suzuki et al, 1997) and was therefore a well-suited fluorescent agent to mark the cells. A-amanitin-treated HeLa cells were microinjected with the RNA/PI mixture and the RNAs were allowed to redistribute and to perform their activity for 30 minutes before cells were fixed. The number and size of nucleolar domains per cell – visualized by immunostaining of the nucleolar protein nucleophosmin (NPM) – was taken as a measure for the extent to which the RNA pool was able to trigger nucleolar reassembly. As depicted in the exemplary CLSM images in Figure 9A, microinjection of RNA from purified nucleoli led to a decrease in the number of nucleolar bodies. This correlated with an increase in their size. A quantitative analysis of the images (Figure 9B) indicated that indeed only the nucleolar RNA fraction but not the buffer or the total RNA of the cells could trigger nucleolar reassembly.

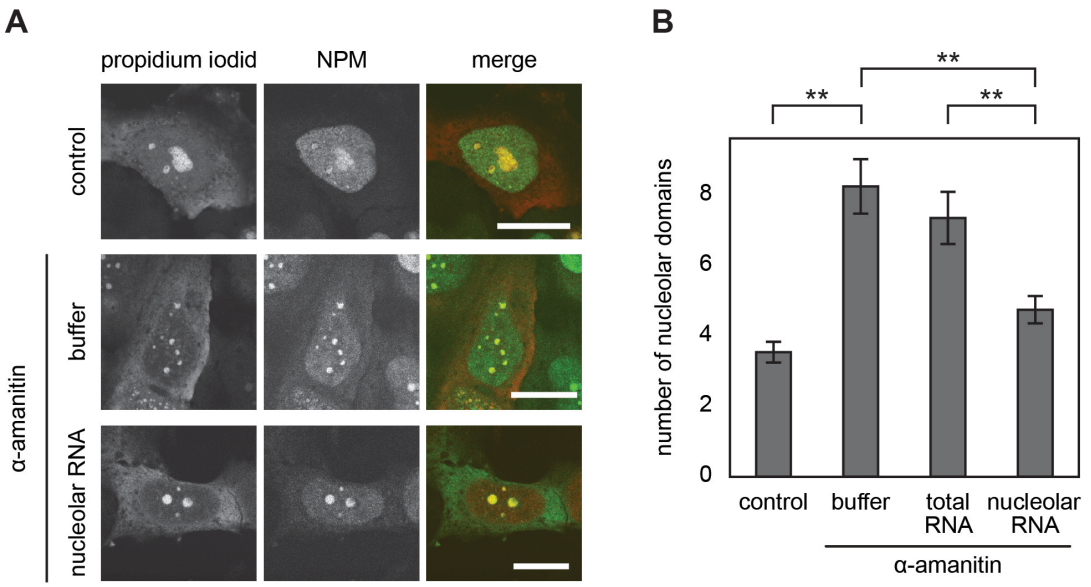


Figure 9: Microinjection of purified nucleolar RNA into the nuclei of living cells. HeLa cells were treated with α-amanitin for 5 hours and then microinjected with buffer, total RNA or the purified nucleolar RNA fraction. After fixation nucleolar domains were visualized by immunofluorescence of NPM. **(A)** CLSM images of an untreated cell injected with PI containing buffer only (top panel) and α-amanitin-treated cells microinjected with either buffer only (middle panel) or the nucleolar RNA fraction (bottom panel). Red: PI, green: NPM, scale bars: 10 µm. **(B)** The graph represents the average number of nucleolar domains (+/- 95% confidence interval (CI)) based on the analysis of 90, 87, 80 and 86 cells, respectively. **p<0.01 (Student's t-test). Adapted from Caudron-Herger et al, 2015.

First, this indicates that the nucleolar RNA fraction contained RNA transcripts that counteracted the α-amanitin-induced segregation of nucleolar regions. Second, this experiment demonstrates that the microinjected nucleolar RNA pool was active in the cell and could perform its functions. Finally, it shows that microinjection is a powerful technique to differentiate the physiological roles of different RNA pools in the nucleus.

Single in vitro transcribed RNAs can perform their function after microinjection

The previous section suggested that a whole RNA pool purified from the nucleolus can perform its nucleolus-assembling function after microinjection into α-amanitin-treated cells. It was therefore hypothesized that single transcripts that are highly enriched in this RNA fraction could also be able to show this phenotype after microinjection. The most enriched transcripts found in the nucleolar RNA fraction were RNA Pol II-transcribed repetitive intronic Alu elements, termed *alu*RNAs (Caudron-Herger et al, 2015). Therefore, a prototypic *alu*RNA was produced *in vitro*, mixed with PI and microinjected into HeLa cells that had been treated with α-amanitin for 5 hours. Again, the number of nucleolar domains per cell was taken as a measure for the extent, to which the *alu*RNA was able to trigger nucleolar reassembly. Nucleolar domains were identified by PI staining. As depicted in Figure 10A, microinjection of *in vitro* transcribed *alu*RNAs triggered reassembly of nucleolar particles as manifested by a decrease in the number of nucleolar domains. As controls, the reverse sequence of the prototypic *alu*RNA and an RNA originating from another frequent repetitive element, the L1 repeat, were used. Notably, the forward *alu*RNA was most efficient in rescuing the nucleolar dispersion in a quantitative comparison to the reverse *alu*RNA, the L1-repeat RNA transcript, or the buffer only Figure 10B). As described above for the whole pool of nucleolar RNA, this shows that also single *in vitro* transcribed RNAs can perform their functions when microinjected into living cells. In conclusion, microinjection is a well-suited technique to investigate the cellular localizations and functions of single RNA transcripts as well as RNA pools.

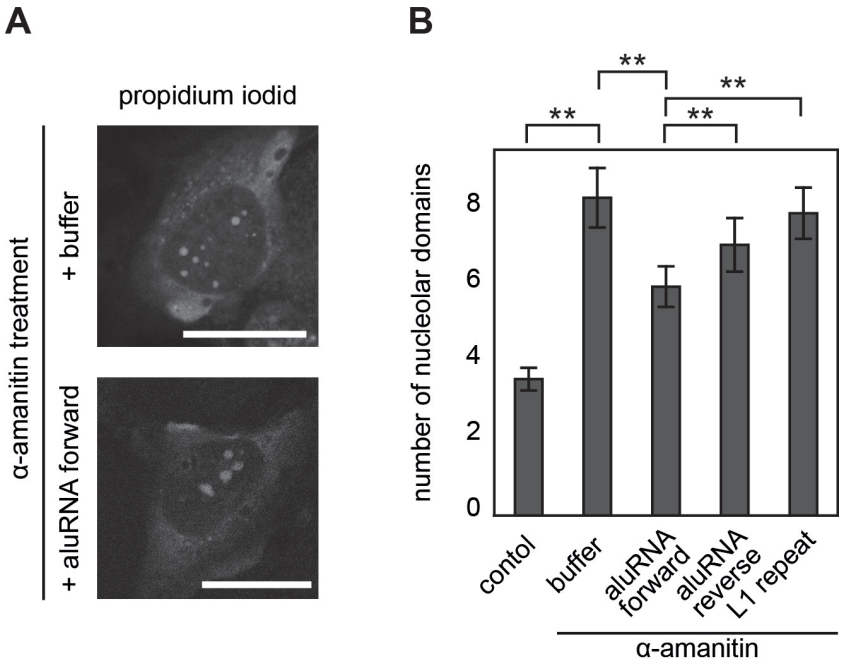


Figure 10: Microinjection of an *in vitro*-transcribed prototypic *alu*RNA into the nuclei of living cells. HeLa cells were treated with α-amanitin for 5 hours and microinjected with forward *alu*RNA, reverse *alu*RNA and L1-repeat RNA or buffer only. Nucleoli were visualized by their enrichment in PI. **(A)** CLSM images of an α-amanitin-treated cell microinjected with buffer only (top) or the forward *alu*RNA transcript (bottom). Scale bars: 10 µm. **(B)** The graph represents average number of nucleolar domains (+/- 95% CI) base on the analysis of 90, 87, 86, 83 and 86 cells, respectively. **p < 0.01 (Student's t-test). Adapted from Caudron-Herger et al, 2015.

1.2 Specific RNAs immobilized at a *lac*O array in living cells can act as chromatin modifiers

In order to assess more than the global effect of RNA on nuclear structure, it is also very informative to quantify the effects of a specific RNA at a defined genomic locus. Amongst others, the following features can be investigated with a model system such as the *lac*O arrays that were used here: recruitment of putative protein interaction partners, the impact on epigenetic chromatin modifications such as posttranslational modifications of histones, the effect on chromatin compaction, or the localization of the genomic locus in respect to other nuclear sub-compartments. Here, I established an experimental system to immobilize RNAs at a specific genomic locus. It made use of two highly affine and highly specific nucleic acid-protein interactions: First, mammalian cells that have stable genomic integrations of an array of the bacterial *lac*O sequence, which are bound by the LacI repressor protein, were employed (Belmont &

Straight, 1998; Jegou et al, 2009; Robinett et al, 1996) (Figure 7A). Second, the RNAs of interest were tagged with the bacteriophage MS2 stem loop sequence, which is recognized and bound by the MS2 coat protein (Johansson et al, 1998; Shevtsov & Dundr, 2011). The two were combined by transfecting a plasmid encoding the MS2-tagged RNA and a fusion protein of MS2 and LacI into cells with stably integrated *lac*O arrays. The localization of the *lac*O array within the nucleus could be visualized by introducing a fluorescent protein, e.g. GFP, to the fusion construct of MS2 and LacI. To ensure efficient recruitment, plasmids encoding for 18 sequential MS2 stem loops were used. Taken together, this provided a toolkit that could be used to immobilize any RNA at a specific, known chromatin locus and to evaluate its activity at this locus. Methods to detect the RNA enriched at the array and to visualize potential protein interaction partners are described in the following (Figure 7B).

Various cell lines containing stable *lac*O integrations are available, in which the arrays are integrated at different locations in the genome. Here, two human cell lines were selected, called EC4 and F4IIB8, both of which are derived from the human osteosarcoma cell line U2OS (Jegou et al, 2009). Additionally, a Chinese hamster ovary (CHO) cell line, called AO3, was used (Luijsterburg et al, 2012; Tumbar et al, 1999). All three cell lines have a single integration of multiple copies of the *lac*O repeat.

Ethynyl-uridine incorporation or RNA-FISH can be used to visualize RNA enrichment at lacO arrays in the MS2-LacI-mediated recruitment system

The first step of establishing the MS2/LacI-mediated recruitment system was to find suited methods to detect the enrichment of RNA at the arrays. There are several methods to do so, two of which will be described in the following. First, RNA enrichment at the array was visualized by using ethynyl-uridine (EU), which is incubated with the cells before fixation so that RNA Pol II incorporates it into freshly made RNA. The EU containing RNA transcripts were then made visible by attaching a fluorophore to the EU using an azide-alkyne cycloaddition reaction, called “click” chemistry (Liang & Astruc, 2011). EC4 cells were transfected with

the MS2-GFP-LacI fusion protein and a plasmid encoding for an MS2 stem loop-tagged RNA oligonucleotide of 60 nucleotides in length. 6 hours after transfection EU was added to the medium of the transfected cells and allowed to be incorporated into RNAs overnight. Figure 11A shows a cell, in which a clear enrichment of the EU signal could be seen at the arrays, visualized by the GFP-tagged LacI protein. In this experimental setup EU is incorporated into all newly synthesized RNAs, not just the MS2-tagged RNAs. Nonetheless, the enrichment of the RNA at the array was strong enough to identify the array via its EU signal in comparison to the nuclear background signal. It is therefore concluded that EU incorporation and labeling by “click” chemistry is a suited method to assess whether RNA is enriched at the *lac*O array when recruiting it there via the MS2/LacI-mediated recruitment system.

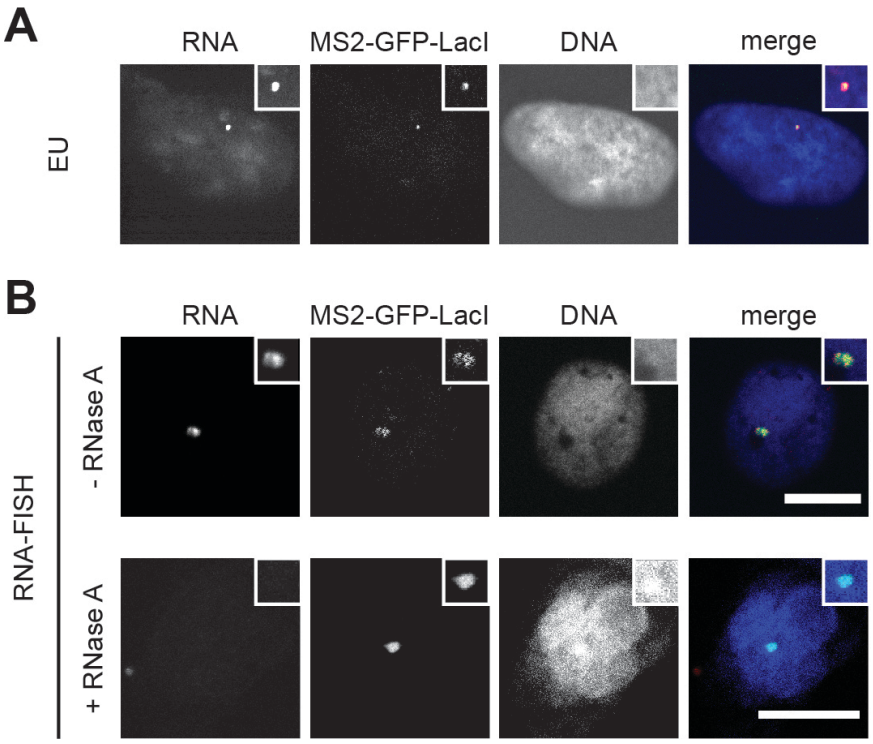


Figure 11: Visualization of MS2 stem loop-tagged RNA recruited to *lac*O arrays with a fusion protein of MS2-GFP-LacI. Cells were transfected with the MS2-GFP-LacI fusion protein and a plasmid encoding for an MS2 stem loop-tagged RNA oligonucleotide of 60 nucleotides. **(A)** EC4 cells were treated with EU overnight and EU was marked with a fluorophore using “click” chemistry after fixation. **(B)** AO3 cells were fixed 24 hours after transfection and RNA-FISH with an ATTO-565-labeled probe directed against the MS2 stem loop sequence was performed. RNase A-treated cells were included in order to rule out unspecific binding of the RNA probe to DNA sequences. Insets show enlargements of the *lac*O arrays. Red: RNA, green: MS2-GFP-LacI, blue: DAPI, scale bars: 10 µm.

The second method involves RNA fluorescence *in situ* hybridization (RNA-FISH) with specific probes directed against the MS2 stem loop sequence (schematically depicted in Figure 7B). AO3 cells were transfected with the MS2-GFP-LacI fusion protein and the same plasmid encoding for a short MS2 stem loop-tagged RNA as above. Cells were fixed 24 hours after transfection and an RNA-FISH with a fluorescently labeled probe directed against the MS2 stem loop sequence was performed (Goodier et al, 2010). To ensure that the sequence detected by the RNA-FISH is in fact the RNA sequence and not the DNA sequence of the plasmid, an RNase A-treated control was included in the analysis. Since RNase A cleaves all single stranded RNAs including the MS2 stem loop-tagged RNA no RNA-FISH signal was expected after RNase A treatment. As depicted in Figure 11B, an enrichment of the RNA-FISH signal that co-localizes with the MS2-GPF-LacI fusion protein could be seen (Figure 11B**,** upper panel), whereas no RNA-FISH signal was detected at the array after RNase A treatment (Figure 11B**,** bottom panel). Due to the high affinity of the MS2 stem loops to the MS2 coat protein immobilized at the array, most of the MS2-tagged RNA was expected to be found at the array whereas very little should be distributed over the nucleus. Indeed, the background signal was very low. It is therefore reasoned that RNA-FISH directed against the MS2 stem loop sequence is a well-suited method to detect enrichment of the MS2 stem loop-tagged RNA at genomic *lac*O sequences. From both the RNA detection by EU and by RNA-FISH it can furthermore be concluded that the *lac*O system works well to locally tether and enrich an RNA of interest at the *lac*O arrays.

*Immobilized RNAs at the lac*O *arrays recruit protein interaction partners*

The next step was to investigate whether the RNA of interest is able to recruit its known protein interaction partners and to perform its function. As before, the cell lines that have stable *lac*O array integrations and the MS2-GFP-LacI fusion protein were used. As shown above *alu*RNA is a nucleolar enriched RNA that could rescue the nucleolar dispersion induced by RNA Pol II inhibition. Given that *alu*RNA has been shown to play a role in nucleolar structure, it was hypothesized

that it could also interact with prominent nucleolar proteins such as the nucleolin (NCL). Here, it was therefore tested whether *alu*RNA interacts with NCL in the living cell. F4BII8 cells were transfected with MS2-GFP-LacI and the MS2 stem loop-tagged *alu*RNA. 24 hours after transfection cells were fixed and an immunostaining directed against NCL was performed. Figure 12 shows that MS2-*alu*RNA induced recruitment of NCL to the *lac*O array. This observation suggests that indeed *alu*RNA interacts with the nucleolar protein NCL to play a role in nucleolar organization. Furthermore, this shows that the MS2/LacI-mediated recruitment system can be used to study RNA – protein interactions in living cells.

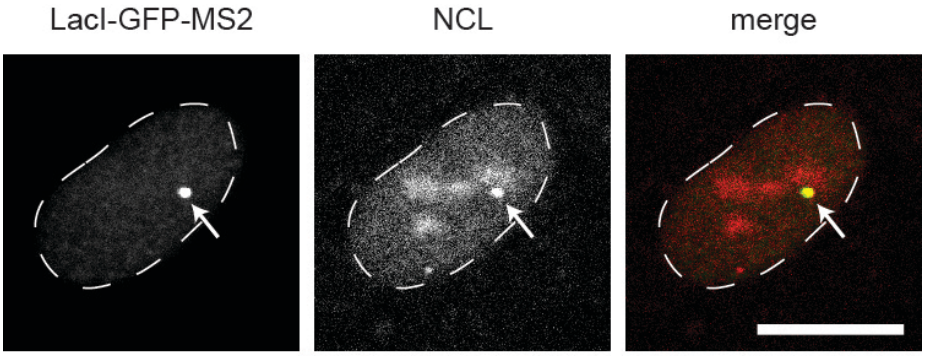


Figure 12: Ectopically expressed and *lac*O array-immobilized MS2-*alu*RNA recruited the nucleolar protein NCL. F4IIB8 cells were transfected with forward MS2-aluRNA and MS2-GPF-LacI (shown in green) and an immunostaining against NCL (shown in red) was performed. Localization of MS2-GPF-LacI and NCL are shown and the arrow indicates the *lac*O array, which is associated with a nucleolar domain. Green: LacI-GFP-MS2, red: NCL, scale bar: 10 µm. Adapted from Caudron-Herger et al, 2015.

Immobilized RNAs at the lacO array influence nuclear localization of the arrays

Next, it was investigated whether the RNAs also perform their function in terms of influencing the sub-nuclear localization of the arrays depending on which RNA is tethered to the array. In this experiment, it was therefore analyzed whether tethering *alu*RNA to the arrays also has an influence on the localization of the arrays within the nucleus. It was previously shown that *alu*RNAs are nucleolar localized RNAs and that they interact with NCL. Hypothetically, *alu*RNA could therefore be able to recruit the *lac*O arrays to the nucleoli driven by their own and

their protein interaction partner's affinity for the nucleolus. To test whether tethering the *alu*RNAs to the *lac*O arrays is a suited system to investigate this, F4IIB8 cells were again transfected with MS2-*alu*RNA and MS2-GFP-LacI. As controls, the MS2 stem loops alone, the two separate arms of the Alu sequence (left, L and right, R) and a selected set of further MS2-tagged RNAs of various lengths were included in the analysis. Using an immunostaining of NPM the nucleolar localization of the *lac*O arrays with respect to the nucleolus was analyzed quantitatively. Figure 13A displays an exemplary CLSM image of a cell where a *lac*O array was located in close proximity to a nucleolus. A quantitative analysis of the percentage of arrays that show nucleolar localization after MS2-*alu*RNA recruitment is depicted (Figure 13B). It is noted that both the full-length *alu*RNA and its right arm fragment *alu*RNA$_R$ significantly increased the number of *lac*O arrays localizing in the nucleoli.

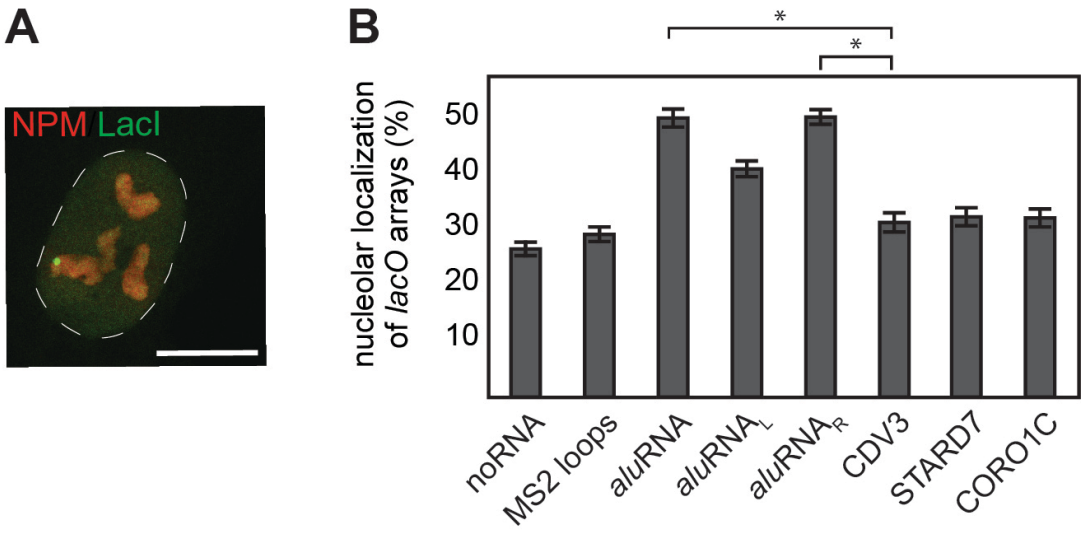


Figure 13: Ectopically expressed and *lac*O array-immobilized MS2-*alu*RNA caused nucleolar localization of the *lac*O arrays. F4IIB8 cells were transfected with the MS2 loops-containing *alu*RNA, its left arm (L), or right arm (R), and the MS2-GFP-LacI fusion protein. The MS2 stem loops only, MS2-tagged 3'-UTRs of *CDV3*, *STARD7* and *CORO1C* served as controls. An immostaining against NPM was performed. **(A)** Exemplary CLSM image after recruitment of MS2-*alu*RNA showing the array in proximity of a nucleolus. Green: MS2-GFP-LacI, red: NPM, scale bar: 10 µm. **(B)** Plot of the percentage of nucleolar localization as evaluated by the average number of *lac*O arrays detected in nucleoli. n>100, *p<0.05 (Student's t-test), (+/- 95% CI). Adapted from Caudron-Herger et al, 2015.

This suggests that *alu*RNA not only interacts with NCL as shown above but also targets genomic loci to the nucleoli. This also demonstrates that the experimental model system used here allows for investigation of the localization of the arrays in dependence of the recruited RNA. These results reinforces that the MS2/LacI-mediated recruitment system is well-suited to evaluate the physiological behaviors of immobilized RNAs at the chromatin.

The above sections describe the methodologies of microinjection and MS2/LacI-mediated recruitment to chromatin in the context of the interplay of the nucleolus with *alu*RNA. In the following parts of this thesis these two methods were used to characterize the role of 3'-UTRs in chromatin organization. Microinjection was employed to determine the sub-cellular localization of putative chromatin-organizing 3'-UTR candidates. The recruitment of RNAs to the *lac*O arrays was used to investigate the effect of specific 3'-UTR sequences on the local chromatin at the array: their effect on chromatin compaction, histone modifications and recruitment of chromatin modifying enzymes was examined.

2 3'-UTRs have chromatin organizing function

3'-UTRs are known for their roles in regulating mRNA stability and protein translation (Matoulkova et al, 2012). A number of publications have suggested additional functions of 3'-UTRs. For example, specific 3'-UTRs have been reported to exist independently of the protein coding sequence of the corresponding mRNA transcripts (Furuno et al, 2006). Moreover, 3'-UTR transcripts can be found in the nucleus independently of the protein coding part of the mRNA (Mercer et al, 2011). Also, it has been reported that digestion of single stranded RNAs in the nuclei of mammalian cells causes a distinct micrometer scale chromatin aggregation of decondensed regions (Caudron-Herger et al, 2011). It is suggested that a homogeneous chromatin distribution can be restored by a soluble nuclear RNA fraction that is enriched in long 3'-UTR transcripts, referred to as the F2 fraction. It is proposed that these RNAs could act as

architectural genome organizing factors. Based on these observations, there is evidence that 3'-UTRs play an important role in nuclear architecture and chromatin organization. The goal of the following sections was to identify particularly interesting 3'-UTR sequences with a chromatin organizing function. To do so, the most enriched 3'-UTRs from the above named F2 RNA fraction were investigated. First, they were tested for their potential to rescue RNase A-induced chromatin compaction. Second, their sub-cellular localization was investigated to find out whether these RNAs can be found in the nucleus as expected from an RNA with chromatin organizing function. Third, using the above described and tested MS2/LacI-mediated recruitment system, the effect of these 3'-UTR sequences, when immobilized at a specific chromatin locus, was examined. Combining the results of these three experiments provided criteria for selecting a particularly interesting 3'-UTR candidate, the 3'-UTR of *CDV3* (carnitine deficiency-associated gene expressed in ventricle 3) that was then further investigated with respect to its chromatin organizing function.

2.1 Specific long 3'-UTR sequences can rescue chromatin aggregation induced by RNase A treatment

In the rescue experiments described by Caudron-Herger and colleagues (Caudron-Herger et al, 2011), HeLa cells were permeabilized and treated with RNase A. The RNase A was then washed away and inhibited by specific RNase A inhibitors. Different RNAs were then added to test their propensity to rescue the initial chromatin distribution. The RNase A-induced changes of chromatin organization and their rescue could be visualized by observing changes in the distribution of the DAPI fluorescent signal (Figure 14).

It was tested whether specific 3'-UTRs that where highly enriched in the F2 fraction could rescue RNase A-treated chromatin. For this experiment, the transcripts were produced *in vitro* and tested in the rescue experiment as described Figure 14.

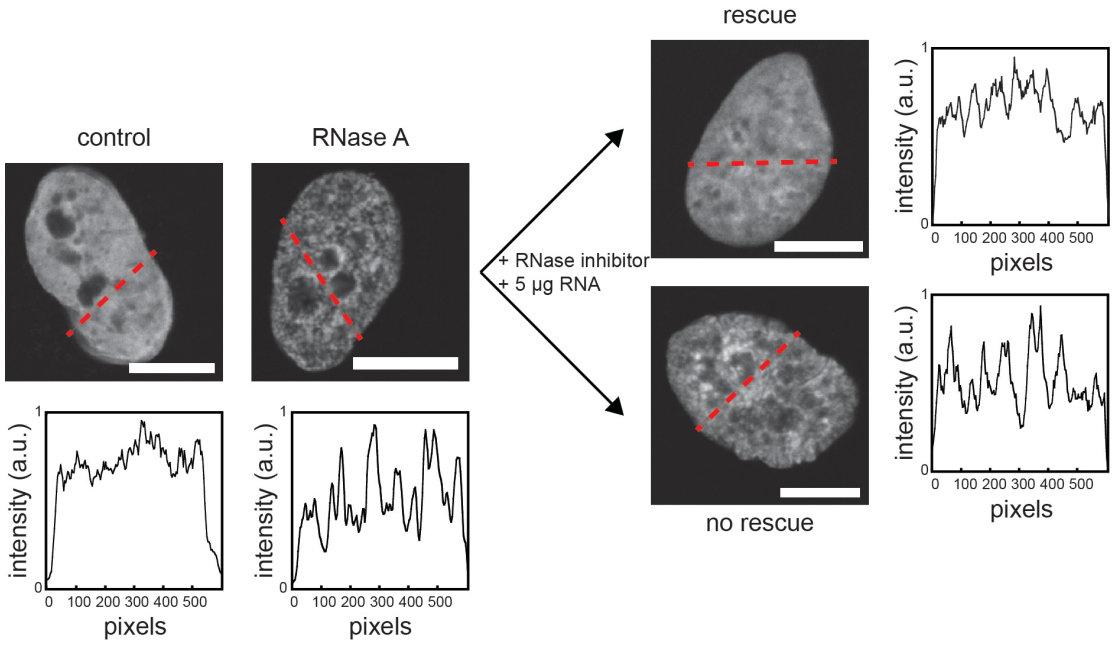


Figure 14: Principle of the rescue experiment. Left panel: DAPI stained control and RNase A-treated cells and intensity profiles along the indicated lines show the more heterogeneous chromatin distribution after RNase A digestion in comparison to the control cell. Right panel: After RNase A treatment, RNase inhibitor and 5 µg of the RNA of interest were added and incubated for 15 minutes to assess the putative potential of the RNA to rescue the initial and more homogenous chromatin distribution. Adapted from Caudron-Herger et al, 2011.

The phenotype of the chromatin was evaluated by visual inspection of CLSM images of at least 100 cells per RNA transcript. The percentage of cells with rescued chromatin was calculated and compared to control cells. As controls, RNAs with similar sizes derived from the intergenic spacer of the rRNA genes were selected. These transcripts were not enriched in the F2 fraction. 3'-UTR transcripts were considered to have rescuing capacity if more than 50% of the cells showed a homogeneous chromatin distribution after rescue. They were considered to have no effect if less than 35% of the cells had homogeneous chromatin. Figure 15A depicts representative CLSM images of five 3'-UTRs that were able to fully reverse the chromatin phenotype after RNase A treatment. Furthermore, five transcripts that partly rescued this phenotype and four transcripts – including the negative controls – that showed no homogeneous chromatin distribution are displayed. A quantification of these results is given in Figure 15B.

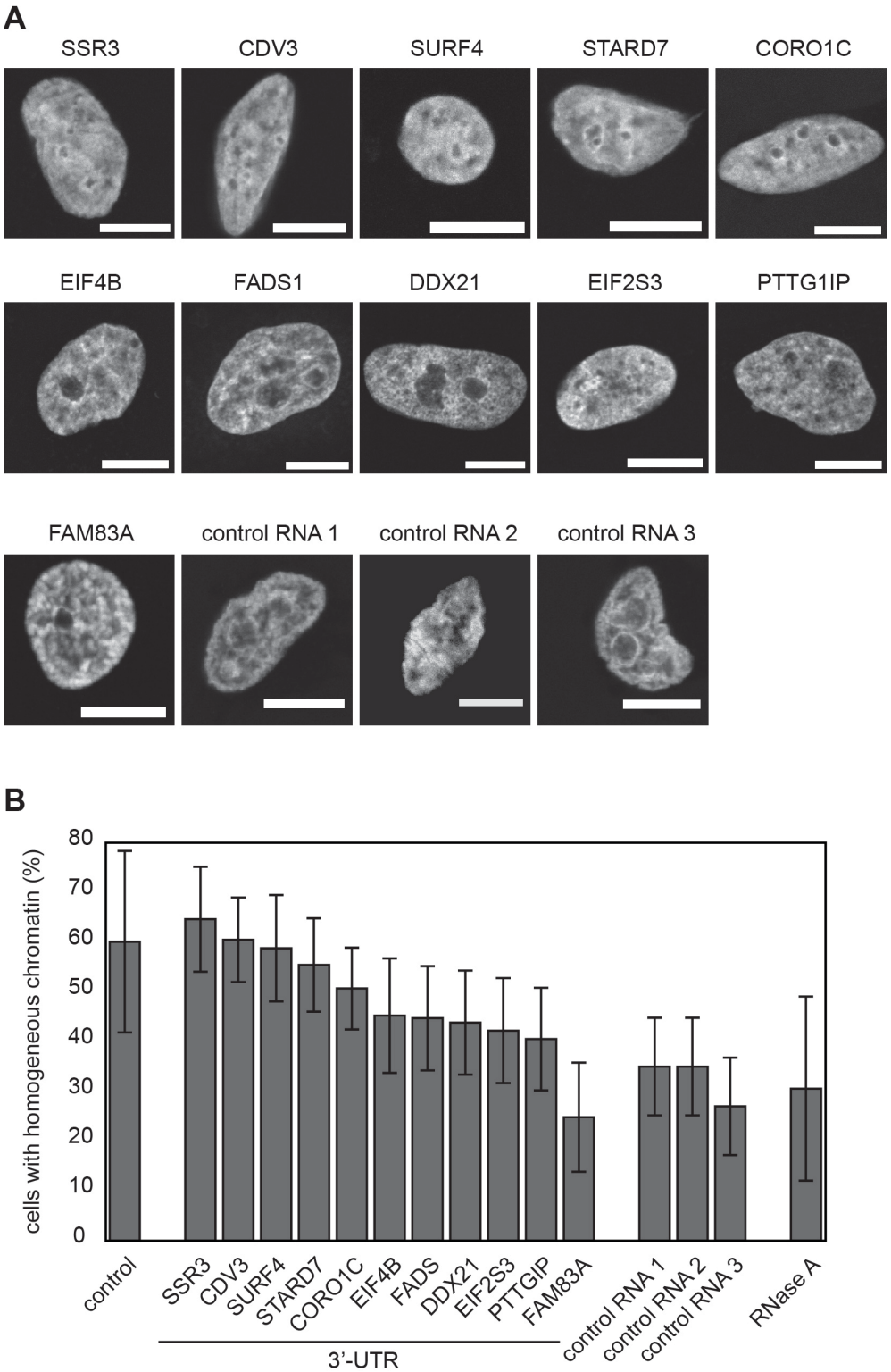


Figure 15: The 3'-UTRs of the *SSR3*, *CDV3*, *SURF3*, *STARD7* and *CORO1C* transcripts rescued chromatin aggregation induced by RNase A treatment. (A) Representative CLSM images of DAPI stained nuclei that were permeabilized, treated with RNase A, and treated with RNase inhibitor and the respective *in vitro*-transcribed 3'-UTR of the gene indicated above the images. RNA sequences derived from the intergenic spacer of the rRNA genes were used as controls. They were not enriched in the RNA pool that can rescue RNase A-treated chromatin. The nuclei were counterstained with DAPI. **(B)** Percentage of cells, in which addition of the indicated RNA rescued the initial homogeneous chromatin distribution after RNase A treatment. n>100.

The 3'-UTRs of the *SSR3* (signal sequence receptor, gamma), *CDV3* (carnitine deficiency-associated gene expressed in ventricle 3), *SURF4* (surface 4 integral membrane protein), *STARD7* (star-related lipid transfer (START) domain containing 7) and *CORO1C* (coronin, actin binding protein, 1C) transcripts were able to rescue the homogeneous chromatin distribution. This suggests that within the F2 fraction, there are specific long 3'-UTR sequences that can reverse the RNase A-induced chromatin aggregation, when used in the rescue assay as single *in vitro* transcribed transcripts. The 3'-UTRs of the *EIF4B* (eukaryotic translation initiation factor 4B), *FADS1* (fatty acid desaturase 1), *DDX21* (DEAD (Asp-Glu-Ala-Asp) box helicase 21), *EIF2S3* (eukaryotic translation initiation factor 2, subunit 3 gamma), *PTTG1IP* (pituitary tumor-transforming 1 interacting protein) and *FAM83A* (family with sequence similarity 83, member A) transcripts proved little or no rescuing capacity. Therefore, clearly, not all 3'-UTR sequences that were enriched in the F2 RNA fraction had the same potential to rescue RNase A-treated chromatin.

2.2 Fluorescently labeled 3'-UTR transcripts, which remain in the nucleus after microinjection, are selected

In the following, it was investigated, whether the 3'-UTRs that could rescue RNase A-treated chromatin had nuclear localization. Only those were more likely to have an *in vivo* nuclear function such as chromatin organization. The 3'-UTR sequences that were tested in the rescue experiment above were therefore produced *in vitro* with fluorescein-labeled UTP and microinjected into the nuclei of HeLa cells as described in section 1.1 of the results. The RNAs were allowed to distribute over the cells within 15 minutes and cells were fixed and counterstained with DAPI for inspection by confocal microscopy. As illustrated in Figure 16, a number of the putative chromatin organizing 3'-UTRs resided in the nucleus (Figure 16A). In contrast, others distributed over the whole cell, i.e. nucleus and cytoplasm (Figure 16B), or even left the nucleus entirely and were exclusively localized in the cytoplasm (Figure 16C).

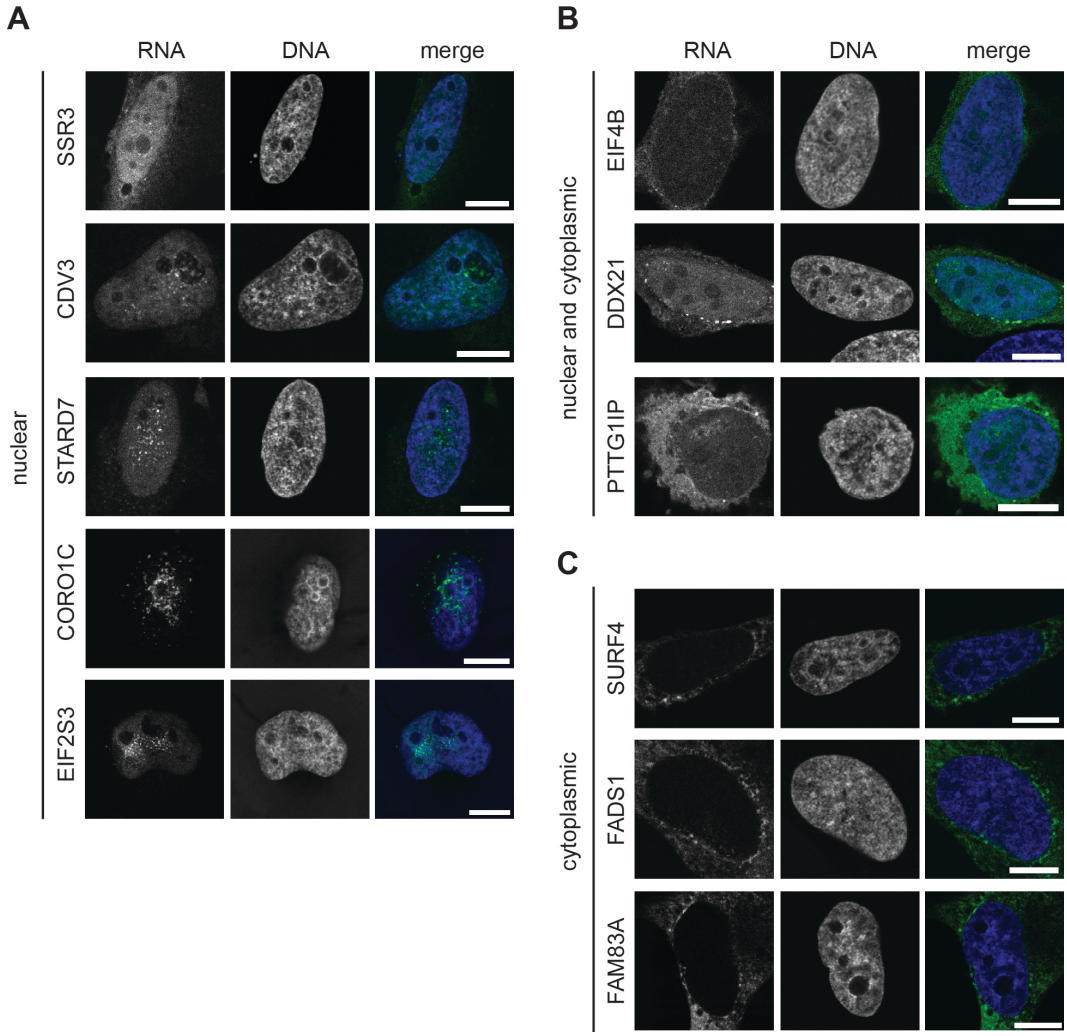


Figure 16: Fluorescently labeled 3'-UTRs of the *SSR3*, *CDV3*, *STAR7* and *CORO1C* transcripts remained in the nucleus after microinjection. HeLa cells were microinjected with the *in vitro* transcribed and fluorescently labeled 3'-UTR sequences of the indicated RNAs. The transcripts were allowed to redistribute within 15 minutes before cells were fixed and counterstained with DAPI. Representative CLSM images are depicted. **(A)** Nuclear retained 3'-UTR sequences, **(B)** 3'-UTR sequences that distribute over nucleus and cytoplasm, **(C)** 3'-UTR sequences that are exclusively found in the cytoplasm after microinjection into the nucleus. Green: RNA, blue: DAPI, scale bars: 10 µm.

Four of the five RNAs that performed well in the rescue experiment described above (Figure 15) were also RNAs that reside in the nucleus after microinjection: the 3'-UTRs of *SSR3*, *CDV3*, *STARD7* and *CORO1C*. SURF4-3'-UTR was the only exception and was exclusively found in the cytoplasm after microinjection into the nucleus. Additionally, EIF2S3-3'-UTR, a 3'-UTR that was not able to fully rescue RNase A-treated chromatin, was also found to remain in the nucleus after microinjection. All other 3'-UTR transcripts that demonstrated no or only limited

rescuing capacity showed distribution over the entire cell or even completely redistributed into the cytoplasm. In summary, this procedure further restricted the selection of 3'-UTRs sequences that are considered as relevant transcripts with chromatin organizing function. Only 3'-UTRs that had the capacity to rescue RNase A-induced chromatin perturbation and that showed nuclear localization after microinjection into the nucleus were further considered in the following analyses. This included the 3'-UTRs of *SSR3*, *CDV3*, *STARD7* and *CORO1C*.

2.3 The 3'-UTR of *CDV3* induces changes in chromatin compaction at the *lac*O array

Next, the chromatin organizing function of the *SSR3*-, *CDV3*-, *STARD7*- and *CORO1C*-3'-UTRs was examined in more detail. The MS2/LacI-mediated recruitment system described in section 1.2 of the results was applied, since it is a well-suited assay to examine the behavior of RNAs at the chromatin. Based on the above observations, it was hypothesized that immobilization and local enrichment of the chromatin organizing 3'-UTRs at a genomic locus would affect the chromatin compaction at the array.

AO3 cells containing a stable insertion of the bacterial *lac*O repeat sequence were transfected with a fusion protein of MS2-GFP-LacI and the RNA of interest tagged with 18 MS2 stem loop sequences. AO3 cells were chose for this experiment for two reasons: On the one hand, the usual compact state of their arrays allows for observation of decondensation of the chromatin at the array as demonstrated by Luijsterburg and colleagues (Luijsterburg et al, 2012). On the other hand, their arrays per se are relatively loose so that also potential condensation could be observed. To detect RNA enrichment at the array, RNA-FISH with fluorescently labeled probes directed against the MS2 stem loop sequence was performed. The size of the arrays served as an indicator for the degree of chromatin compaction. It was calculated after segmentation via thresholding and by applying the function "Analyze Particles" in ImageJ

(Schneider et al, 2012). Cells transfected with no RNA or with the MS2 stem loops sequence only were used as controls.

Figure 17A shows representative CLSM images of cells that were transfected with no RNA, the MS2 stem loops and the MS2-tagged 3'-UTR of *CDV3*. The corresponding samples treated with RNase A (bottom panel) are also depicted demonstrating that the RNA-FISH only specifically hybridized to the MS2 RNA and not the plasmid DNA. By visual inspection, it became clear that recruitment of the 3'-UTR of *CDV3* decreased the size of the array. Recruiting just the MS2 loops did not affect the size of the array. This observation was substantiated by a quantitative analysis of the size of the arrays as displayed in Figure 17B. Only the 3'-UTR of *CDV3* but not those of *SSR3*, *STARD7* or *CORO1C* was able to significantly influence the degree of compaction of the *lac*O array.

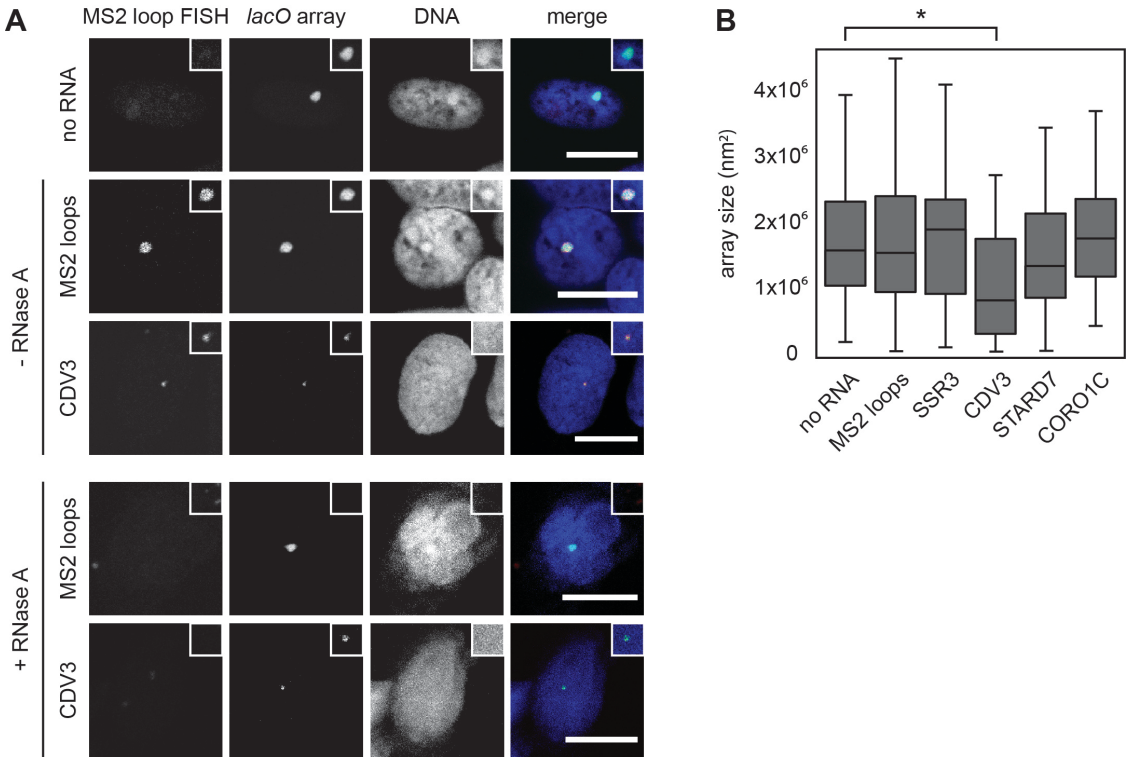


Figure 17: The 3'-UTR of *CDV3* compacted chromatin when immobilized at the *lac*O arrays. AO3 cells containing a stable insertion of the bacterial *lac*O repeat sequence were transfected with a fusion protein of MS2-GFP-LacI and the RNA of interest tagged with the MS2 stem loop sequences. The MS2 loops were detected by RNA-FISH and nuclei counterstained with DAPI. **(A)** CLSM images of cells transfected with no RNA, the MS2 stem loops only or the 3'-UTR of *CDV3* tagged with the MS2 stem loops. Each RNA was transfected twice and one sample of each was treated with RNase A (lower two panels). Insets show enlargements of the *lac*O arrays. Red: RNA-FISH, green: MS2-GFP-LacI, blue: DAPI, scale bar: 10 µm. **(B)** Quantitative analysis of the size of the *lac*O array depending on the RNA recruited to it. The median, the upper and lower quartiles and the 95% confidence intervals are shown. n>50, *p<0.05 (Wilcoxon test**).**

Thus, only the 3'-UTR of the *CDV3* gene fulfilled three criteria that make it a particularly interesting 3'-UTR with chromatin organizing function: i) the transcript can rescue RNase A-treated chromatin as an *in vitro* transcribed RNA, ii) it resides within the nucleus after being microinjected directly into the nuclei of HeLa cells and iii) it compacts chromatin when immobilized at the *lac*O array, that is stably integrated into the chromatin of AO3 cells. Combining the results of these three experiments provided me with criteria for selecting the 3'-UTR of *CDV3* for further analyses on the mechanism by which it influences chromatin and on its endogenous role in chromatin organization.

2.4 *CDV3* is a nuclear and cytoplasmic RNA with several splice variants

CDV3 codes for a protein that is up-regulated in the ventricles of mice with carnitine-deficient juvenile visceral steatosis, a condition, in which mice suffer from fatty liver, hyperammonemia, hypoglycemia and growth retardation (Fukumaru et al, 2002). Earlier reports had already identified the human homolog of *CDV3* as being upregulated in HER-2/NEU over expressing breast tumors (Oh et al, 1999) and as a potential carrier of a small nucleotide polymorphism (SNP) that might affect colorectal cancer risk (Abuli et al, 2011). Intriguingly, one study suggests that the 3'-UTR of *CDV3* is expressed separately from the associated protein-coding sequence, to which it is normally linked, in human and mice (Mercer et al, 2011). All of these previous reports agree that *CDV3* has two splice variants both, of which are expressed as detected by northern blotting and by using anti-sera raised against the protein. Although it carries a nuclear localization signal, the protein is found mainly in the cytoplasm (Tsuchiya et al, 2003). From genome-wide sequencing studies it is now known that there are at least three splicing forms of the *CDV3* transcript, one of which has an alternative transcription start site. The other two differ in their last exon with the *CDV3a* transcript carrying a longer 3'-UTR and the *CDV3b* transcript an alternative exon with a shorter 3'-UTR (Figure 18A). Since this thesis focuses on putative

chromatin organizing signals in the 3'-UTR of the RNA only the latter two splicing variants will be considered in the following and referred to as *CDV3a* and *CDV3b*. For convenience the 3'-UTR of *CDV3a* will in the following be called "CU-RNA", for **C**DV3-3'-**U**TR-RNA. The following experiments aimed at a closer characterization of the *CDV3* transcripts with special focus on their 3'-UTRs.

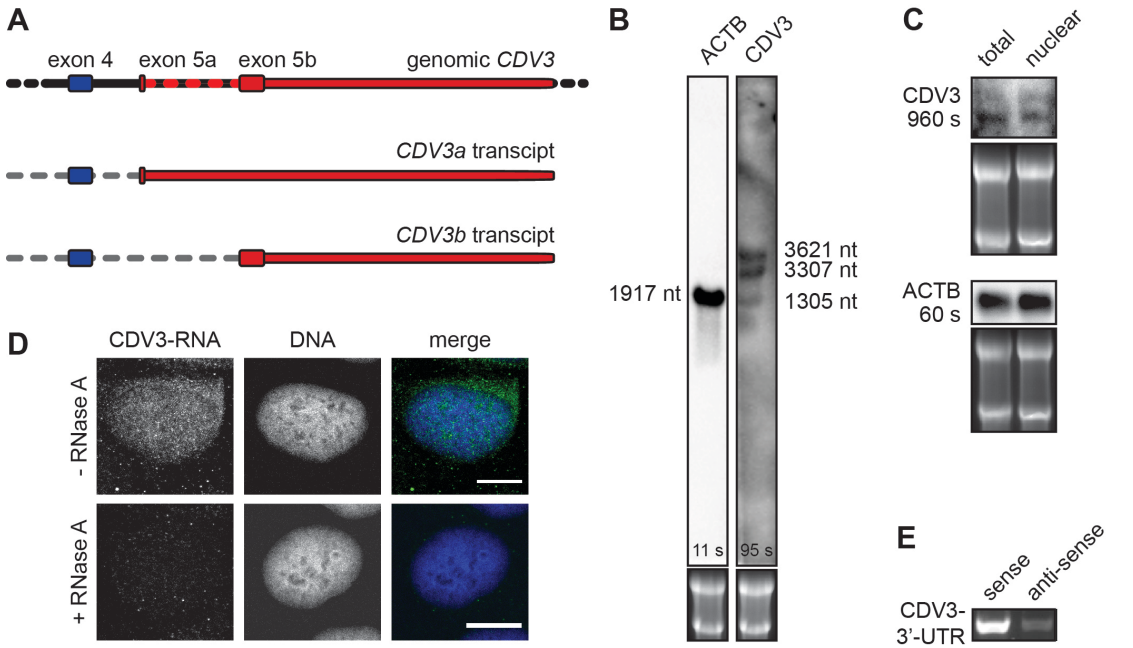


Figure 18: The *CDV3* transcript is a nuclear and cytoplasmic RNA with two transcript variants that differ in their last exon and the length of their 3'-UTR. (A) Schematic illustration of the 3' end of the genomic sequence of *CDV3* and two of its splice variants that differ in the last exon (exon 5) and in the length of their 3'-UTRs. The blue exon represents their last common exon (exon 4). Red indicates the last exon including the 3'-UTR of both transcripts. The dotted grey line in between represents intronic sequences that are spliced out from the pre-mRNA. **(B)** Northern blot of total RNA extracts with DIG-labeled probes directed against *β-actin* (*ACTB*) and both splicing forms of *CDV3*. The expected sizes of the bands and the exposure time of the blot are indicated. The lower panel shows an image of the agarose gel used for blotting demonstrating the integrity of ribosomal RNA in the samples. **(C)** Same as (B) comparing equal amounts of total and nuclear RNA. **(D)** RNA-FISH of *CDV3* RNA in HeLa cells. Specific DIG-labeled probes targeting both splicing forms of *CDV3* were used and visualized with an immunostaining against DIG. Nuclei were counterstained with DAPI. Untreated and RNase A-treated cells are shown. Green: *CDV3* RNA, blue: DAPI, scale bar: 10 µm. **(E)** Agarose gel of strand specific RT-PCRs of the 3'-UTR of *CDV3*.

First, a northern blot of the transcript was performed using total RNA extract from HeLa cells. A combination of three DIG-labeled probes that could bind in the 3'-UTRs of all splicing forms of the *CDV3* transcripts revealed two strong and one weak band, presumably corresponding to the three splicing variants (Figure 18B). The two stronger bands could also be observed in a nuclear RNA extract as

depicted in Figure 18C. Northern blot analysis, revealed no obvious difference in expression of these two RNAs when comparing equal amounts of total RNA with nuclear RNA. Additionally, the endogenous RNA was detected by RNA-FISH with the same DIG-labeled probes used for northern blotting. As illustrated in Figure 18D, the endogenous RNA could be found in the nucleus as well as in the cytoplasm. Finally, it was tested whether there was significant anti-sense transcription happening within CU-RNA. Strand specific RT-PCR revealed that the majority of the transcript is made in the sense direction and only a neglectable amount of PCR product was observed in the anti-sense direction (Figure 18E). From this set of experiments it can be summarized that *CDV3* is expressed in 3 splicing forms, the pool of which can be detected in the nucleus and in the cytoplasm and that the majority of its 3'-UTR is produced in the sense direction with respect to the gene.

3 CU-RNA recruits EZH2 and induces H3K27me3 when tethered to *lac*O arrays

As described in section 2.3 of the results, it was observed that CU-RNA can induce compaction of the chromatin at the *lac*O arrays, a feature that was distinct for CU-RNA and not measured for any of the other RNAs tested. Presumably, CU-RNA functions together with chromatin modifying proteins, which are either recruited to the arrays by the RNA, or which need CU-RNA to properly perform their chromatin modifying function. The following sections focus on changes of the chromatin on the level of posttranslational histone modifications enforced by the presence of CU-RNA. Additionally, it was investigated which protein interaction partner is recruited to the array by CU-RNA to perform the RNA-mediated chromatin changes.

3.1 CU-RNA induces H3K27me3 at the *lac*O arrays

Chromatin compaction is generally associated with heterochromatin (Fraser & Bickmore, 2007). Because CU-RNA induced chromatin compaction when

recruited to the *lac*O arrays, the next step was to investigate the presence of the typical heterochromatic marks, H3K9me3 and H3K27me3. To do so, AO3 cells were transfected with the MS2-GFP-LacI fusion protein and the MS2 stem loop-tagged CU-RNA. Cells transfected with the MS2 loops alone and the *SSR3*-3'-UTR, which does not induce chromatin compaction, were included as controls. The RNA enrichment at the array was detected by RNA-FISH directed against the MS2 stem loops. H3K9me3 and H3K27me3 were visualized by immunofluorescent stainings. H3K9me3 was not found at the arrays (Figure 19).

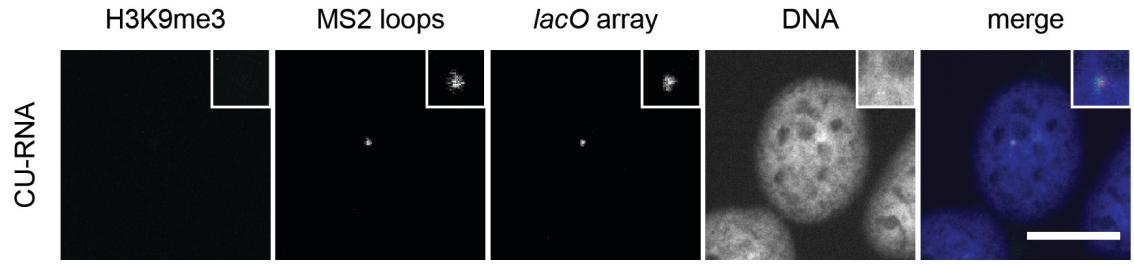


Figure 19: CU-RNA did not promote H3K9me3 when immobilized at the *lac*O array. AO3 cells containing a stable insertion of the bacterial *lac*O repeat sequence were transfected with the MS2-GFP-LacI fusion protein and the MS2 stem loop-tagged CU-RNA. The MS2 loops were detected by RNA-FISH and H3K9me3 by immunostaining. Nuclei were counterstained with DAPI. A representative CLSM image is shown. Insets show enlargements of the *lac*O arrays. Grey: H3K9me3, red: RNA-FISH, green: MS2-LacI-GFP, blue: DAPI, scale bar: 10 µm.

However, as depicted in Figure 20A, CU-RNA led to an enrichment of H3K27me3 at the array. To quantify this finding, the fold enrichment of H3K27me3 at the arrays was determined. This was achieved by comparing the fluorescent signal intensity of the H3K27me3 immunofluorescence at the array to the surrounding background levels in the nucleus. Additionally, the RepA RNA, a 420 nucleotides long transcript that originates from the first exon of the XIST RNA, was included into the analysis as a positive control. It is proposed to be important for H3K27me3 deposition in X inactivation (Zhao et al, 2008). Interestingly, the H3K27me3-promoting activity was not observed to the same extend for all RNAs tested. As depicted in Figure 20B RepA RNA induced H3K27me3 as expected, albeit not as efficiently as CU-RNA. The 3'-UTR of the *SSR3* gene also showed H3K27me3 enrichment but the enrichment was lower in comparison to CU-RNA.

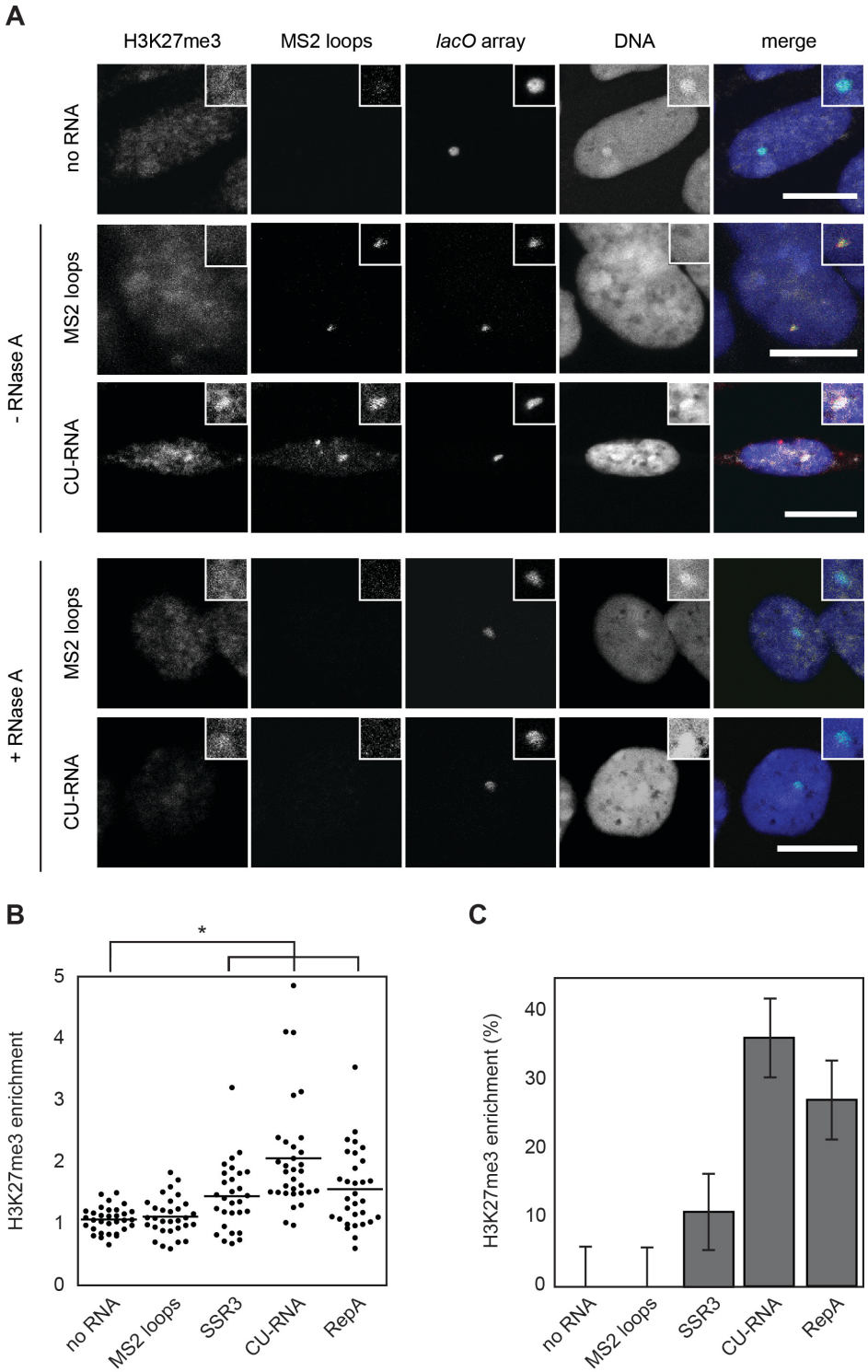


Figure 20: CU-RNA promoted H3K27me3 when immobilized at the *lac*O array. AO3 cells containing a stable insertion of the bacterial *lac*O repeat sequence were transfected with the MS2-GFP-LacI fusion protein and the RNA of interest tagged with the MS2 stem loop sequences. The MS2 loops were detected by RNA-FISH and H3K27me3 by immunostaining. Nuclei were counterstained with DAPI. **(A)** CLSM images of cells transfected with no RNA, the MS2 stem loops only or CU-RNA tagged with the MS2 stem loops. Each RNA was transfected twice and one sample of each was treated with RNase A (lower panel). Insets show enlargements of the *lac*O arrays. Grey: H3K27me3, red: RNA-FISH, green: MS2-LacI-GFP, blue: DAPI, scale bar: 10 µm. **(B)** Quantitative analysis of the enrichment of H3K27me3 at the array as compared to the background, based on the fluorescent intensity of the H3K27me3 signal. n>30, *p<0.05 (Wilcoxon test). **(C)** Percentage of cells that show a two-fold enrichment of H3K27me3.

The MS2 stem loops alone were not able to induce any H3K27me3 enrichment in comparison to the basal level of H3K27me3 at the arrays. These differences became even more obvious when comparing the percentage of cells that had a two-fold enrichment in H3K27me3 (Figure 20C). Thus, of the RNAs tested, CU-RNA recruitment to the *lac*O array promoted the highest enrichment of H3K27me3, but also RepA could efficiently induce H3K27me3. Therefore, it is concluded that the H3K27me3-promoting activity represents a characteristic feature of CU-RNA that is not shared by another 3'-UTR, as for example *SSR3*-3'-UTR. It points towards a difference in the RNAs tested that potentially could explain the differences observed with respect to chromatin compaction.

3.2 RNAs immobilized at the *lac*O array recruit EZH2 with low specificity

H3K27me3 is set by EZH2, which is part of the PRC2 complex. Having observed H3K27me3-induction at the arrays after recruitment of CU-RNA, the next step was to investigate the presence of EZH2. To do so, the same experimental setup and controls as described above were used with an immunostaining directed against EZH2. As illustrated in Figure 21A, CU-RNA induced the recruitment of EZH2 to the *lac*O array when immobilized there. Figure 21A and Figure 21B show that not only CU-RNA but all other RNAs, including the MS2 stem loops alone, recruited EZH2 to the arrays. All RNAs that were analyzed here led to a significant enrichment of EZH2 at the arrays in comparison to the basal level of EZH2 found at the arrays when no RNA was tethered there. Setting a threshold of 1.4 and calculating the percentage of cells that show an enrichment of EZH2 at the arrays above this threshold further supported this observation (Figure 21C). From the set of transcripts investigated here, it is thus concluded that the interaction of EZH2 with the RNAs tested seemed to be rather unspecific. The mere recruitment of EZH2 to the array could consequently not explain why CU-RNA also induced H3K27me3 while other RNAs did not. Instead there must be additional factors that determined when H3K27me3 was set.

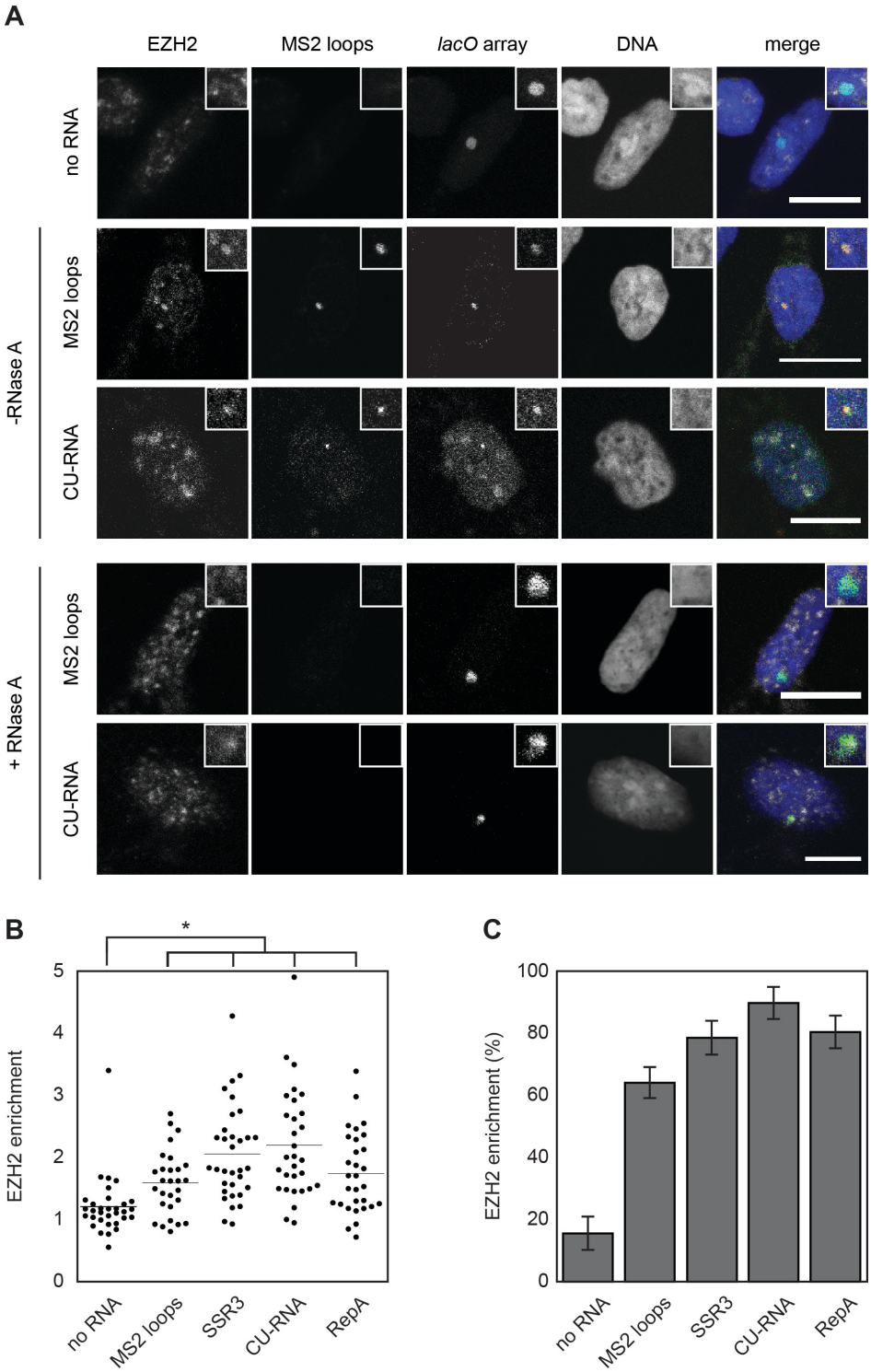


Figure 21: CU-RNA recruited EZH2 when immobilized at the *lac*O array. AO3 cells containing a stable insertion of the bacterial *lac*O repeat sequence were transfected with the MS2-GFP-LacI fusion protein and the RNA of interest tagged with the MS2 stem loop sequences. The MS2 loops were detected by RNA-FISH and EZH2 by immunostaining. Nuclei were counterstained with DAPI. **(A)** CLSM images of cells transfected with no RNA, the MS2 stem loops only or CU-RNA tagged with the MS2 stem loops. Each RNA was transfected twice and one sample of each was treated with RNase A (lower panel). Insets show enlargements of the *lac*O arrays. Grey: EZH2, red: RNA-FISH, green: MS2-LacI-GFP, blue: DAPI, scale bar: 10 µm. **(B)** Quantitative analysis of the enrichment of EZH2 at the array as compared to the background, based on the fluorescent intensity of the EZH2 signal. n>30, *p<0.05 (Wilcoxon test). **(C)** Percentage of cells that show an enrichment of EZH2 above a threshold of 1.4.

3.3 CU-RNA contains a 250 nucleotide long functional element, T0, that is needed for H3K27me3 deposition

Of all RNAs tested the 3097 nucleotides long CU-RNA was the transcript that most significantly led to an enrichment of the H3K27me3 mark at the arrays. The next goal was to narrow down its H3K27me3-promoting activity to a shorter functional sequence. The full length CU-RNA was thus systematically fragmented into smaller transcripts that were cloned together with the MS2 stem loop tag. Their EZH2-recruiting and H3K27me3-inducing capacities were then tested with the MS2/LacI-mediated recruitment system as described above. Each sequence element that still carried the EZH2-recruiting and H3K27me3-promoting capacity was in the next step cut down to even smaller fragments. Figure 22 depicts the fragments of CU-RNA that were tested for their activity in this deletion study.

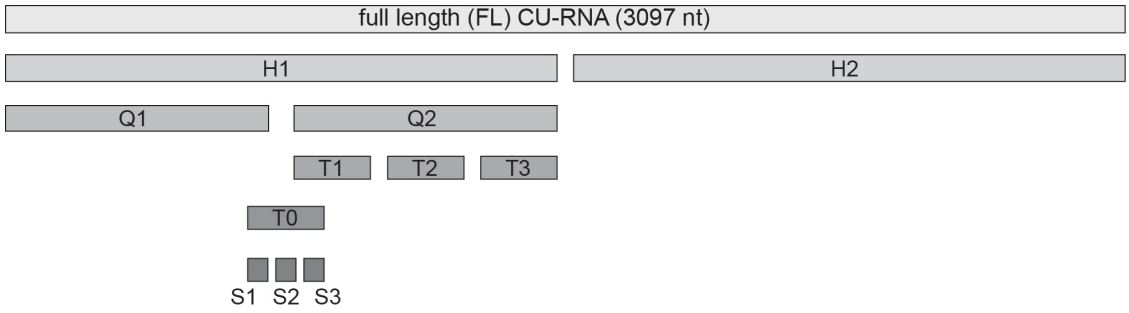


Figure 22: Fragments of CU-RNA that were used in the deletion study. Each fragment was cloned together with the MS2 stem loops sequence and tested for EZH2 recruitment and H3K27me3 induction in the MS2/LacI-mediated recruitment system.

Figure 23 depicts the results of the recruitment assays using H3K27me3 enrichment as the readout. A 250 nucleotides long transcript was identified that induced H3K27me3 as efficiently as the full length CU-RNA. In the following this sequence element will be referred to as “CU-RNA-T0”. All other fragments derived from CU-RNA were less efficient than the full length CU-RNA itself. Tethering of smaller fragments derived from CU-RNA-T0 did not lead to H3K27me3 enrichment at the array, suggesting that CU-RNA-T0 is the minimal element that is necessary for CU-RNA to induce H3K27me3.

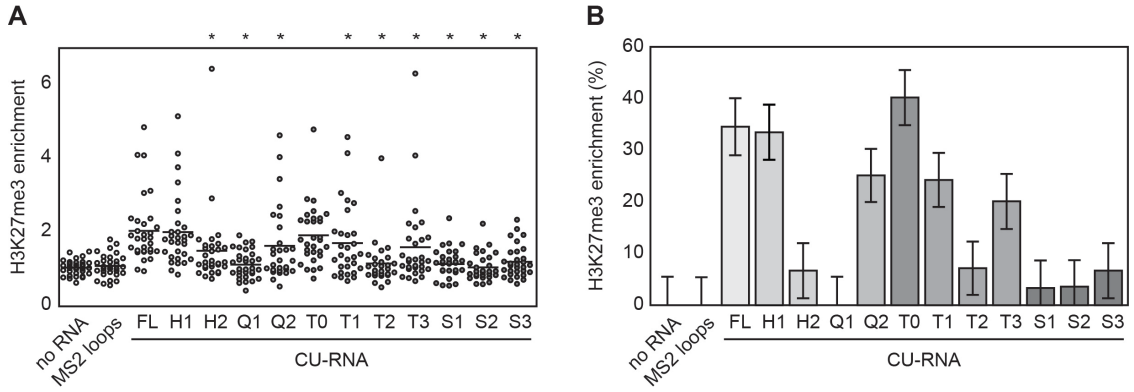


Figure 23: The CU-RNA-T0 fragment recruits H3K27me3 to the *lac*O array as efficiently as full length CU-RNA. AO3 cells containing a stable insertion of the bacterial *lac*O repeat sequence were transfected with the MS2-GFP-LacI fusion protein and the RNA of interest tagged with the MS2 stem loop sequences. The MS2 loops were detected by RNA-FISH and H3K27me3 by immunostaining. **(A)** Quantitative analysis of the enrichment of H3K27me3 at the array as compared to the background, based on the fluorescent intensity of the H3K27me3 signal. The propensity of each fragment to recruit H3K27me3 was compared to that of the full length CU-RNA. n>30, *p<0.05 (Wilcoxon test). **(B)** Percentage of cells that show a two-fold enrichment of H3K27me3.

The experiment was repeated testing for EZH2 enrichment at the array depending on the CU-RNA fragment that is tethered. It can be seen that all RNAs had the ability to recruit EZH2 to the array (Figure 24). Although some had a lower efficiency than the full length CU-RNA or even than the MS2 stem loops alone they still led to an enrichment of EZH2 that is twice as high as the basal level of EZH2 at the array without tethering any RNA.

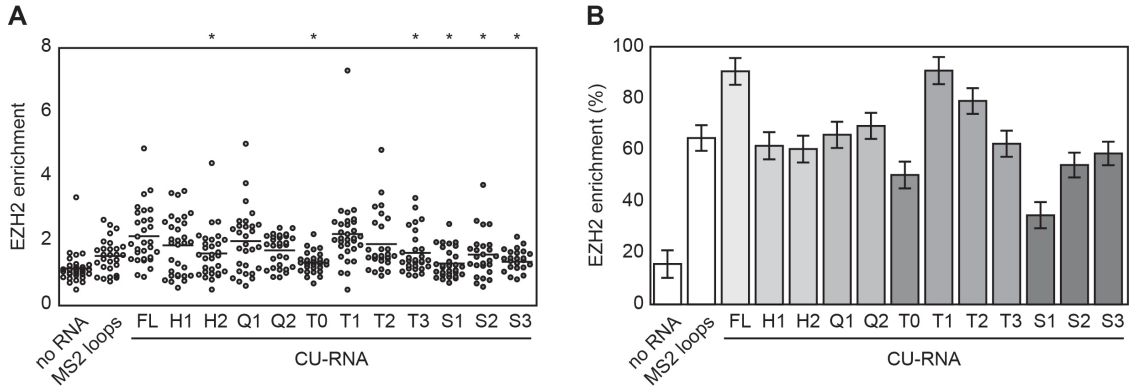


Figure 24: A deletion study of CU-RNA revealed that all fragments derived from CU-RNA recruit EZH2 to the *lac*O array. AO3 cells containing a stable insertion of the bacterial *lac*O repeat sequence were transfected with the MS2-GFP-LacI fusion protein and the RNA of interest tagged with the MS2 stem loop sequences. The MS2 loops were detected by RNA-FISH and EZH2 by immunostaining. **(A)** Quantitative analysis of the enrichment of EZH2 at the array as compared to the background, based on the fluorescent intensity of the EZH2 signal. The propensity of each fragment to recruit EZH2 was compared to that of the full length CU-RNA. n>30, *p<0.05 (Wilcoxon test). **(B)** Percentage of cells that show an enrichment of EZH2 above a threshold of 1.4.

3.4 CU-RNA-ΔT0, a transcript variant lacking the full length T0 element, cannot induce H3K27me3 deposition

In the next step, CU-RNA-T0 was investigated with respect to its position within the *CDV3a* transcript and the alternative splicing form *CDV3b*. It is noted that T0 is not fully represented in the splice form *CDV3b*, because the sequence overlaps with the alternative splice site of exon 5b (Figure 25). To investigate the role of CU-RNA-T0 as part of the *CDV3a* transcript in relation to the *CDV3b* transcript, a second RNA fragment originating from the *CDV3b* was designed. In this fragment the first 80 nucleotides of the 250 bases long CU-RNA-T0 were replace by the corresponding sequence from *CDV3b*. This fragment was termed CU-RNA-ΔT0, as depicted in Figure 25.

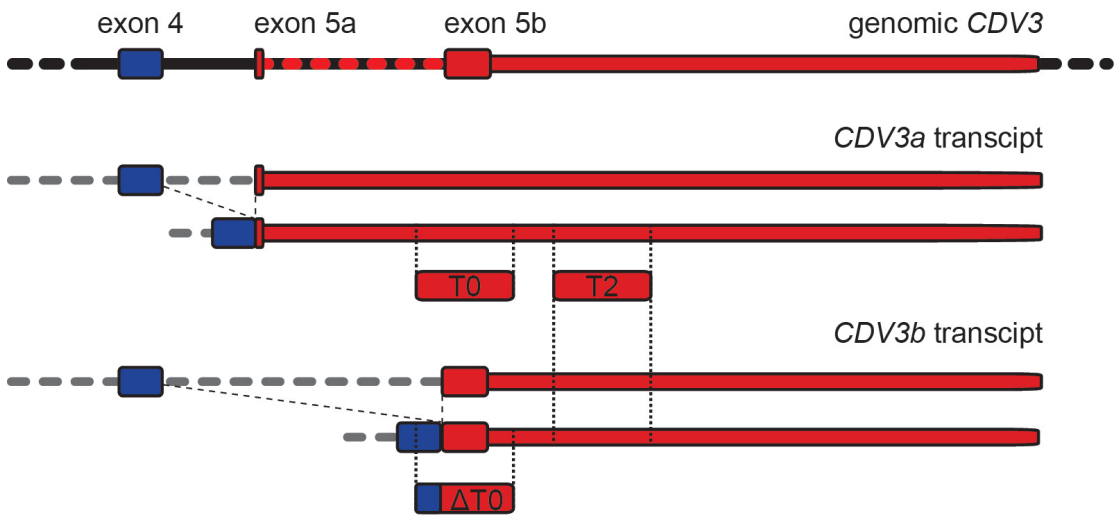


Figure 25: Schematic illustration of the position of CU-RNA-T0 and CU-RNA-ΔT0 with respect to *CDV3a* and *CDV3b*. Genomic *CDV3* (top panel) and the two transcript variants in their unspliced and spliced forms (bottom panels) indicating the positions of CU-RNA-T0 (T0), CU-RNA-T2 (T2) and CU-RNA-ΔT0 (ΔT0) are depicted (not drawn to scale). CU-RNA-ΔT0 shares the 3' part with CU-RNA-T0, the part of the sequence that is present in both the *CDV3a* and *CDV3b* transcripts. The first 80 nucleotides of the CU-RNA-T0 are not present in the *CDV3b* transcript. They are replaced with the corresponding sequence from the *CDV3b* transcript to form the CU-RNA-ΔT0 fragment.

In the following, CU-RNA-ΔT0 was investigated with respect to its H3K27me3-promoting activity and its EZH2-recruiting ability. To this end, CU-RNA-ΔT0 was recruited to and tethered at the stably integrated *lac*O arrays in AO3 cells. The enrichment of H3K27me3 and EZH2 was investigated as described above

(sections 3.1 and 3.2 of the results). Figure 26A and Figure 26B suggest that the efficiency of CU-RNA-ΔT0 to promote H3K27me3 at the array was significantly lower compared to CU-RNA-T0. From Figure 26C and Figure 26D it can be concluded that CU-RNA-ΔT0 recruited EZH2 to the arrays at least as efficiently as CU-RNA-T0. Therefore, the H3K27me3-promoting capacity of CU-RNA-T0 is indeed highly specific and can be abolished by exchanging the first 80 nucleotides as shown by recruiting CU-RNA-ΔT0.

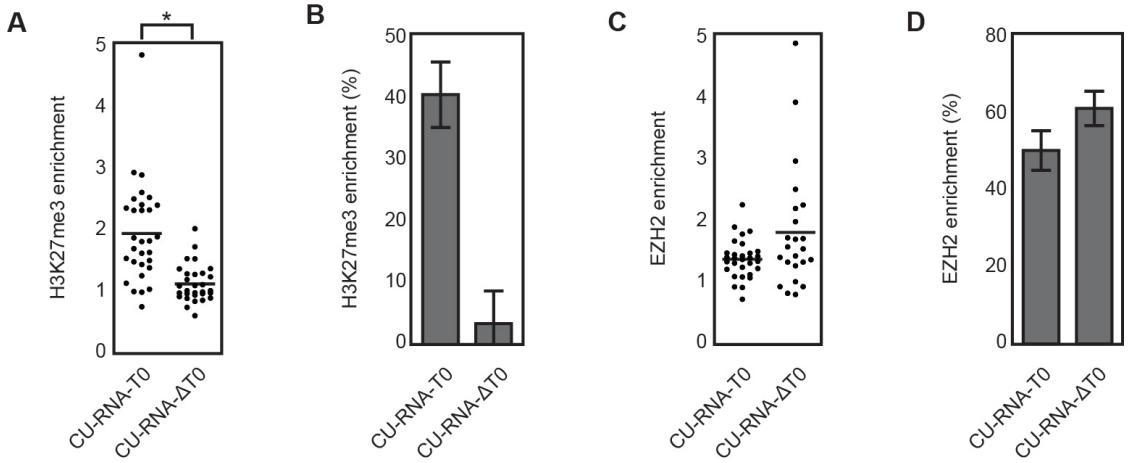


Figure 26: CU-RNA-ΔT0 recruits EZH2 but fails to induce H3K27me3. AO3 cells containing a stable insertion of the bacterial *lac*O repeat sequence were transfected with the MS2-GFP-LacI fusion protein and the RNA of interest tagged with the MS2 stem loop sequences. The MS2 loops were detected by RNA-FISH and H3K27me3 (A and B) and EZH2 (C and D) by immunostaining. **(A)** Quantitative analysis of the enrichment of H3K27me3 at the array as compared to the background, based on the fluorescent intensity of the H3K27me3 signal. n>30, *p<0.05 (Wilcoxon test) **(B)** Percentage of cells that show a two-fold enrichment of H3K27me3. **(C)** Same as in (A) with an immunostaining against EZH2**. (D)** Same as in (B) with an immunostaining against EZH2. A threshold of 1.4 was used**.**

3.5 RNA transcripts have a non-specific inhibitory effect on EZH2 activity *in vitro*

It was observed that EZH2 was catalytically active and methylated H3K27 in the MS2/LacI recruitment assay, only when in complex with certain RNAs such as CU-RNA-T0. Therefore, an *in vitro* methylation assay was used to measure the activity of EZH2 in the presence of different RNA transcripts. The assay comprises of the purified enzyme in complex with its physiological interaction partners EED, SUZ12, RBAP48 and AEBP2 (Margueron & Reinberg, 2011).

Immobilized histone H3 peptide substrate was incubated with the 5-component enzyme complex and S-adenosyl methionine (SAM) as a methyl donor. The amount of H3K27me3 as detected by a horseradish peroxidase (HRP)-coupled antibody was used as an indication for EZH2 activity (Figure 27A).

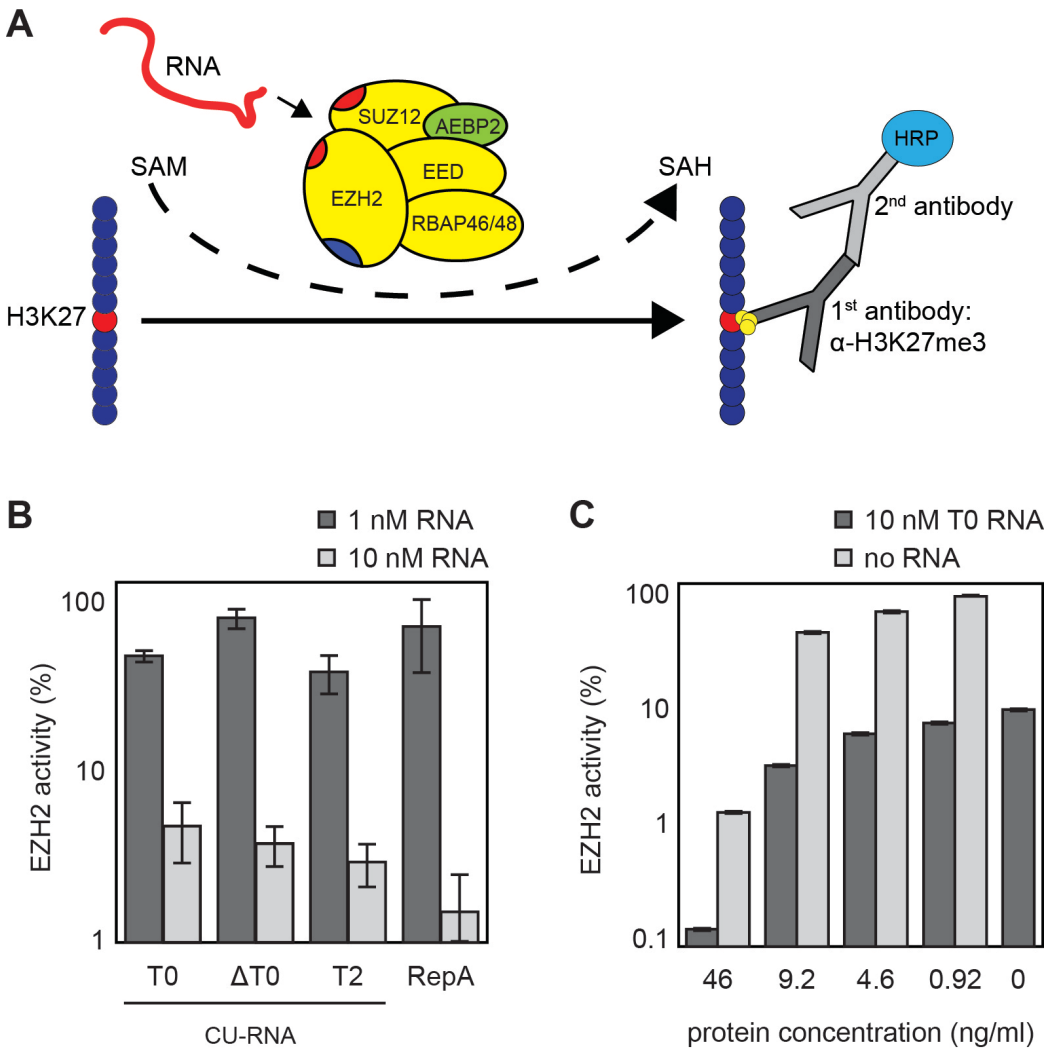


Figure 27: CU-RNA-T0 has an inhibitory effect on EZH2 activity in an *in vitro* methylation assay. (A) Principle of the EZH2 *in vitro* methylation assay. EZH2 in complex with EED, SUZ12, RBAP48 and AEBP2, the methyl donor S-adenosylmethionine (SAM) and the RNAs of interest were added to immobilized H3 peptide substrate. The amount of H3K27me3 was detected with a specific antibody directed against H3K27me3 and a secondary antibody coupled to HRP. The amount of emitted chemiluminescence was used as a measure for the amount of H3K27me3 in the sample. **(B)** 10 nM of the EZH2 enzyme complex were incubated with the indicated amounts of CU-RNA-T0, CU-RNA-ΔT0, CU-RNA-T2 and RepA RNA. The activity of EZH2 is plotted as the percentage of the total EZH2 activity achieved when no RNA is added to the reaction. **(C)** The indicated amounts of nuclear lysate extracted from AO3 cells were incubated with 10 nM of CU-RNA-T0 in the EZH2 *in vitro* methylation assay. The activity of EZH2 is plotted as the percentage of the total EZH2 activity achieved when no lysate and no RNA were added to the reaction. SAH: S-adenosyl-homocysteine.

CU-RNA-T0 was added to the reaction in approximately equimolar concentration to the EHZ2 complex. Additionally, a ten times lower concentration of CU-RNA-T0 was tested. As negative controls, CU-RNA-ΔT0 and CU-RNA-T2 fragments were included as transcripts of similar length that did not harbor the H3K27me3-promoting activity. The earlier mentioned RepA transcript was chosen as the positive control. This RNA has been described to inhibit EZH2 catalytic activity in an *in vitro* methylation assay using EZH2 in a 4-component complex with EED, SUZ12 and RBAP48 (Cifuentes-Rojas et al, 2014). Figure 27B shows that when EZH2 and CU-RNA-T0 were present at approximately equimolar concentrations (10 nM), EZH2 activity was abolished to less than 10 %. This was the case with all three RNAs tested. Adding 1 nM of any RNA to the reaction, thus only a tenth of the amount of EZH2, left EZH2 almost fully active. Unlike the initial hypothesis that CU-RNA-T0 could have a positive effect on the activity of EZH2, it is concluded that in the *in vitro* assay CU-RNA-T0 negatively influences EZH2 H3K27 methylation ability. This setting also confirmed previous results that RepA inhibited EZH2 activity.

Based on these data, it was proposed that there must be other factors to explain the differential H3K27me3-promoting capacity of CU-RNA-T0 that was observed in the experiments using the MS2/LacI model system. Therefore, it was then tested whether a nuclear lysate extracted from AO3 cells could rescue the inhibitory effect of CU-RNA-T0 on EZH2 in the *in vitro* methylation assay. Hypothetically, the factors, that together with CU-RNA-T0 enhance the H3K27me3-activity of EZH2, should be present in this lysate and should be able to also perform this function *in vitro*. The lysate was added to the methylation reaction in increasing concentrations as depicted in Figure 27C. For comparison, the lysate was loaded twice, once in combination with CU-RNA-T0 and once without (Figure 27C). The nuclear lysate alone had only limited effect on the activity of EZH2 up to a protein concentration of 9.2 ng/ml. However, at a concentration of 46 ng/ml the lysate alone almost fully inhibited EZH2 catalytic activity. Adding 10 nM CU-RNA-T0 to the reaction in combination with the lower lysate concentrations inhibited EZH2 activity to the same extend as observed

without the nuclear extract (Figure 27B). At the highest lysate concentration, that itself already had an inhibitory effect, CU-RNA-T0 reduced EZH2 activity even further to less than 1%. The AO3 nuclear lysate thus did not rescue the EZH2 inhibitory effect of CU-RNA-T0. The factor that can rescue EZH2 activity in the living cells at the *lac*O arrays was thus either not present in the lysate or only inactive in the chromatin-independent context. Thus, this factor remains to be determined.

3.6 The inhibitory effect of RNAs on EZH2 activity might be rescued by JARID2

It is proposed that the PCR2 accessory factor JARID2 relieves EZH2 repression by RNA, boosting its methylation rate in an RNA and dose dependent manner (Cifuentes-Rojas et al, 2014). Here, it is therefore tested whether JARID2 could be the factor that differentiates CU-RNA-T0 from other RNAs that bind EZH2 but do not promote H3K27me3. To this end, the following experiments were conducted: First, RNA affinity purification using immobilized biotinylated CU-RNA-T0 in combination with western blotting was used to test for the specific interaction of CU-RNA-T0 with JARID2. Second, the MS2/LacI-mediated recruitment system was used to investigate whether JARID2 is also recruited to the arrays by CU-RNA-T0 in the living cell.

JARID2 interacts with CU-RNA-T0

To investigate the differential interaction of CU-RNA-T0 and CU-RNA-ΔT0 with JARID2, RNA affinity purification was performed: The biotinylated RNAs were immobilized on streptavidin coated magnetic beads. These were incubated with nuclear protein lysate extracted from HeLa S3 cells. Specifically bound proteins were eluted with RNase A and processed for western blotting (Figure 28A). As controls, the proteins bound to the beads alone, a control RNA of similar size (*alu*RNA_R), CU-RNA-T2 and RepA were included. Moreover, as a proof of principle, EZH2 binding was monitored by western blotting, as it was previously

shown to bind all RNAs tested with similar affinity. As depicted in Figure 28B, both CU-RNA-T0, CU-RNA-ΔT0, CU-RNA-T2 and RepA pulled down EZH2 while the beads alone and the control RNA showed no EZH2 enrichment. Despite the low contrast, the western blot against JARID2 showed that CU-RNA-T0, CU-RNA-ΔT0, CU-RNA-T2 and RepA bound JARID2 (Figure 28C) while the beads alone and the control RNA did not. The blot also displayed that CU-RNA-T0 and RepA bound slightly more JARID2 than CU-RNA-ΔT0 and CU-RNA-T2. This suggests that the affinity of CU-RNA-T0 and RepA for JARID2 is higher than that of the other RNAs tested. Therefore, JARID2 could be the factor distinguishing CU-RNA-T0 from CU-RNA-ΔT0 and CU-RNA-T2.

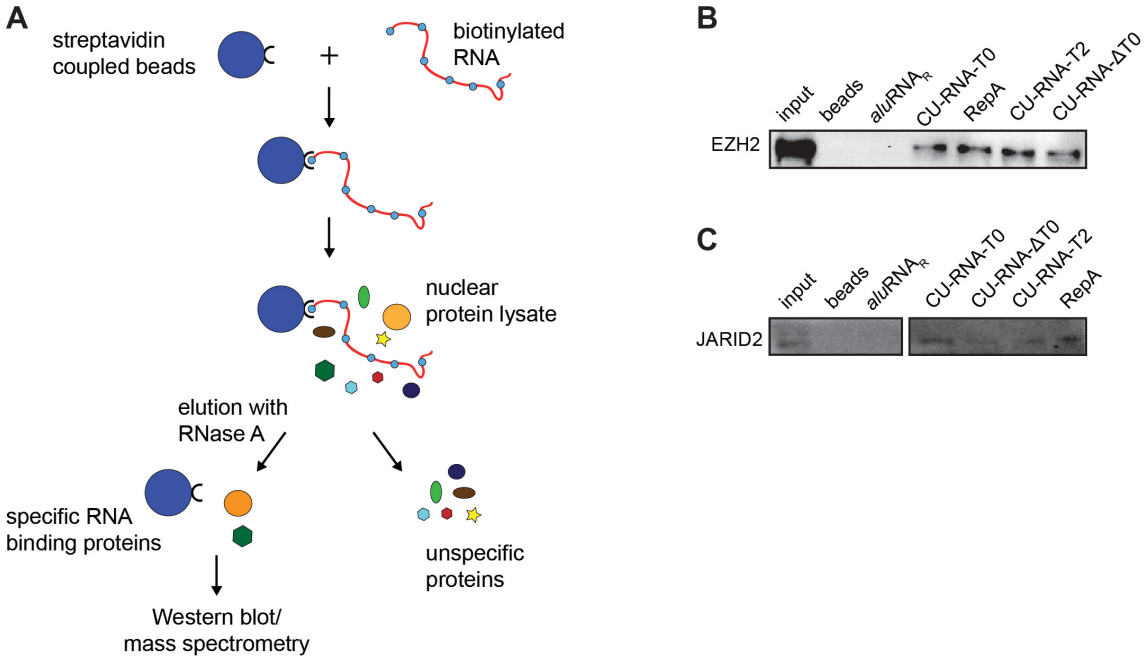


Figure 28: CU-RNA-T0, CU-RNA-ΔT0, CU-RNA-T2 and RepA pulled down EZH2 and JARID2 in an *in vitro* binding assay. (A) Schematic representation of the RNA affinity purification workflow. The RNAs of interest were *in vitro* transcribed with biotinylated UTP and immobilized on streptavidin-coated magnetic beads. Nuclear protein lysate was incubated with the RNAs, unbound proteins were washed off and proteins that specifically bound to the RNAs were eluted with RNase A. The identity of these proteins could be analyzed by western blot or mass spectrometry. **(B)** Western blot of EZH2 showing its enrichment through CU-RNA-T0, CU-RNA-ΔT0, CU-RNA-T2 and RepA using RNA affinity purification. **(C)** Same as (B) with an antibody against JARID2 (Abcam antibody). The blot was cut to show only the RNAs of interest here. All fragments shown were treated equally and not processed differently.

JARID2 recruitment to the lacO arrays by CU-RNA fragments

To investigate whether JARID2 is also recruited to the arrays by CU-RNA-T0 in the living cell, the MS2/LacI-mediated recruitment system was used. AO3 cells were transfected with the MS2-GFP-LacI fusion protein and the MS2 stem loop-tagged CU-RNA-T0 and CU-RNA-ΔT0. CU-RNA-T2 and the MS2 stem loops were used as negative controls, while RepA served as a positive control. RNA at the *lac*O arrays was detected by RNA-FISH. Two different antibodies were used to investigate differential enrichment of JARID2 at the arrays. The enrichment of JARID2 was calculated as the ratio of the intensity of the fluorescent signal at the array against the background levels in the nucleus. Figure 29A and Figure 29C depict the enrichment analysis of this experiment. From the analysis using the Abcam antibody (Figure 29A) it could be concluded that all RNAs tested including the MS2 stem loops slightly enriched JARID2 at the arrays. The variations in enrichment were subtle. Furthermore, tethering the positive control, RepA RNA, which was shown to interact with JARID2 (Cifuentes-Rojas et al, 2014), did not increase JARID2 enrichment as compared to its basal level at the array. Another antibody was tested in the same experimental setup and lead to inconsistent results with respect to the RNA dependent enrichment of JARID2 at the arrays (Figure 29C). Here, the MS2 stem loops, CU-RNA-T0, CU-RNA-T2 and RepA showed a slight enrichment of JARID2. However, tethering full-length CU-RNA did not change JARID2 levels.

Having observed the inconsistency of the data retrieved from the recruitment assay above, the quality of both antibodies was assessed in a western blot analysis with nuclear protein extract of HeLa cells. The Abcam antibody (Figure 29B) showed a single band that could potentially represent the 138 kDa large JARID2, although the band ran slightly higher than expected. The Novusbio antibody (Figure 29D) however, gave rise to a multitude of bands of various sizes. Although the expected band at approximately 138 kDa was also faintly visible, this demonstrated low specificity of the antibody and made it useless for immunofluorescent analyses. The Novusbio antibody was therefore not used in subsequent experiments. The Abcam antibody was apparently showing enough

specificity, but there is no guaranty that it is suited for immunofluorescence. From this experimental setup JARID2 could not be proven to be the factor that differentiated CU-RNA-T0 from other RNAs that bind EZH2.

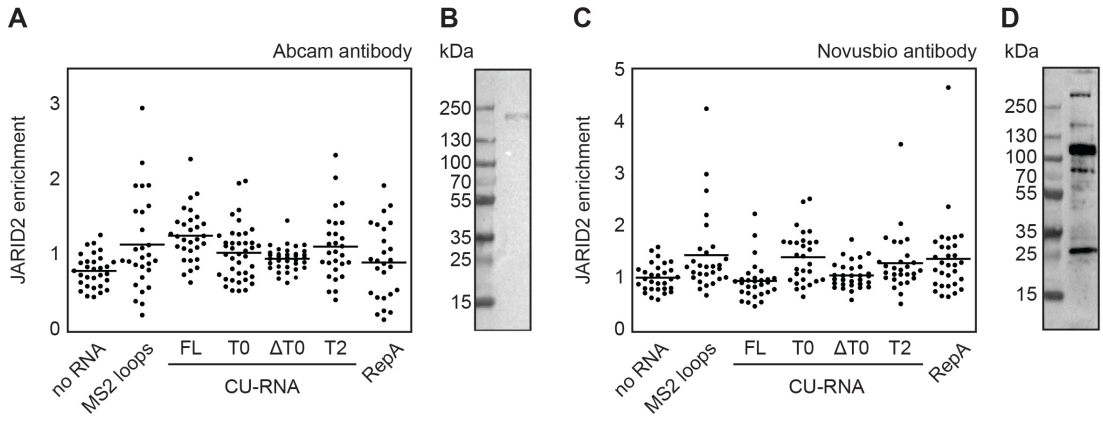


Figure 29: CU-RNA-T0 dependent JARID2 enrichment at the *lac*O arrays. (A) and **(C)** AO3 cells containing a stable insertion of the bacterial *lac*O repeat sequence were transfected with the MS2-GFP-LacI fusion protein and the RNA of interest tagged with the MS2 stem loop sequences. The MS2 loops were detected by RNA-FISH and JARID2 by immunostaining using an Abcam antibody (A) or a Novusbio antibody (C). A quantitative analysis of the enrichment of JARID2 at the array as compared to the background, based on fluorescent intensity of the JARID2 signal is shown. **(B)** and **(D)** Western blot analysis of the two antibodies used to detect enrichment of JARID2 at the *lac*O arrays using nuclear protein lysate of HeLa cells.

4 Splicing of the *CDV3* transcript variants is dependent on H3K27me3

It has been shown that the H3K27me3-promoting capacity of CU-RNA-T0 is highly specific to its 250 nucleotide sequence. Exchanging the first 80 nucleotides of the 5' sequence, giving rise to CU-RNA-ΔT0, led to almost complete depletion of the H3K27me3-promoting activity in the MS2/LacI-mediated recruitment system. Given that these two variants of CU-RNA originate from the endogenous, different splice variants of the *CDV3* transcript, I hypothesized that the H3K27me3-promoting ability of CU-RNA-T0 is intrinsically linked to the regulation of alternative splicing resulting in *CDV3a* or *CDV3b*. In support of this possibility, it has been postulated by other groups that indeed the local level of H3K27me3 can influence alternative splicing (Gonzalez et al, 2015; Luco et al, 2010). To test whether H3K27me3 levels may influence *CDV3* splicing, EZH2 was inhibited by

the small molecule inhibitor GSK343 and the levels of the *CDV3* transcripts were measured by qRT-PCR.

4.1 H3K27me3 levels at the *CDV3* locus determine the ratio of the *CDV3a* and *CDV3b* splice variants

To be able to measure the expression levels of *CDV3a* and *CDV3b*, with regards to the H3K27me3 levels, two conditions had to be satisfied: First, it was required to show that the inhibition of EZH2 by the small molecule inhibitor GSK343 effectively inhibits EZH2 and leads to reduction in H3K27me3 globally and at the *CDV3* locus. Second, *CDV3a* and *CDV3b* specific primer pairs had to be designed, which are able to differentially detect the expression levels of the two splice variants.

GSK343 globally reduces H3K27me3 levels

From what is known to date, EZH2 is the only methyltransferase that trimethylates H3K27. Therefore, a knockdown or inhibition of EZH2 should be sufficient to globally alter the levels of H3K27me3. The specific small molecule inhibitor GSK343 has been shown to be a highly potent, selective and cell-active inhibitor of EZH2 (Verma et al, 2012). To test the inhibitor, HeLa cells were treated with two concentrations of GSK343, 1 µM and 5 µM, or the corresponding volume of the solvent, DMSO, as a control, for three days. The global levels of H3K27me3 were measured by FACS analysis and western blotting. Figure 30A depicts the levels of H3K27me3 in DMSO-treated HeLa cells in comparison to HeLa cells that have been treated with the EZH2 inhibitor for three days as determined by FACS analysis. It was observed that with both concentrations, the level of H3K27me3 was reduced to the intensity of the background signal, which was determined by measuring the signal intensity of cells that were not incubated with the H3K27me3 primary antibody. The same result was obtained by measuring the amount of H3K27me3 in DMSO- and GSK343-treated (5 µM) HeLa cells by western blotting of histone extractions (Figure 30B). A reduction to

only 13% of H3K27me3 can be observed suggesting that GSK343 is a potent inhibitor of the enzymatic activity of EZH2.

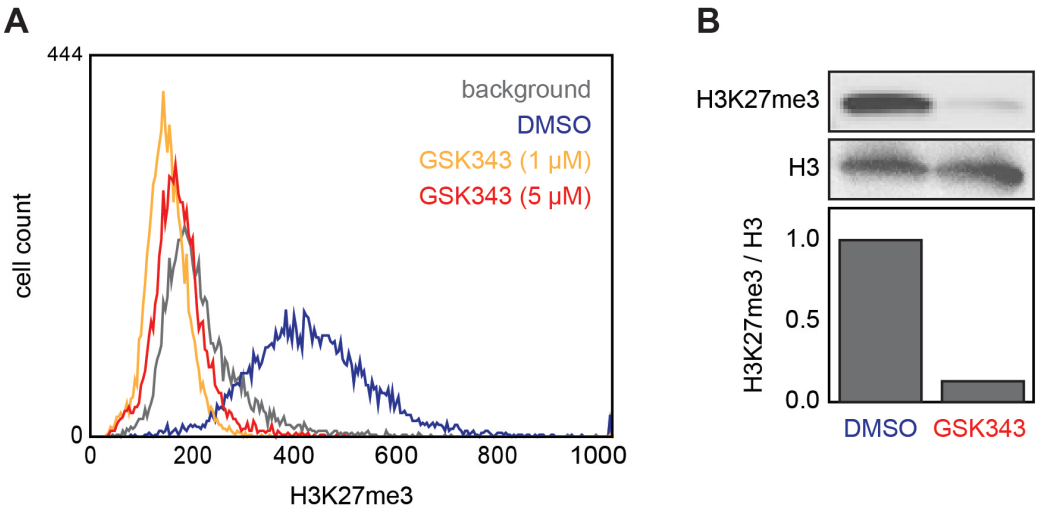


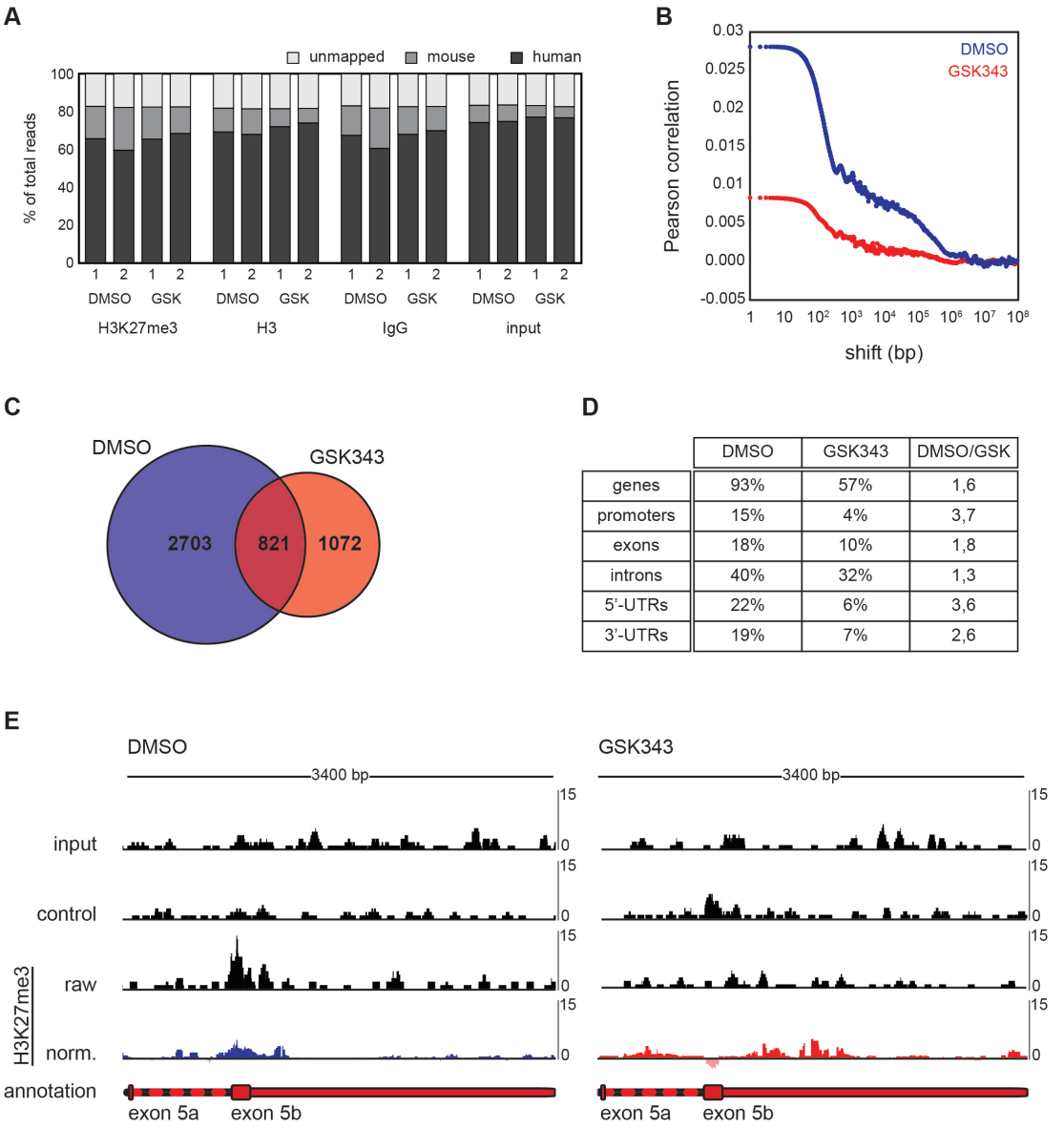
Figure 30: Treatment with the EZH2-small molecule inhibitor GSK343 globally reduced the H3K27me3 levels. (A) HeLa cells were treated with 1 µM and 5 µM GSK343 or the equivalent volume of DMSO for 72 hours. H3K27me3 was stained with a specific antibody and FACS analysis was applied to measure the H3K27me3 levels of the cells. A sample that was not incubated with the primary antibody against H3K27me3 served as the background control. The histogram shows the intensity of the H3K27me3-staining in each of the samples. **(B)** HeLa cells were treated with 5 µM GSK343 or the corresponding volume of DMSO. The histones were extracted and subjected to western blot analysis using specific antibodies against H3K27me3 and histone H3. The H3K27me3 levels were quantified and are displayed as the ratio of histone H3 in each sample.

GSK343 treatment leads to a genome-wide redistribution of H3K27me3 and reduces H3K27me3 at the CDV3 locus

To examine the influence of EZH2 inhibition by GSK343 on a genome-wide scale, H3K27me3 specific chromatin immunoprecipitation and sequencing (ChIP-seq) of cells that were treated with 5 µM GSK343 for six days was performed in two replicates. The input chromatin, an immunoglobulin G (IgG) control, and a ChIP against histone H3 were included as controls. Furthermore, a spike-in control was included by adding 10% of chromatin from untreated mouse embryonic stem cells to all samples (Bonhoure et al, 2014). The number of total reads yielded in each reaction is shown in Appendix 1. The reads were mapped to the human and mouse genomes using bowtie (Langmead et al, 2009), allowing for two mismatches. Ambiguous reads mapping to both the human and mouse genome

were removed from the analysis. Figure 31A depicts the distribution of reads mapping to human and mouse. 59% to 77% of all reads could be assigned to the human genome, whereas 5% to 22% mapped to the mouse genome. 16% to 18% of all reads could not be mapped to either of the genomes. In the H3 ChIP, IgG control, input, and untreated H3K27me3 samples, the percentage of reads mapping to the mouse genome was expected to be approximately 10% corresponding to the amount of mouse chromatin that was spiked in. With 5% to 22% the variability of the amount of reads mapping to the mouse genome in those samples was thus higher than expected. Figure 30 shows that GSK343 treatment reduced the global level of H3K27me3. It was therefore expected that in the GSK343-treated samples more reads would align with the mouse genome, since in relation less human material should be pulled down in the H3K27me3 ChIP. This trend could, however, not be observed here.

First, to compare the two replicates of the ChIP-seq experiment, all H3K27me3 ChIP data sets were normalized to obtain comparable read occupancy profiles. This was done by subtracting the weighted control IP (IgG) and calculating the enrichment of the signal over the input sample. Next, a correlation function of the two DMSO- and GSK343-treated samples respectively was calculated using multi-scale correlation evaluation (MCORE) (Molitor et al, 2015). To do so, the normalized read occupancy profiles from each H3K27me3-IP were shifted with respected to each other along the genomic coordinates. The Pearson correlation coefficient for each shifting distance along the human chromosome 3, the chromosome *CDV3* is located on, was calculated and plotted in Figure 31B. The positive correlation of the co-localization of signal in the two replicates indicates co-occurrence at the given shift distance. The amplitude of the correlation function is a measure for the abundance of H3K27me3 domains. The correlation amplitude was higher for the two DMSO-treated samples (0.028) than for the GSK343 replicates (0.008) reflecting a higher signal to noise ratio in the DMSO- than in the GSK343-treated cells. It can therefore be concluded that under GSK343 treatment, the localization of H3K27me3 over the genome redistributes, resulting in less well-defined H3K27me3 domains.



	DMSO	GSK343	DMSO/GSK
genes	93%	57%	1,6
promoters	15%	4%	3,7
exons	18%	10%	1,8
introns	40%	32%	1,3
5'-UTRs	22%	6%	3,6
3'-UTRs	19%	7%	2,6

Figure 31: Inhibition of the enzymatic activity of EZH2 by GSK343 redistributes H3K27me3 genome wide and reduces H3K27me3 at the *CDV3* locus. HeLa cells were treated with 5 µM GSK343 for six days and ChIP-seq was performed against H3K27me3. H3, IgG, and input chromatin were included as controls. A spike-in control consisting of 10% chromatin from untreated mouse embryonic stem cells was added to each sample. **(A)** Distribution of reads mapped to the human and mouse genomes. Both replicates (1 and 2) of DMSO- and GSK343-treated samples are shown for the H3K27me3 and H3 ChIPs, IgG control, and input. **(B)** MCORE Pearson correlation functions of the DMSO- and GSK343-treated replicates on the human chromosome 3, where *CDV3* is located (Molitor et al, 2015). **(C)** H3K27me3 peaks called with Sicer in replicate 1 of the DMSO- and GSK343-treated samples. The numbers of differentially identified and overlapping peaks are indicated. **(D)** Percentage of Sicer peaks overlapping with genes, promoters, exons, introns, 5'-UTRs, and 3'-UTRs in replicate 1 of both the DMSO- and GSK343-treated cells. The ratio of the percentages in the control and treated cells is indicated. **(E)** Interactive genome viewer (IGV) trace of the read distribution in CU-RNA. The left panel depicts replicate 1 of the DMSO-treated sample and the right panel replicate 1 of the GSK343-treated sample. Raw read distributions of the input, IgG control, and H3K27me3 samples are shown in black. The normalized H3K27me3 occupancy is depicted in blue for the DMSO-treated cells and in red for the GSK343-treated cells. All traces are scaled equally.

To identify regions that are enriched in H3K27me3, peaks were called using Sicer (Zang et al, 2009). No peaks could be called for the second replicates of both the DMSO- and GSK343-treated samples with the parameters used (window size: 200 and gap size: 600). This indicates a poor enrichment of the H3K27me3-IP over the background signal and low quality of the data. The second replicate was therefore not used for further analyses. For the first replicate, a total of 3524 peaks could be called in the DMSO-treated sample and 1893 in the GSK343-treated sample. 821 of the peaks overlapped whereas the others were unique to either of the samples (Figure 31C). First, the fact that a large proportion of peaks did not overlap indicates a redistribution of H3K27me3 over the genome after GSK343 treatment. This confirms the findings from the MCORE analysis (Figure 31B). Second, the fact that less peaks could be called in the GSK343-treated sample, shows a reduction in the enrichment of H3K27me3 at certain domains, as expected after GSK343 treatment. Subsequently, all peaks were assigned to genomic regions of coding genes. It was found that in the DMSO-treated sample the majority of peaks (93%) overlapped with genes, while this fraction was only 57% in the GSK343-treated sample. The most profound differences in overlaps between the treated and untreated samples were found in promoters and 5'-UTRs (3.6 and 3.7 times more in the DMSO sample). Furthermore, also significantly more peaks were found in 3'-UTRs in DMSO-treated cells (2.6 times more). This analysis implies that many H3K27me3 domains in annotated genes were affected by the GSK343 treatment and reinforces the fact that H3K27me3 is redistributed by inhibition of EZH2.

Lastly, the level of H3K27me3 at the CU-RNA locus was examined. From the raw read occupancy profiles but even more clearly from the MCORE normalized occupancies (Figure 31E, row 4) it could be deduced that there was an enrichment of H3K27me3 at the start of the alternative exon of *CDV3* (exon 5b) in DMSO-treated cells (Figure 31E**,** left panel, blue trace). This enrichment disappeared entirely after GSK343 treatment (Figure 31E**,** right panel, red trace). This indicates a strong reduction of H3K27me3 by GSK343 treatment at the CU-RNA locus. This occurs specifically at the site where *CDV3* can be

alternatively spliced. At the same time, more H3K27me3 was found in the upstream intron and the 3'-UTR after GSK343 treatment suggesting that H3K27me3 is very specifically altered at the alternatively spliced exon and its surrounding. In conclusion, GSK343 is thus a potent inhibitor of EZH2 leading to a global reduction and redistribution of H3K27me3 over the genome, reduction of H3K27me3 in annotated genes, and a local reduction of H3K27me3 at the CU-RNA locus, at the start of the alternative exon of *CDV3*.

Transcript levels of alternative CU-RNA variants can be quantified by real-time PCR

In order to measure whether the local level of H3K27me3 at the *CDV3* locus has an effect on the ratio between the two *CDV3* transcript variants, the first step was to set up an assay to stably detect each of the variants. Quantitative reverse-transcriptase PCR (qRT-PCR) was ran with primers pairs that specifically detected only one of the two splicing forms (Figure 32).

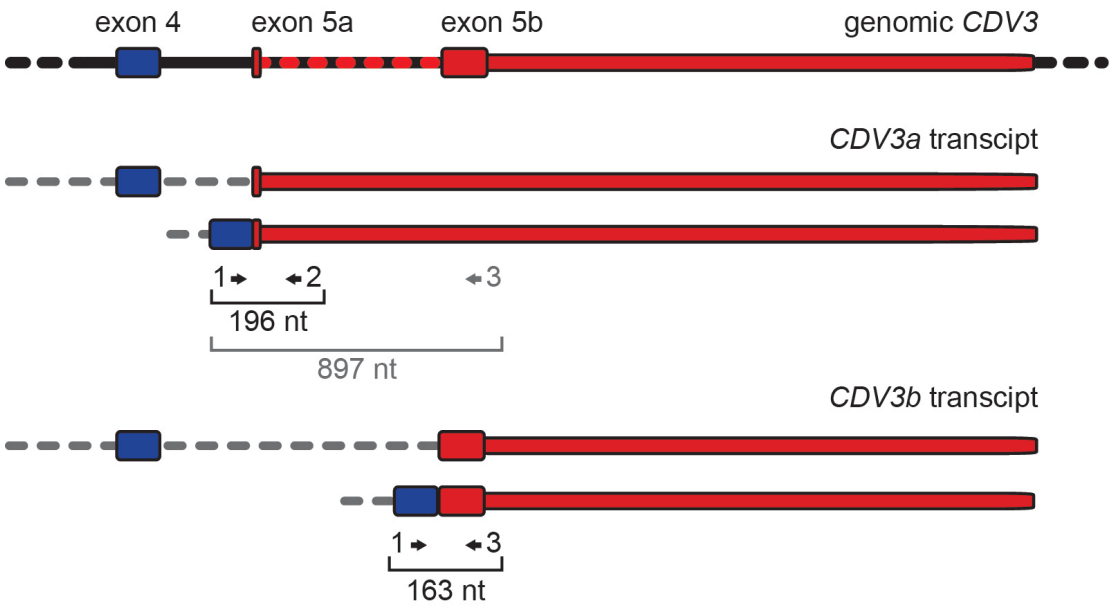


Figure 32: Detection of the the two splicing forms of the *CDV3* gene via qRT-PCR analysis. Pimers 1 and 2 were used to specifically detect a 196 nucleotide long PCR product amplified from the *CDV3a* transcript. Primers 1 and 3 were used to specifically amplify a 163 nucleotide PCR product amplified from the *CDV3b* transcript. Theoretically, the primer pair 1 and 3 could also amplify an 897 nucleotide product from the *CDV3a* transcript. However, the time given for the polymerase to elongate the products was short enough to not allow for the amplification of the longer transcript.

As depicted in Figure 32 the *CDV3a* transcript was detected by a forward primer in exon 4 and a reverse primer in its 3'-UTR giving rise to a PCR product of 196 nucleotides. This set of primers could exclusively detect *CDV3a* and not *CDV3b* because the hybridization site of the reverse primer is spliced out of *CDV3b* and is not present in the mature transcript. For detection of *CDV3b* the same forward primer was used and a reverse primer positioned in exon 5b amplifying a PCR product of 163 nucleotides. Theoretically, this primer could also bind the *CDV3a* transcript and amplify a second product of 897 nucleotides. However, the elongation time for the polymerase to produce this PCR product was set short enough to allow only for the amplification of the shorter transcript.

To test for the specificity of the primers, synthetic PCR templates corresponding to either of the two splice variants were used in combination with each of the primer pairs. Figure 33A shows that only primer pair 1 and 2 led to a PCR product with *CDV3a* as a template and conversely only primer pair 1 and 3 produced a PCR product from *CDV3b* as the template. Furthermore, the melting curves of both PCR products using total HeLa cDNA as the template were measured to demonstrate that only one PCR product is amplified. The melting curve was obtained by measuring the fluorescence signal of the double-strand DNA (dsDNA) dye in dependence of increasing temperatures. This resulted in decreasing fluorescence due to dissociation of the dsDNA binding dye when the DNA became single stranded. The change in slope of this decrease was plotted and depicted in Figure 33B for *CDV3a* and Figure 33C for *CDV3b*. Both curves showed a single peak indicating that only one PCR product was present. Lastly, as a proof of principle that the levels of *CDV3a* and *CDV3b* can be distinguished by qRT-PCR, HeLa cells were treated with PlaB, an inhibitor of splicing. PlaB treatment should leave the transcripts unspliced as precursor mRNA. Due to the fact that the primers should not be able to detect unspliced pre-mRNA, a difference in expression of both transcript variants should be visible. qRT-PCR for *β-actin* (*ACTB*) and both transcripts was performed. In the case of *ACTB*, less PCR product was expected because one of its primers binds at an exon-exon junction that is not present in the unspliced transcript. Figure 33D shows that

indeed the amount of *ACTB* RNA was reduced to one third. The primer pairs for *CDV3a* and *CDV3b* theoretically could still produce a PCR product on the pre-mRNA of *CDV3* and *CDV3a*, however they would be too long to be elongated in the chosen PCR program. Thus, no *CDV3* PCR product was expected in the PlaB-treated samples. The expression levels of *CDV3a* (Figure 33E), *CDV3b* (Figure 33F) and the ratio of the expression level of the two in respect to each other (Figure 33G) are plotted. Both the levels of *CDV3a* and *CDV3b* mRNA decreased after PlaB treatment in comparison to cells that had been treated with the solvent control. As expected, the RNA extract from PlaB-treated HeLa cells contained significantly less *CDV3a* and *CDV3b* mRNA than that from untreated cells. Moreover, the ratio between the two splice variants remained unchanged after PlaB treatment demonstrating that both transcript variants were equally affected. In summary, a qRT-PCR-based assay with specific primer pairs that can distinctly measure the endogenous abundance of each of the two splicing variants of *CDV3* was established.

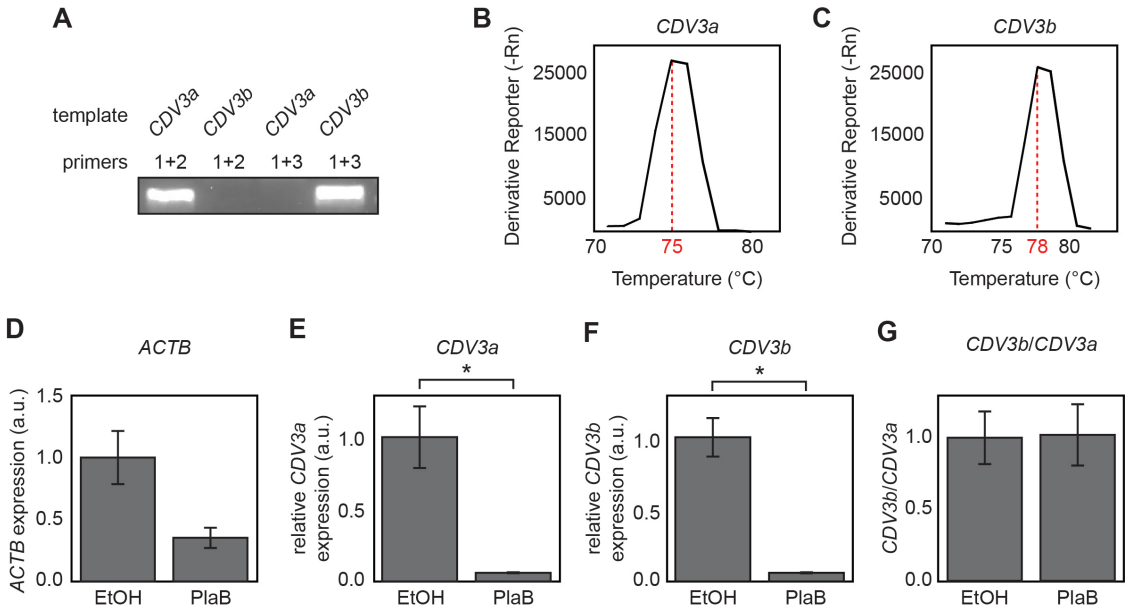


Figure 33: Establishment of a qRT-PCR-based assay to specifically detect the *CDV3a* and *CDV3b* splice variants. (A) Synthetic PCR templates were used to demonstrated that with the primer pairs used and the chosen PCR program only primers 1 and 2 can detect *CDV3a* and primers 1 and 3 *CDV3b* (see **Figure 32** for the position of the primers). **(B)** and **(C)** Melting curves for the *CDV3a* and the *CDV3b* PCR product. **(D)** - **(G)** HeLa cells were treated with the splicing inhibitor PlaB or the solvent control ethanol (EtOH). **(D)** Expression levels of *ACTB* are depicted, normalized to the solvent-treated cells. **(E)** and **(F)** Plots showing the relative levels of *CDV3a* and *CDV3b* RNA normalized to *ACTB* mRNA and the solvent-treated cells. **(G)** The ratio of the two transcript variants is plotted. n=3, *p<0.05 (Student's t-test).

EZH2 inhibition changes the ratio between the CDV3 transcript variants

The above-established qRT-PCR assay was applied to measure whether the local level of H3K27me3 at the *CDV3* locus had an effect on the ratio between the *CDV3a* and *CDV3b* transcript variants. To this end, HeLa cells were treated with 5 µM GSK343 to reduce the levels of H3K27me3 (Figure 30 and Figure 31). *CDV3a* and *CDV3b* levels and the ratio of the two splice variants were determined by qRT-PCR after three and six days of treatment. As plotted in Figure 34A and Figure 34C, the levels of *ACTB* remained unchanged by the inhibitor treatment so that it could be used for normalization of the *CDV3a* and *CDV3b* expression levels. Figure 34B shows that *CDV3a* expression increased after three days of EZH2 inhibition, while *CDV3b* expression remained unchanged (Figure 34C). The ratio of *CDV3b* over *CDV3a* expression decreased after GSK343 treatment as depicted in Figure 34D.

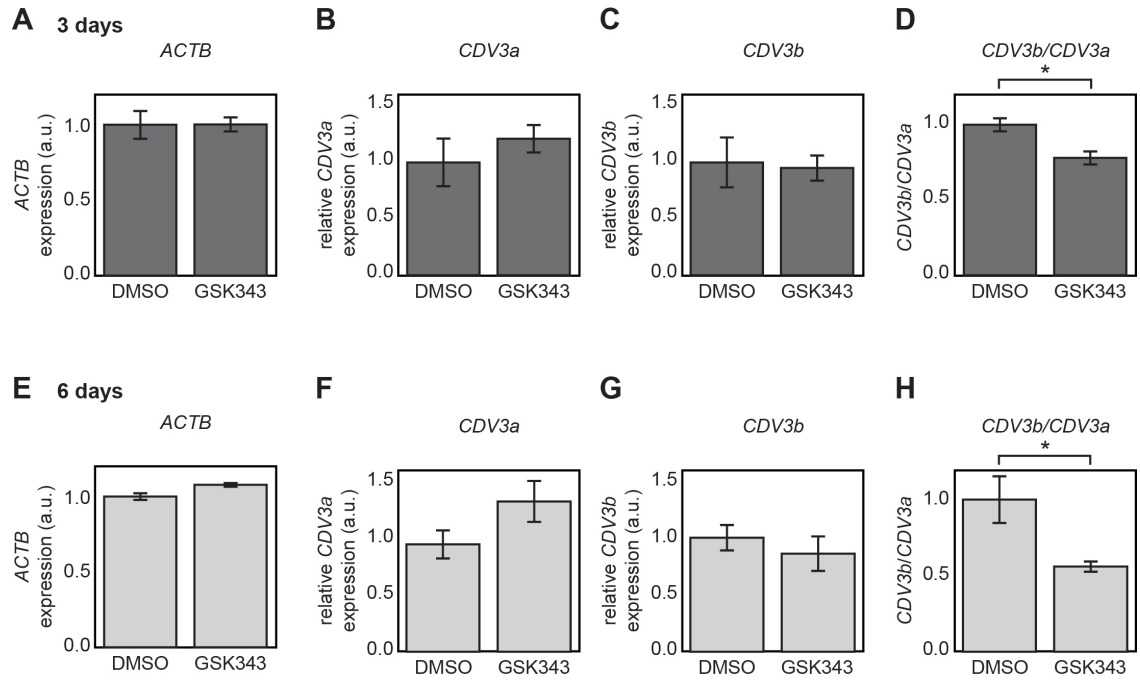


Figure 34: EZH2 inhibition changed the *CDV3a* and *CDV3b* ratio. HeLa cells were treated with 5 µM of the EZH2 inhibitor GSK343 for three and six days. **(A)** qRT-PCR of *ACTB* after inhibitor treatment for three days. **(B) – (D)** qRT-PCR of the two splice variants *CDV3a* and *CDV3b* was performed after three days. The relative mRNA levels of *CDV3a* (B) and *CDV3b* (C) normalized to *ACTB* mRNA and the ratio of *CDV3b* over *CDV3a* (D) are shown**. (E) – (H)** Same as (A) – (D) after six days of inhibitor treatment. n=5, *p<0.05 (Student's t-test).

The same trend was observed after 6 days of EZH2 inhibition (in Figure 34E - Figure 34H). Here, the increase in expression level of *CDV3a* was even more pronounced, whereas *CDV3b* expression remained unaffected. The ratio of the two splice variants was thus significantly influenced by the inhibition of EZH2, supporting the hypothesis that the local H3K27me3 level indeed affects the expression ratio of the two *CDV3* splicing forms.

4.2 CU-RNA-T0 is enriched in splicing and H3K27me3 promoting factors

Next, the aim was to differentially identify proteins that bind CU-RNA-T0 and CU-RNA-ΔT0 to gain insight into the mechanism that connects alternative splicing of the *CDV3* transcripts to the H3K27me3-promoting activity of CU-RNA-T0. To this end, RNA affinity purification of both RNAs was carried out as described in Figure 28 and the protein interaction partners were compared. An RNA of similar size (*alu*RNA$_R$), CU-RNA-T2 and RepA were included as controls. As depicted in Figure 35A, CU-RNA-T0, CU-RNA-ΔT0, CU-RNA-T2 and RepA pulled down more protein than the control RNA and the beads. Moreover, the patterns of bands that were enriched with CU-RNA-T0, CU-RNA-ΔT0, CU-RNA-T2 and RepA differed for all four RNAs. This suggests that the RNAs pulled down diverse proteins with various affinities and thus had distinct protein interaction partners. The entire lanes were subjected to mass spectrometry. For analysis of the mass spectrometry data, stringent parameters with regards to protein and peptide probabilities were applied. Only proteins, for which at least two peptides were identified with high confidence, were taken into consideration. Many proteins that were identified were known RBPs, demonstrating the specificity of the assay. While EZH2 and JARID2 were earlier found to bind CU-RNA-T0 as detected by western blotting (Figure 28), they could not be identified in the mass spectrometry data using these criteria. It remains unclear how the discrepancies between mass spectrometry and western blotting arise. However, explanations could be found in the high binding specificity of the

antibodies used for western blotting and the high abundance of proteins in the pull-down samples.

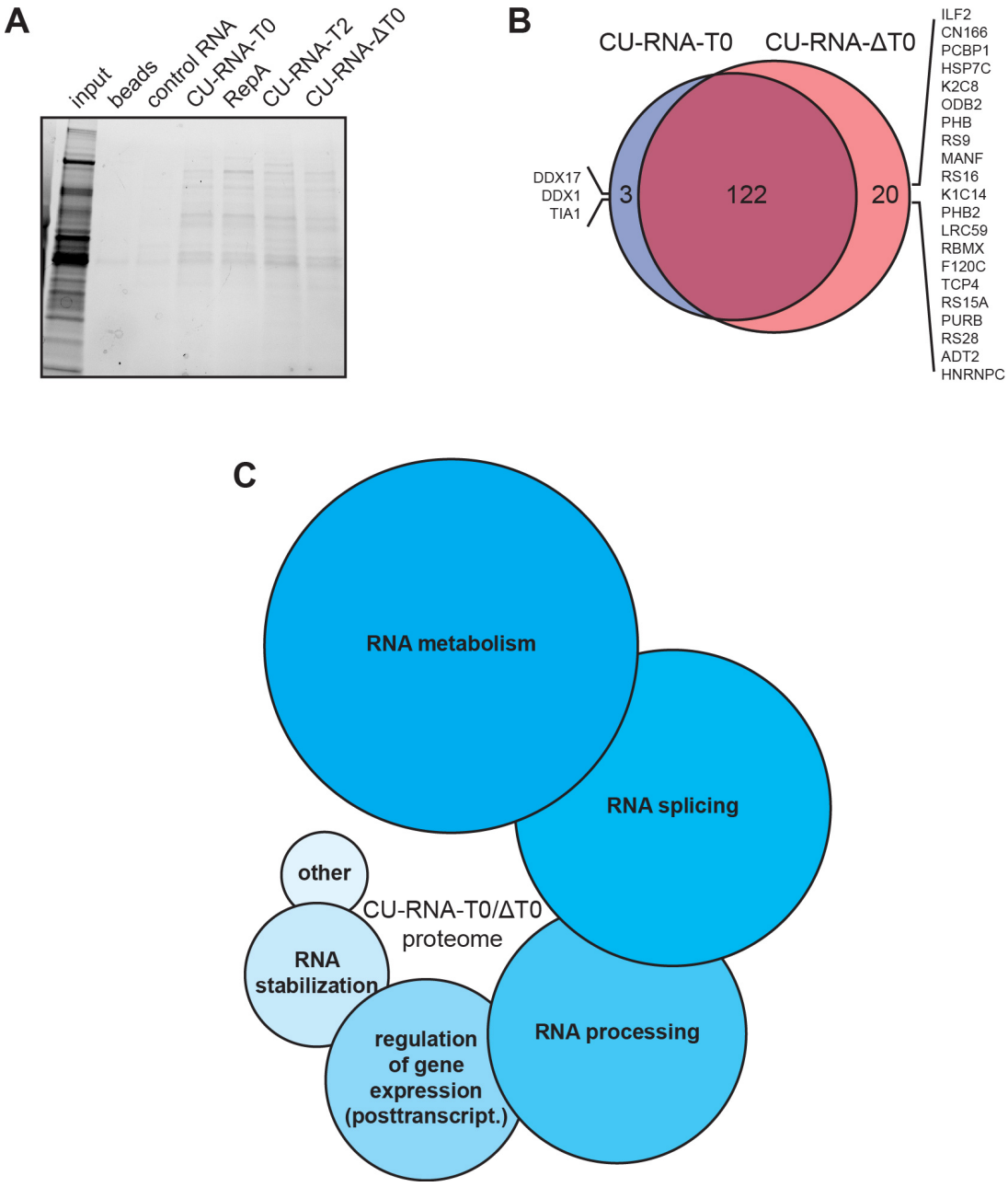


Figure 35: RNA affinity purification of CU-RNA-T0 and CU-RNA-ΔT0 enriched proteins that are involved in RNA metabolism, splicing, processing, stabilization and posttranscriptional regulation of gene expression. RNA affinity purification and mass spectrometry was carried out as described in **Figure 28**. **(A)** Proteins that were retrieved in the indicated samples by RNA affinity purification were separated by gradient SDS-PAGE. The entire lanes were subjected to mass spectrometry. **(B)** Venn diagram depicting the overlap and unique proteins that were pulled down by CU-RNA-T0 and CU-RNA-ΔT0. **(C)** GO term analysis of the CU-RNA-T0/ΔT0 interacting proteins. The five GO terms identified with the lowest p-values ($7x10^{-66}$ – $7x10^{-35}$) are depicted. The sizes of the circles are assiged according to the p-values. Most of the proteins interacting with CU-RNA-T0 could be assigned to play a role in RNA metabolism, splicing, processing, stabilization and postranscriptional regulation of gene expression.

Figure 35B gives an overview over all proteins that were pulled down by CU-RNA-T0 as compared to CU-RNA-ΔT0. The majority of proteins was pulled down by both transcripts and only a minority was uniquely pulled down by either of the two transcripts. A gene ontology (GO) term analysis of the common CU-RNA-T0/ΔT0 proteome was performed based on biological processes. Figure 35C shows the five most relevant GO terms. This analysis revealed that CU-RNA-T0 and CU-RNA-ΔT0 were primarily bound by proteins that play a role in RNA metabolism, splicing, processing, and stabilization and in posttranscriptional regulation of gene expression (for a complete list of all proteins identified by mass spectrometry please refer to Appendix 2). Also proteins, which bound uniquely to CU-RNA-T0, could be assigned to functions in RNA processing, metabolism or splicing. The proteins that were uniquely found to bind CU-RNA-ΔT0 enriched for the GO terms RNA metabolism and gene expression but also other very diverse other biological processes. Most of these proteins were, however, only identified with very low peptide count so that it remains to be determined whether they have functions that are necessary to distinguish CU-RNA-ΔT0 from CU-RNA-T0 in terms of the H3K27me3-promoting capacity.

To gain insight into the mechanism that might be involved in linking alternative splicing at the *CDV3* locus to the H3K27me3 inducing-capacity of CU-RNA-T0, all proteins that were enriched at CU-RNA-T0 in comparison to CU-RNA-ΔT0 with a factor of at least 1.3 were investigated more closely (Table 15). In combination with a literature research, the following proteins from this list were considered to be of particular interest fort this thesis: PTBP1, U2AF2 and HNRNPK. The splicing regulator PTBP1 (polypyrimidine tract binding protein 1) has been described to play a role in linking alternative splicing to chromatin via a protein-protein interaction with MRG15 (mortality factor 4 like 1), a reader of H3K36me3 (Gonzalez et al, 2015; Luco et al, 2010). U2AF2 (U2 small nuclear RNA auxiliary factor 2), also a regulator of splicing, has not directly been linked to chromatin modifications at the locus of the transcript that is to be spliced. However, it is interesting in this context because it plays a role in defining the functional 3’

splice site of introns and is an opponent of HNRNPC (heterogeneous nuclear ribonucleoprotein C) by competing for the same binding sites on the pre-mRNA (Zarnack et al, 2013). HNRNPC interestingly was found to bind CU-RNA-ΔT0 (see Appendix 2). Thus, the interplay between U2AF2 and HNRNPC might play a role in defining whether the pre-mRNA is spliced in favor of *CDV3a* or *CDV3b*. The nuclear matrix protein HNRNPK (heterogeneous nuclear ribonucleoprotein K) was found to bind the XIST RNA and has been described to play a role in the specific XIST-mediated recruitment of repressive chromatin marks such as H3K27me3 (Chu et al, 2015; Minajigi et al, 2015). It is therefore interesting in this context because it might play a role in H3K27me3 transmission mediated by CU-RNA-T0 at the *CDV3* locus. Furthermore, HNRNPK has been described to interact with PTBP1 (King et al, 2014), potentially linking the splicing reaction directed by PTBP1 to H3K27me3-induction by CU-RNA-T0. Incidentally, it is also interesting that DDX17 (DEAD (Asp-Glu-Ala-Asp) box helicase 17) only binds to CU-RNA-T0. While it has been described to act as a transcriptional co-activator (Rossow & Janknecht, 2003) it can also, in some contexts, act as a transcriptional repressor by interaction with HDAC1 (histone deacetylase 1) (Wilson et al, 2004). By interacting with CU-RNA-T0 it could therefore link deacetylation of H3K27 by HDAC1 to its methylation by EZH2 (Reynolds et al, 2011).

Table 15: RNA affinity purification of CU-RNA-T0 enriched proteins that are involved in splicing regulation and H3K27me3 induction in comparison to CU-RNA-ΔT0. RNA affinity purification was carried out as described in **Figure 28**. The table shows the peptide counts and all proteins that were at least 1.3 fold enriched at CU-RNA-T0 in comparison to CU-RNA-ΔT0. For a complete list of proteins that were identified by RNA affinity purification and mass spectrometry of CU-RNA-T0 and CU RNA-ΔT0 and the controls please refer to Appendix 2. *Proteins of particular interest for which the CU-RNA-T0 binding capacity was confirmed by western blotting (**Figure 36**).

protein	T0	ΔT0	T0/ΔT0
TIA1	7	0	> 7
DDX17	6	0	> 6
U2AF2*	28	6	4,7
PSPC1	6	2	3,0
AQR	3	1	3,0
DRG1	6	2	3,0
SND1	58	21	2,8
SF01	5	2	2,5
FUBP1	85	42	2,0
GRSF1	2	1	2,0
DDX1	2	0	> 2
RS5	4	2	2,0
RL22	4	2	2,0
IF2B3	13	7	1,9
HNRNPF	20	11	1,8
PTBP1*	43	24	1,8
TADBP	14	8	1,8
FUBP2	61	35	1,7
PUF60	5	3	1,7
HNRNH1	28	17	1,6
SMD3	11	7	1,6
G3BP1	35	23	1,5
IF2B1	12	8	1,5
RBM14	3	2	1,5
HNRNH2	19	13	1,5
SNRPA	16	11	1,5
TIAR	13	9	1,4
U2AF1	7	5	1,4
RENT1	4	3	1,3
KHDR1	12	9	1,3
EWS	4	3	1,3
NONO	84	65	1,3
HNRNPK*	57	45	1,3

The interactions of PTBP1, U2AF2, HNRNPC, and HNRNPK with CU-RNA-T0 and CU-RNA-ΔT0, respectively, were validated by western blotting as depicted in Figure 36A and Figure 36B. Both the beads alone and the control RNA interacted with the tested proteins to a much lesser extent than CU-RNA-T0, CU-RNA-ΔT0, CU-RNA-T2 and RepA suggesting that the observed interactions are specific to the RNAs tested. Moreover, the western blot against the histone H3 loading control showed that none of the RNAs, nor the beads, interacted with histone H3. PTBP1 was pulled down by RepA and CU-RNA-T2 most strongly, but importantly CU-RNA-T0 pulled down more PTBP1 than CU-RNA-ΔT0 confirming the data obtained by mass spectrometry. U2AF2 showed a similar pattern. HNRNPC, the U2AF2 opponent, was pulled down approximately equally strongly by CU-RNA-T0, CU-RNA-ΔT0, CU-RNA-T2 and RepA whereas the mass spectrometry data suggested that it was enriched at CU-RNA-ΔT0 in comparison to CU-RNA-T0. It is noted however, that the peptide count for HNRNPC is very low in the mass spectrometry data so that no real quantitative statement can be made. HNRNPK is pulled down most strongly by the RepA RNA and in agreement with the mass spectrometry data shows a slightly stronger binding to CU-RNA-T0 than CU-RNA-ΔT0. Finally, the western blot against EZH2 is depicted again for comparison. It shows that EZH2 bound all RNAs except the control RNA, although it could not be detected by mass spectrometry.

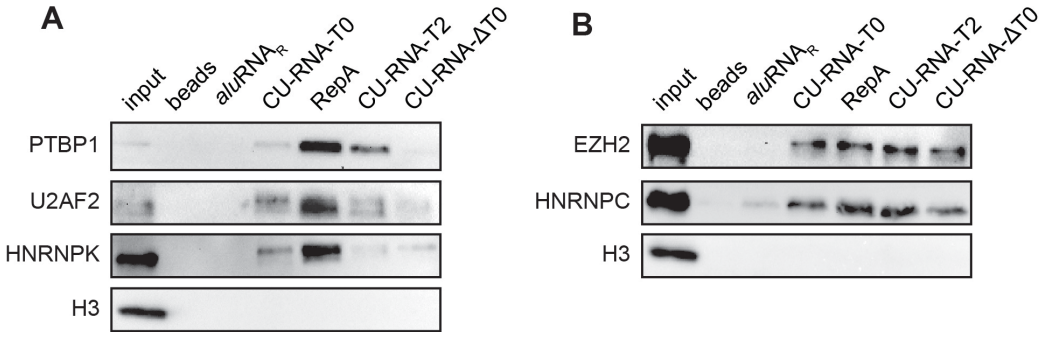


Figure 36: RNA affinity purification of CU-RNA-T0 confirmed PTBP1, U2AF2 and HNRNPK as its protein interaction partners. RNA affinity purification was carried out as described in **Figure 28**. **(A)** and **(B)** Proteins that were identified by mass spectrometry and were found to be of particular interest for this thesis were validated by western blot analysis. The same sample was equally distributed over two blots and antibodies against the indicated proteins were incubated with either one of the two blots. The EZH2 blot is the same shown in **Figure 28**.

Combining the mass spectrometry data with the western blot validations gives a comprehensive overview, over which proteins bind CU-RNA-T0 and CU-RNA-ΔT0, respectively. In summary, both RNAs bound to proteins, which are important for the processing and splicing of RNA. Additionally, CU-RNA-T0 bound HNRNPK, which has been described to play a role in RNA-mediated H3K27me3. This gives insight into the putative protein players that regulates alternative splicing of the *CDV3* transcript variants in dependence of the local level of H3K27me3, which in turn is presumably regulated by CU-RNA-T0 itself.

4.3 Knock-down of HNRNPK and MRG15 changes the ratio between the *CDV3* transcript variants

To investigate the role of HNRNPK and MRG15 in the regulation of alternative splicing of the CDV3 transcript variants, an siRNA-mediated knockdown of the two proteins was performed and the levels of *CDV3a* and *CDV3b* were measured by qRT-PCR as described in section 4.1 of the results (Figure 32). As determined by western blotting HNRNPK was knocked down to 1% of the protein level in comparison to the non-target control siRNA (Figure 37A). After HNRNPK knockdown the level of *CDV3a* did not change significantly, whereas the relative abundance of *CDV3b* decreased to 86% of the level in the control siRNA treated sample (Figure 37B and Figure 37C). The ratio of the two transcript variants – *CDV3b* over *CDV3a* – was therefore also decreased (Figure 37D). The fact that the ratio of the two transcript variants changes indicates that indeed HNRNPK is involved in the regulation of alternative splicing of the *CDV3* transcripts. Furthermore, these results suggest that HNRNPK might be involved in the regulation of the local H3K27me3 levels as proposed since its knockdown shows a similar effect as inhibiting the setting of H3K27me3.

The same experiment was repeated with a knockdown against MRG15. A knockdown to 37% of the protein level in comparison to the non-target control siRNA could be observed (Figure 37E). Furthermore the MRG15 knockdown increased the relative abundance of both the *CDV3a* and *CDV3b* transcript

variants significantly (Figure 37F and Figure 37G). Due to the fact, that the increase in the *CDV3a* variant was moderately higher compared to the *CDV3b* isoform, a small change in in the ratio of *CDV3b* over *CDV3a* was observed (Figure 37H). This suggests that also MRG15 influences the alternative splicing regulation of *CDV3*.

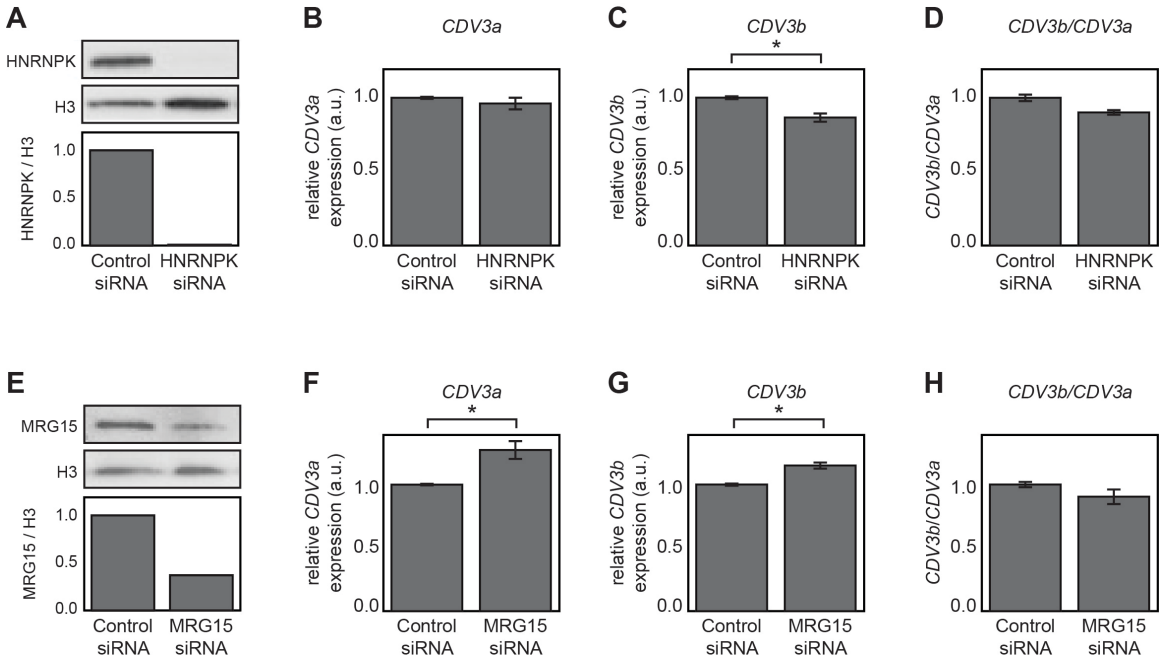


Figure 37: Knockdowns of HNRNPK and MRG15 changed the ratio between expression of the two transcript variants *CDV3a* and *CDV3b*. HeLa cells were transfected with an siRNA against HNRNPK, MRG15, or a non-target control siRNA. Cells were harvested after 72h. **(A)** and **(E)** The efficiency of the knockdown was determined by western blotting. The protein levels were quantified and the amounts of HNRNPK or MRG15 relative to histone H3 are depicted. **(B)** – **(D)** qRT-PCR of *CDV3a* and *CDV3b* was performed. Plotted are the relative expression levels of *CDV3a* (B) and *CDV3b* (C) normalized to *ACTB* mRNA expression and the ratio of *CDV3b* expression over *CDV3a* expression (D). **(F)** – **(H)** Same as (B) – (D) with a knockdown of MRG15. n=3, *p<0.05 (Student's t-test).

Concluding the results from the knockdown experiments, the mass spectrometry data and their western blot validations an alternative splicing regulatory mechanism of the *CDV3* transcript variants is proposed: It involves the specific binding of HNRNPK to CU-RNA-T0 which might be a regulatory factor for the local H3K27me3 level. Furthermore, the chromatin reader MRG15 as the connecting factor between the chromatin and the splicing machinery seems to be involved. Thereby, the *CDV3* transcript could regulate its own alternative splicing,

presumably through a sequence element in its own 3'-UTR via an H3K27me3-mediated mechanism.

Discussion

3'-UTRs are well-known for their roles in regulating mRNA stability and protein translation. However, several reports point to additional nuclear activities of 3'-UTRs (Caudron-Herger et al, 2011; Mercer et al, 2011). Therefore, it was investigated in this thesis whether and how 3'-UTRs are linked to chromatin modifications. Towards this goal, methods for the detection and functional characterization of nuclear RNAs were established. These included microinjection of *in vitro*-transcribed RNAs and RNA pools purified from distinct cellular compartments. Moreover, specific RNAs were tethered to *lac*O arrays stably integrated in the chromatin using an MS2-LacI-mediated recruitment system to investigate their effect on chromatin. The advantages and limitations of these methods are discussed in section 1. In the second part of this thesis, these methods were used for the detailed analysis of a number of specific 3'-UTR sequences and in particular the 3'-UTR sequence originating from the *CDV3* gene, called CU-RNA. A mechanism was proposed, in which CU-RNA induces H3K27me3 of its own locus and thereby establishes a feedback loop regulating its own nuclear processing. The experiments leading to this model are discussed in sections 2, 3, and 4. Finally, all experiments from this thesis are integrated with previously published observations in section 5 to present a comprehensive model of how the 3'-UTR-mediated H3K27me3 modification is connected to splicing and thereby might regulate 3'-UTR dependent cellular stability of the *CDV3* transcript. A perspective outlook is given in section 6, on how this mechanism might control nuclear RNA processing and mRNA stability, including the implications this might have for biomedical applications.

1 Microinjection and chromatin recruitment of RNAs were established as methods to investigate nuclear RNAs

Two microscopy-based methods were established in order to detect and functionally characterize nuclear RNAs. Generally, this has the advantage that

cells are studied at the single-cell level as opposed to population averaging done with biochemical and deep sequencing-based methods. The two techniques discussed here involve microinjection as a tool to monitor the subcellular localization of RNAs and to link this to their function. The second technique applies a model system that can be used to immobilize a transcript of interest at a specific genomic locus and to examine its effect on the local chromatin environment. The following sections discuss the advantages and challenges of these methods. In addition, the techniques used to visualize the RNAs in each case are put into perspective.

1.1 Microinjection as a tool to dissect the structural and functional roles of nuclear RNAs

In the first set of experiments, microinjection was established as a tool to investigate the subcellular localization of RNAs. It could be shown that RNAs of different origin and sequence behave differently after microinjection into the nucleus (Figure 8): Some transcripts remained in the nucleus whereas others distributed over both the nucleus and the cytoplasm, or were entirely found outside of the nucleus. To validate that the localizations observed by microinjection also represent the natural physiological distribution of the RNAs, methods to label the endogenous transcript, such as RNA-FISH, are informative. For CU-RNA an RNA-FISH against the endogenous transcript was also performed (Figure 18). Comparing its distribution after microinjection (Figure 16) to the RNA-FISH revealed that in both cases the RNA is found in the nucleus. RNA-FISH also gave a signal in the cytoplasm. This discrepancy is, however, not surprising given that RNA-FISH also stains the full-length *CDV3* mRNA, which can of course also be found in the cytoplasm as functional mRNA. In summary, it is concluded that microinjection is a powerful technique to analyze the subcellular localization of RNAs in living cells as also other groups have demonstrated previously (Ainger et al, 1993; Sato et al, 2015; Theurkauf & Hazelrigg, 1998; Wilkie & Davis, 2001; Yuan & Sun, 2009).

Additionally, it was investigated whether microinjected RNAs can perform their function. As a proof of principle, it could be shown that microinjection of the nucleolar RNA fraction counteracted the α-amanitin-mediated segregation of nucleolar domains (Figure 9). This provided a quantifiable measure for the extent, to which microinjected RNAs perform their function. Moreover, a single *in vitro*-transcribed prototypic *alu*RNA transcript, which was enriched in the nucleolar RNA fraction, was able to significantly promote nucleolar reassembly after microinjection (Figure 10). Biologically, these observations are very interesting as they give a strong indication that RNA influences nuclear and nucleolar structure as recently also indicated by other publications (Bergmann & Spector, 2014). Although transcripts originating from the intergenic spacer of rDNA have been implicated in the nucleolar stress response (Audas et al, 2012; Jacob et al, 2012; Jacob et al, 2013), these observations, for the first time, provide insight into the identity of regulatory nucleolar transcripts that are needed for maintenance of nucleolar structure. Furthermore, these finding collectively show that various RNAs behave differently after microinjection into the nucleus and that they have a structural influence on the nucleus, suggesting that they also perform their functions after microinjection. Microinjection was therefore used to study the subcellular localization of specific 3'-UTRs and to determine, which ones have putative functions in nuclear organization (Figure 16).

1.2 The MS2/LacI-mediated recruitment system as a tool to study RNA-protein interactions and their effect on chromatin in single cells

In this thesis, the focus was directed on exploiting the MS2/LacI system for studying the effect of RNA immobilized at chromatin. To test whether this system is suited for this purpose, two readouts were investigated in more detail. First, it was tested whether *lac*O-immobilized RNA can recruit protein interaction partners. Again as a proof of principle, it could be shown that recruitment of the *alu*RNA, the RNA that also rescued α-amanitin-induced nucleolar segregation, to the *lac*O array led to enrichment of nucleolin at the array (Figure 12). This finding

supported previous observations that *alu*RNA bound nucleolin in an *in vitro* binding assay (Caudron-Herger et al, 2015) and provided evidence that this interaction also exists in the living cell. Moreover, also nucleophosmin, another nucleolar protein, was shown to interact with *alu*RNA tethered at the array (Figure 13). This suggests that *alu*RNA exerts its nucleolar maintaining activity by interacting with nucleolar proteins similar to the previously reported interaction of rRNA with nucleolin (Tajrishi et al, 2011). Of note, these interactions might also partly explain the enrichment of *alu*RNA in the nucleus after microinjection (Figure 9 and Figure 10). As also previously demonstrated by other groups (Shevtsov & Dundr, 2011), chromatin-tethered RNA can thus interact with and recruit its protein interaction partners. This demonstrates the well-suited application of the LacI/MS2-mediated recruitment system to study RNA-protein interactions in the nucleus in the living cell.

Second, it was investigated whether *lac*O-tethered RNA could also affect nuclear localization of the arrays. As depicted in Figure 13, arrays enriched in *alu*RNA showed an increased localization within or in proximity to the nucleoli supporting the model of *alu*RNA-promoted nucleolar assembly. The MS2/LacI-mediated recruitment system has thereby been demonstrated to be a valuable system to study the contribution of RNA to chromatin organization. Since the *lacO* sequence is part of the E.coli lac operon, it offers a unique, well-defined binding site after incorporation into the genome of mammalian cells. However, a potential disadvantage might be that the array consists of repetitive DNA sequences, which might not properly mimic the mammalian chromatin at coding regions. To address this issue, one could make use of TALENs- (Miyanari et al, 2013) or CRISPER/Cas-mediated tagging of chromatin (Chen et al, 2013). Nevertheless, both systems would also require the choice of a specific sequence in the chromatin, which might still introduce bias into the observations.

In addition, the MS2 stem loop sequences and the binding of the MS2 coat protein might present a rather artificial tethering of the RNA. Alternative methods for tagging the RNA could be used as described in other publications (Baron-Benhamou et al, 2004; Coller & Wickens, 2002; Daigle & Ellenberg, 2007; De

Gregorio et al, 1999; Keryer-Bibens et al, 2008; Lim et al, 2001; Wakiyama et al, 2012) (see section 4.3, introduction). Nonetheless, all of these systems would also include tagging of the RNA with a foreign sequence and binding to a specific protein, and thus present no effective improvement of this point.

This study shows that the MS2/LacI model system used in this thesis is well-suited to investigate the involvement of RNA in chromatin organization in the living cell. Therefore, it was used to study RNA-dependent chromatin compaction, changes in histone methylation levels at the arrays and the RNA-dependent recruitment of EZH2 to the arrays.

1.3 Labeling and visualizing ectopic RNAs

Visualizing microinjected RNAs in living cells

When investigating nuclear RNAs using microscopy-based methods such as microinjection or immobilization at chromatin, it is also essential to be able to make them visible. In the case of microinjection, two RNA labeling methods were chosen that bypass labeling of microinjected RNA with ectopic sequences such as aptamers or hybridization: First, *in vitro*-transcribed, fluorescein-labeled RNAs were microinjected to test for their cellular localization (Figure 8 and Figure 16). In some microinjected cells, clusters of the fluorescein signal were detected in the nucleus (for example Figure 8A, upper panel or Figure 16A, third panel). It remains to be determined whether these foci represent enrichments of the microinjected RNAs due to the transcripts physiological properties and their function in the nucleus. Alternatively, these could be artifacts caused either by aggregation of the fluorescein itself or too high concentrations of the microinjected RNAs. Additional experiments, e.g. co-staining for proteins that are known to be found in distinct foci in the nucleus, would be required to deepen this question. Generally, fluorescein did, however, allow for distinct localization of the RNAs of interest, making it a suitable compound to label microinjected RNAs.

In a second set of experiments, the microinjected RNAs were labeled by mixing with nucleic acid-binding PI. From Figure 9 and Figure 10 it can be seen that PI

labeled preferentially RNA, as the cellular compartments that are most enriched in RNA, namely the cytoplasm and the nucleoli, are stained most intensively. Furthermore, the nucleoli were also stained with high intensity due to the fact that nucleolar RNA and *alu*RNA were microinjected that relocalized to the nucleoli after being injected. The fact that *alu*RNA is an RNA that enriches in the nucleolus was supported by the MS2/LacI-mediated recruitment studies (see above in section 1.2, discussion).

Visualizing ectopic RNAs tethered at the lacO array

In the process of establishing the MS2/LacI-mediated recruitment system, an appropriate method for labeling the RNA at the array had to be developed. In contrast to the microinjection experiments, however, the staining of RNA was conducted after fixation of the cells. Labeling the RNA enriched at the array proved to be essential in all experiments conducted with this system to ensure that not only the MS2-GFP-LacI fusion protein but also the RNA of interest was enriched. Two ways of visualizing the RNA were tested: First, EU incorporation and fluorescent labeling by “click” chemistry was applied. Second, a fluorescently labeled probe was hybridized to the MS2 stem loops sequence (Goodier et al, 2010) that is part of the RNA transcript. Enrichment of RNA at the arrays could be successfully detected with both methods (Figure 11). Both techniques have their advantages and disadvantages: While incorporation of EU is experimentally easy to handle and can be readily combined with immunofluorescence, it is less specific than directly detecting the MS2 stem loops-tagged RNA. In principle, the EU is incorporated into all newly synthesized transcripts in the cell. Thus, although unlikely, the signal from the EU-labeled RNA at the array might not originate solely from the MS2 stem loop-tagged RNA. In contrast, detecting the MS2 stem loops tagged-RNA with RNA-FISH shows the presence of a specific transcript. This technique, however, requires more caution in the experimental handling. Due to incubation with formamide-containing buffers, this method is not compatible with every antibody used in subsequent immunofluorescent stainings. Additionally, controls that have been treated with RNase A have to be included to

ensure that the probe only hybridized to the RNA and not the plasmid DNA. To be even more specific in the RNA detection, one could also make use of RNA-FISH probes directly against the RNA of interest. However, using probes against the MS2 stem loops makes the system more versatile by being able to use the same probes for all RNAs of interest. Finally, since every RNA is tagged with more than one MS2 stem loop, the fluorescent signal is amplified accordingly by the hybridization of several probes per transcript. Therefore, using RNA-FISH with probes directed against the MS2 stem loops was the method applied in the course of this thesis.

2 Nuclear 3'-UTRs rescue aggregation of decondensed chromatin and compact chromatin when tethered to a specific locus

In order to dissect the mechanism, by which 3'-UTRs modify chromatin, selected 3'-UTR sequences were investigated. One of them, the CU-RNA, was classified as particularly interesting according to the following criteria: (i) the *in vitro*-transcribed RNA was able to rescue the chromatin aggregation induced by RNase A treatment, (ii) it remained in the nucleus after microinjection, (iii) its corresponding endogenous sequence was present in the nucleus as detected by northern blotting of RNA from purified nuclei and RNA-FISH, and (iv) it induced chromatin compaction at the *lac*O arrays. These findings are discussed and put into perspective in the following sections.

2.1 Role of nuclear 3'-UTRs in rescuing the RNase A-induced aggregation of decondensed chromatin regions

It was reported that digestion of nuclear single-stranded RNAs causes distinct micrometer scale chromatin aggregation of decondensed chromatin regions (Caudron-Herger et al, 2011). The RNase A-induced chromatin compaction could be rescued by a soluble nuclear RNA fraction that is enriched in long 3'-UTR

transcripts, called the F2 fraction. In this thesis, it was observed that also single transcripts that were highly enriched in this RNA fraction were able to rescue the RNase A-induced chromatin aggregation (Figure 15). This activity was specific for transcripts such as the 3'-UTRs of *SSR3*, *CDV3*, *SURF4*, *STARD7*, and *CORO1C* and could not be observed to the same extend for all other 3'-UTRs and non-3'-UTR transcripts tested.

It is interesting that single *in vitro*-produced transcripts from the F2 fraction could rescue the effect of the RNase A treatment. This points to a very strong chromatin rescuing phenotype of the 3'-UTRs tested here. Observations made in the microinjection experiments supported this possibility: With the exception of *SURF4*, all RNAs that rescued RNase A-induced chromatin compaction were also found to reside in the nucleus after their microinjection (Figure 16). This suggests that they have a role in the nucleus. It is, however, noted that the concentration of RNA that is added is likely to be above the physiological concentration. Assuming that several of the RNAs present in the F2 fraction share a common mechanism of action, the fact that only a single type of transcripts is added could therefore be compensated by the concentration used in this experimental setup.

While the rescue experiment points to a structural role of these specific 3'-UTR in chromatin organization, their mechanism of action remains unclear. In the first place, it cannot be fully explained why RNase A treatment of mammalian interphase nuclei induces such vast rearrangements of the chromatin structure. Since RNA has long been known to be a major component of chromatin, it seems logical that removing the RNA will affect chromatin organization (Huang & Bonner, 1965; Paul & Duerksen, 1975; Pederson & Bhorjee, 1979). As listed by Caudron-Herger et al. (Caudron-Herger et al, 2011), there are a number of perturbations of cellular function that lead to a similar chromatin phenotype: for example treatment of cells with the RNA Pol II inhibitor α-amanitin (Caudron-Herger et al, 2011) or an increased degree of histone acetylation (Görisch et al, 2005; Lleres et al, 2009; Toth et al, 2004) also results in chromatin rearrangements as observed after RNase A treatment. Perturbation of histone acetylation and the involved

enzymes might be regulated by nuclear RNAs, for example by signals found in 3'-UTRs. In fact, mRNAs bind many more proteins than the originally known RNA processing factors. They contain many protein interaction motifs encoded in their sequence or secondary structure (Baltz et al, 2012; Castello et al, 2012; Karapetyan et al, 2013; Mignone et al, 2002). The interaction of the 3'-UTRs with chromatin modifying proteins might explain the changes in chromatin morphology observed here. Thus, one possibility is that the chromatin topology is influenced by the 3'-UTRs via regulations of the histone modification state as suggested by Caudron-Herger et al. (Caudron-Herger et al, 2011). Alternatively, the 3'-UTRs could perform their function indirectly by acting as decoys for proteins that otherwise would induce chromatin compaction. Their digestion by RNase A would therefore cause chromatin aggregation.

In view of the follow-up experiments in this thesis that showed EZH2 binding to the 3'-UTRs (Figure 21), the PRC2 complex is a potential candidate for such an RNA-dependent chromatin-organizing complex. The polycomb complexes have been described to be able to compact chromatin (Engreitz et al, 2013; Francis et al, 2004; Margueron et al, 2008; Naughton et al, 2010; Rego et al, 2008) and to bind RNA (Brockdorff, 2013). Furthermore, it has been shown that transcriptional inhibition in mouse embryonic stem cells is sufficient to induce genome-wide PRC2 recruitment to its target genes (Riising et al, 2014). In this publication, it is suggested that the transcriptional state governs PRC2 binding. Thus, although it has not been confirmed so far and would need further experimental validation, in the unperturbed state, the 3'-UTRs would keep the RPC2 complexes away from the chromatin, disabling them to compact the chromatin. When cells are treated with RNase A the polycomb complexes would no longer be kept in solution by the RNAs and would be recruited to the chromatin leading to a global compaction phenotype. A similar mechanism has previously been suggested by other groups (Davidovich et al, 2014; Davidovich et al, 2013; Kaneko et al, 2013). Furthermore, analogous mechanisms could be envisioned for RNA-binding chromatin modifying complexes other than the polycomb complexes, although direct experimental evidence is lacking.

2.2 Chromatin compaction and H3K27me3 induction by CU-RNA at the *lac*O arrays

From the microinjection and rescue experiments, a number of specific 3'-UTRs were identified as particularly interesting nuclear 3'-UTRs with chromatin organizing function. However, only CU-RNA, the 3'-UTR of the *CDV3* transcript, showed an effect on chromatin compaction when recruited to the *lac*O arrays (Figure 17). It is intriguing that recruitment of a specific RNA can be the trigger for such a drastic rearrangement of chromatin. Classically, chromatin compaction is associated with the formation of heterochromatin (Trojer & Reinberg, 2007). The H3K9me3 mark, which is enriched at silenced genes and is a mark of constitutive heterochromatin, (Lehnertz et al, 2003; Muller-Ott et al, 2014) could not be found at the arrays after CU-RNA recruitment (Figure 19). However, the H3K27me3 mark was found specifically after CU-RNA recruitment (Figure 20) as schematically depicted in Figure 38.

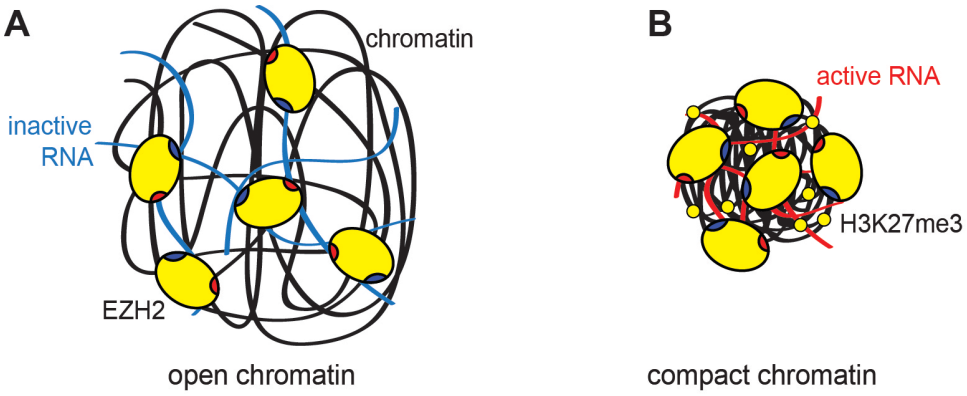


Figure 38: Chromatin states of the *lac*O array with recruitment of different RNAs. **(A)** Tethering of an "inactive" RNA recruits EZH2 but cannot promote H3K27me3 and leaves the chromatin decondensed. **(B)** Tethering of an active RNA, such as CU-RNA, recruits EZH2, promotes H3K27me3 and leads to chromatin condensation.

It has been reported that repressive histone marks such as H3K27me3 alone are not necessarily sufficient for chromatin compaction in heterochromatin formation (Chandra et al, 2012). Thus, other factors in addition to the CU-RNA-promoted H3K27me3 at the array could be important to induce chromatin compaction. In the case of X chromosome inactivation for example, it is still under debate

whether PRC2 recruitment and H3K27me3 are the cause or consequence of chromatin compaction. A number of additional factors such as histone H2A monoubiquitination, the heterochromatin protein 1-binding protein HBiX1 and the structural maintenance of chromosomes hinge domain–containing protein 1 (SMCHD1) have been suggested to be essential for organization of the compact Xi structure (Kalb et al, 2014; Nozawa et al, 2013). Nonetheless, there seems to be a link between PRC2, H3K27me3 and the compaction of the X chromosome as it is proposed that both the PRC2 complex and SMCHD1-HBiX1 are needed for compaction of Xi. Furthermore, the polycomb complexes themselves have been described to be able to compact chromatin (Francis et al, 2004; Margueron et al, 2008; Naughton et al, 2010; Rego et al, 2008). Thus, although no direct connection between EZH2 and chromatin compaction has been found so far, the presence of the enzyme at the arrays after CU-RNA recruitment was tested here and confirmed (Figure 21). In conclusion, H3K27me3 together with EZH2 and potentially further – yet to be discovered – factors might explain the CU-RNA-mediated chromatin compaction of the *lac*O arrays through a mechanism that seems to resemble compaction of the inactive X chromosome.

3 EZH2 binds RNA unspecifically and exerts its histone methylation activity only upon interaction with CU-RNA-T0

By systematically studying fragments derived from the CU-RNA, different activities were assigned: the RNAs were either only able to recruit EZH2 to the *lac*O array or to recruit EZH2 and additionally induce the H3K27me3 modification. The binding of the CU-RNA to EZH2 and its influence on the catalytic activity are discussed here.

3.1 CU-RNA contains a 250 nucleotide long functional element, T0, that is needed for its H3K27me3 inducing activity

After the 3 kb long CU-RNA had been identified as an RNA that not only recruited EZH2 to the *lac*O arrays but also promoted H3K27me3, it was cut down to its

minimal functional element that still performs this function, the CU-RNA-T0. Testing CU-RNA-ΔT0, the CU-RNA-T0-corresponding sequence originating from a *CDV3* alternative transcript variant, *CDV3b*, showed that replacing the first 80 nucleotides of the 5' end fully abolished the H3K27me3-promoting activity. Furthermore, it could be demonstrated that recruiting solely these 80 nucleotides, or further smaller fragments from the CU-RNA-T0 was not sufficient to induce H3K27me3 at the *lac*O array. Therefore, the 80 nucleotides at the 5' end and an additional substantial part of the CU-RNA-T0 are essential to perform the H3K27me3 function (Figure 23 and Figure 26).

Advantages of having identified this 250 nucleotide long transcript from the 3 kb CU-RNA are the following: First, cutting down the long CU-RNA automatically generated CU-RNA-T2 as a negative control that in subsequent experiments could be used as a non-H3K27me3-promoting transcript that has the same length as CU-RNA-T0. Second, the smaller fragment is easier to manipulate in terms of its experimental handling. Third, RNA affinity experiments become much more specific to the function that is actually ascribed to the CU-RNA-T0 transcript in comparison to doing the pull-down with the whole CU-RNA transcript. Fourth, the binding of RNA to the PCR2 complex has been suggested to be a function of the size of the RNA of interest (Davidovich et al, 2013). According to this publication, CU-RNA-T0 is roughly in the proposed size range were maximal binding of RNA to PRC2 is observed but where increasing its length would probably not increase its affinity. In summary, a minimal functional sequence element of the CU-RNA was successfully identified that could be further investigated to reveal the mechanisms by which it operates.

3.2 RNA binding to the PRC2 complex is not specific

Interestingly, all RNAs investigated in the LacI/MS2-mediated chromatin recruitment system, even the MS2 stem loops alone, and also the 80-nucleotide short transcripts derived from the CU-RNA-T0 (CU-RNA-S1, CU-RNA-S2 and CU-RNA-S3) could enrich EZH2 (Figure 22, Figure 24, Figure 26). The positive

control, RepA was also found to recruit EZH2 to the arrays (Zhao et al, 2008). With this assay, it remains unclear whether EZH2 is the component of the PRC2 complex that promotes the interaction with the RNAs. Alternatively, the interaction could also be mediated by one of the other PRC2 components that have RNA-binding capacities (Cifuentes-Rojas et al, 2014). Nonetheless, this finding was in agreement with the promiscuous binding of RNA to the PRC2 complex observed both *in vitro* (Davidovich et al, 2013) and *in vivo* as suggested by a number of publications (Guil et al, 2012; Kaneko et al, 2014b; Kaneko et al, 2013; Kanhere et al, 2010; Karapetyan et al, 2013; Khalil et al, 2009; Zhao et al, 2010). While RNA binding to the PRC2 complex has been suggested to be rather promiscuous, it has been stated that its interactions with RNA are still specific (Davidovich et al, 2015). Having this controversy in mind, the binding specificity of RNA to PRC2 and EZH2 in particular remains to be dissected in detail. As far as the MS2/LacI-mediated recruitment assays conducted here in the living cell can reveal, there seem to be only subtle differences in the binding affinity of EZH2, directly or indirectly, to the tested RNAs.

3.3 RNA affects the catalytic activity of EZH2

While all RNAs tested in this work could recruit EZH2 to the *lac*O arrays, only certain transcripts could also induce H3K27me3, suggesting that the RNAs differ in their effect on the catalytic activity of EZH2. Besides its H3K27me3-promoting activity, CU-RNA was also the only 3'-UTR that also induced chromatin compaction at the arrays as discussed in section 2.2. The influence of RNA on EZH2 catalytic activity both *in vitro* and *in vivo* is only partly understood. This thesis supports previous findings showing in an *in vitro* methylation assay that the EZH2 catalytic activity is inhibited by any RNA transcript at approximately equimolar concentrations (Figure 27) (Cifuentes-Rojas et al, 2014; Davidovich et al, 2013). In contrast, it is observed for the first time using the MS2/LacI-mediated recruitment system that there are differences in the capacity of different RNAs to promote H3K27me3 in living cells (Figure 20 and Figure 23). It can be argued

that the *lac*O array does not represent a natural mammalian chromatin environment and that the endogenous H3K27me3-promoting capacity of CU-RNA might be different. Nevertheless, it was shown that with this system tethering of the positive control RepA, an essential factor for H3K27me3 in X chromosome inactivation (Zhao et al, 2008), also triggers the H3K27me3 modification. This suggests that the MS2/LacI-mediated recruitment assay is a valid system to study the effect of specific RNAs on the posttranslational histone modification H3K27me3. It is remarkable that the capacity of CU-RNA and CU-RNA-T0 to promote H3K27me3 at the array is even higher than that of the positive control RepA.

To explain the inconsistencies between the *in vitro* methylation assay and the observations made in living cells, there must be additional factors in the cell that rescue the inhibitory effect of RNA on EZH2 activity. Figure 39 illustrates a model to explain the discrepancies between the RNAs that induce H3K27me3 at the *lac*O array and those that do not promote H3K27me3. For many RNAs, binding to EZH2 inhibits its catalytic activity (Figure 39A) so that no H3K27me3 is set at the array, even though the PCR2 complex is present. In case of specific “active RNAs” (Figure 39B), it is suggested that additional factors are recruited that activate the PCR2 complex so that H3K27 is trimethylated. In this model it is assumed that EZH2 is not recruited alone but in a multi-protein complex with the other core components of PRC2. Strictly speaking, this has not been shown in this study, since only the presence of EZH2 at the array but not that of the other PRC2 components was tested. It is also possible that with “inactive RNAs” just EZH2 and not the whole PRC2 complex is recruited to the array, which might explain why in the one case EZH2 is catalytically active and in other cases not. Noteworthy, adding nuclear lysate to the *in vitro* methylation assay conducted in this study did not rescue the enzymatic activity of EZH2. Although it was reasoned that the lysate should in principle contain the factors distinguishing the activities of various RNAs, it appears that an *in vivo* environment as in the cells is needed for certain RNAs to rescue EZH2 catalytic activity.

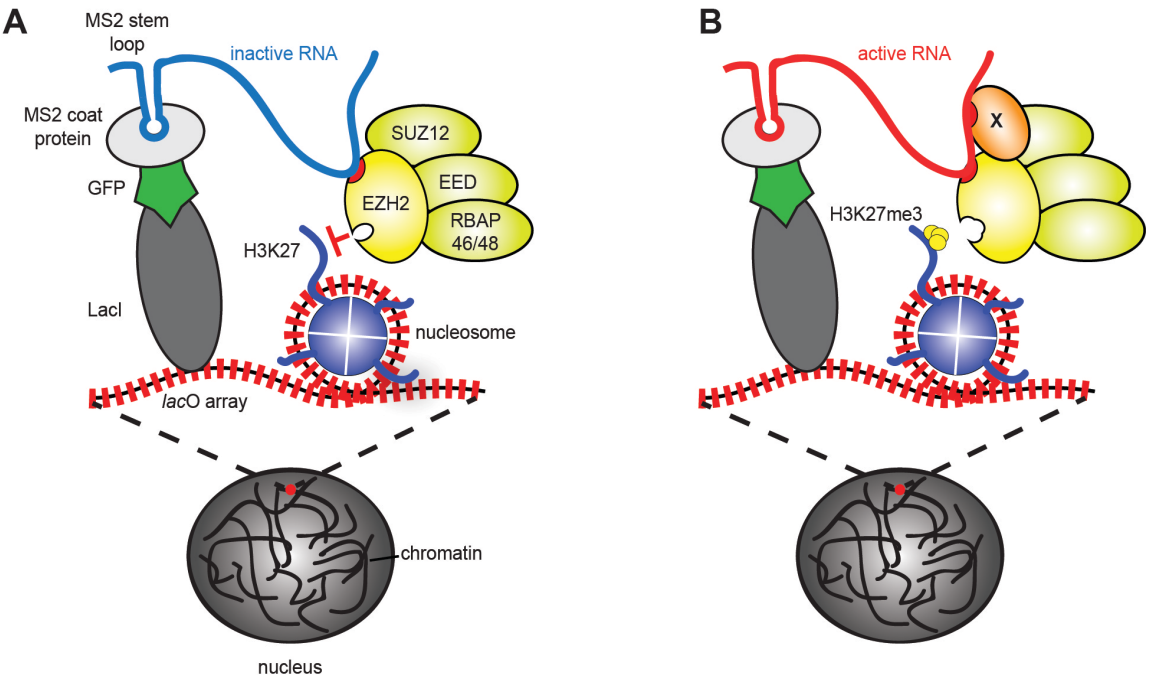


Figure 39: Schematic illustration of the discrepancy between RNAs that do not lead to H3K27me3 and those that promote H3K27me3 at the *lac*O arrays. (A) RNA is recruited to the array via the MS2-GFP-LacI fusion protein and EZH2 is recruited via its interaction with the RNA. For some inactive RNAs this interaction inhibits the catalytic activity of EZH2 and therefore H3K27 is not methylated. **(B)** In the case of "active RNAs", an additional factor, here called "X", is recruited to the array, which activates the PCR2 complex to set H3K27me3 at the *lac*O array.

In vitro studies suggest for example that SUZ12 and EED are mandatory for the enzymatically active complex (Cao & Zhang, 2004; Cloos et al, 2008; Margueron & Reinberg, 2011; Pasini et al, 2010). SUZ12 is a particularly interesting candidate because it has been shown to also bind RNA with high affinity (Cifuentes-Rojas et al, 2014). A recent study applied quantitative mass spectrometry to determine the stoichiometry of the PRC2 complex. They could show that the PRC2 complex contains at least a monomer of each EZH2, EED and SUZ12 and additional interactors, suggesting that *in vivo* EZH2 is always in complex with the whole PRC2 (Smits et al, 2013). This would suggest that SUZ12 is present in the inactive state. On the contrary, for binding of the HOTAIR RNA to PRC2 it was found *in vitro* that only the EZH2-EED heterodimer is necessary (Wu et al, 2013). Thus, it remains to be determined whether all of the PRC2 core components that make the complex catalytically active are recruited to the array with the RNAs that are investigated here.

Assuming that in all cases RNAs recruit the whole PRC2 core complex to the arrays, it is also conceivable that accessory factors distinguish inactive from active PRC2 when certain RNAs are bound. In the recent years, the polycomb-like proteins PCL1 and PCL2 (Cao et al, 2008; Casanova et al, 2011; Sarma et al, 2008), and ATRX (Sarma et al, 2014) have been demonstrated to have a stimulating activity on PCR2 activity. Additionally, two studies investigating the XIST proteome identified the nuclear matrix protein HNRNPK by mass spectrometry (Chu et al, 2015; Minajigi et al, 2015). They demonstrated that it has a crucial function in the accumulation of H3K27me3 on the Xi, which could also point to its involvement in setting the H3K27me3 mark. Interestingly, HNRNPK was found to be slightly enriched at CU-RNA-T0 in comparison to CU-RNA-ΔT0 as shown by mass spectrometry and western blot analysis (Table 15, Figure 36 and Appendix 2) in this thesis. In the same mass spectrometry analysis, it was also shown that RepA is enriched in HNRNPK. While RepA served as a positive control and was thus expected to bind HNRNPK, the very strong enrichment observed with RepA cannot explain why the H3K27me3-promoting capacity of CU-RNA-T0 is higher than that of RepA (Figure 20). It is noted, however, that the mass spectrometry as conducted here presents no quantitative measure for the amount of identified proteins. Large differences in the peptide counts of the same proteins point to a difference in the amount of the protein present in the sample. Yet, alternative methods would be superior for quantification purposes as discussed in more detail in section 4.3. Thus additional experiments are needed to compare the enrichment of HNRNPK at CU-RNA-T0, CU-RNA-ΔT0 and RepA.

In a second experiment the functional impact of HNRNPK on *CDV3* splicing was investigated. It was found that HNRNPK knockdown influences the expression level of the CU-RNA-ΔT0 containing *CDV3b* and thereby the ratio of the two *CDV3* transcript isoforms (Figure 37). Thus, although the enrichment of HNRNPK at the CU-RNA-T0 was small, it could potentially be an important factor that distinguishes the H3K27me3-promoting capacity of the CU-RNA-T0 from that of CU-RNA-ΔT0 or CU-RNA-T2. This does not exclude that there are additional

factors other than HNRNPK that enhance the catalytic activity of EZH2 at the array in combination with CU-RNA and CU-RNA-T0.

Finally, although controversially discussed, the two nucleosome binding factors AEBP2 (Kim et al, 2009) and JARID2 (Lee et al, 2006; Li et al, 2010; Pasini et al, 2010; Peng et al, 2009; Shen et al, 2009) both have been proposed to enhance the catalytic activity of the RPC2 complex on their own and also together in a synergistic manner (Cao & Zhang, 2004; Son et al, 2013). JARID2 represents a particularly interesting accessory factor in view of the above-described model (Figure 38) because it has been also shown to bind RNA (Cifuentes-Rojas et al, 2014). In this thesis, it could not be ultimately determined whether JARID2 is the factor that differentiates CU-RNA-T0 from CU-RNA-ΔT0 and CU-RNA-T2. Nevertheless, there is evidence that JARID2 binds the CU-RNA-T0 *in vitro*, suggesting its involvement in the H3K27me3-promoting activity of the transcript. It will therefore be discussed in more detail in the following paragraph.

3.4 JARID2 binds CU-RNA-T0 and might be a factor that rescues the inhibitory effect of RNA on EZH2

The presence of JARID2 at the *lac*O arrays when recruiting CU-RNA-T0 and CU-RNA-ΔT0 was examined in search for a potential factor that could account for the difference in the H3K27me3-promoting capacities of the two RNAs. JARID2 was considered a good candidate in this context because a previous publication suggested that JARID2 could rescue the inhibitory effect of RNA on EZH2 activity in an *in vitro* methylation assay (Cifuentes-Rojas et al, 2014). Here, two commercial antibodies against JARID2 were tested in the MS2/LacI-mediated recruitment system. With both antibodies no clear enrichment of JARID2 at the arrays could be observed with any of the RNAs (Figure 29). This was either linked to a technical detection problem due to a lack of quality of the antibodies or JARID2 is not or only very weakly enriched. On the one hand, it cannot be fully excluded that the immunofluorescent staining of JARID2 did not work perfectly. Since the Novusbio antibody shows a multitude of bands of different size in a

western blot of nuclear HeLa extract, this antibody was not suited for immunofluorescence in the first place. The Abcam antibody gave a single band of approximately the expected size in the western blot test. Nonetheless, RNA-FISH to detect the MS2 stem loop containing RNA transcript at the array in combination with the immunofluorescent staining against JARID2 might present technical issues. Paraformaldehyde fixation and stringent washing steps as required in the RNA-FISH protocol might reduce the affinity of the Abcam JARID2 antibody for example due to changes in the epitope structure during fixation.

On the other hand, the model by Cifuentes-Rojas and colleagues has not been supported by *in vivo* experiments. They propose that RNA is the factor that targets PRC2 to chromatin but inhibits its enzymatic activity until PRC2 associates with JARID2 on chromatin. JARID2 thereby weakens PRC2's binding to RNA. The only evidence from *in vivo* studies to support this view are findings from other laboratories, reporting that JARID2 enhances PRC2 recruitment to chromatin at the Xi (da Rocha et al, 2014; Kaneko et al, 2014a; Li et al, 2010). The tests conducted here are thus the first experimental approach to study the model proposed by Cifuentes-Rojas et al. in the living cell but did not provide clear results. Nevertheless, it could still be shown that CU-RNA-T0 binds JARID2 *in vitro* in an RNA affinity purification pull-down experiment (Figure 28). Here, despite the lack of a good antibody and ambiguous data, the pull-down experiment indicated an increased affinity of CU-RNA-T0 for JARID2 in comparison to CU-RNA-ΔT0. Thus, although more detailed studies are needed, JARID2 is still a potentially interesting candidate to explain discrepancies between the H3K27me3-promoting activities of different RNAs.

4 CU-RNA regulates its own alternative splicing via establishment of a splicing specific chromatin signature

In search of the endogenous function of the H3K27me3-promoting activity of CU-RNA-T0, it was noted that it is located at the alternative last exon of the

CDV3 transcript. Therefore a connection between the H3K27me3-promoting activity and alternative splicing was investigated.

4.1 EZH2 inhibition influences H3K27me3 distribution at the *CDV3* locus and genome wide

By western blotting and FACS analysis it was found that inhibition of EZH2 with the small molecule inhibitor GSK343 reduced the level of H3K27me3 in HeLa cells (Figure 30). In addition, ChIP-seq experiments of GSK343-treated and control cells were performed to obtain insight into the genome-wide changes in H3K27me3 (Figure 31). Several points suggest that the quality of data was not ideal: First, the expected increase in reads mapping to the mouse genome after GSK343 treatment was not observed. It was expected that in the GSK343 treated samples a higher proportion of reads should map to the mouse genome than in the untreated samples because less human chromatin carrying H3K27me3 is provided to the antibody to bind its epitope. However, the expected increase in reads mapping to the mouse genome after GSK343 treatment was not observed. It remains elusive why this is the case since the GSK343 inhibitor has been shown to perform its function properly on a global level by FACS anaylsis and western blot (Figure 30). A potential explanation could be differences in the binding affinity of the H3K27me3 antibody under the experimental conditions used for ChIP-seq in comparison to FACS and western blotting.

Second, the MCORE correlation functions of H3K27me3 usually have and amplitude of up 0.1 in ChIP-seq experiments of high quality (Molitor et al, 2015), while only 0.03 was reached here. This supports the notion that the enrichment of the H3K27me3 signal over background is relatively low in the present data set. Third, in one of the replicates no peaks could be called using the Sicer algorithm (Zang et al, 2009) again suggesting a low signal to background ratio.

To be able to better quantify and calibrate the findings from the ChIP-seq analysis the so-called internal standard calibrated ChIP (ICeChIP) could be used in stead of the mouse chromatin spike-in (Grzybowski et al, 2015). This method

also employs chromatin spike-ins but uses nucleosomes that are reconstituted from recombinant and semisynthetic histones together with barcoded DNA. The authors claim that unlike spiking in chromatin from a different organism, this method could measure actual amounts of marks using very defined standards and thereby provide an even better means of data correction. Experimentally, ICeChIP is more laborious than spiking in chromatin from a different organism but might be worth the effort to obtain a quantitative overview over the differential H3K27me3 before and after GSK343 treatment.

Nevertheless, the combined analyses using MCORE and SICER showed that GSK343 treatment leads to a genome-wide redistribution of H3K27me3 and a reduction of H3K27me3 in annotated genes. Moreover, comparing the traces at the CU-RNA in the DMSO- and GSK343-treated samples demonstrated a reduction of H3K27me3 at the start of the alternative last exon of *CDV3b* and a slight increase in H3K27me3 at the upstream intron and the 3'-UTR. The findings are substantial to support the hypothesis that splicing of the *CDV3* alternative transcripts is dependent on the differential H3K27me3 at the locus.

4.2 EZH2 inhibition influences the balance between *CDV3a* and its splice variant *CDV3b*

Expression of the two splice variants of the *CDV3* transcript, *CDV3a* and *CDV3b*, was dependent on the activity of EZH2 and the levels of H3K27me3 at the *CDV3* locus. Inhibition of the catalytic activity of EZH2 led to a decrease in the local level of H3K27me3 at the alternative exon of *CDV3b* (Figure 31). Additionally, qRT-PCR of the two transcript variants revealed that the *CDV3b*/*CDV3a* ratio decreased to almost 50% after EZH2 inhibition while they were expressed at equal levels in the unperturbed state (Figure 34). The involvement of H3K27me3 in splicing decisions has been described in a number of studies. For example siRNAs targeted against an intronic sequence next to a fibronectin exon in human cells promote an AGO1-dependent local increase in the H3K27me3. This is suggested to slow down RNA Pol II and to cause the inclusion of an alternative

exon that is skipped in the unmethylated state (Alló et al, 2009). Similarly, H3K27me3 distribution on specific genes is associated with changes in alternative splicing in differentiating neurons (Schor et al, 2013). Many of the models propose a kinetic mechanism according to which the chromatin template modulates the rate of transcription, which in turn impacts splicing decisions (de la Mata et al, 2003; Shukla & Oberdoerffer, 2012). Here it is proposed that the methylation of H3K27 rather functions as a chromatin signal that is read out and affects the ensemble of proteins involved in splicing (see sections 4.3 and 5, discussion).

4.3 HNRNPK, PTBP1, U2AF2 and HNRNPC link splicing of *CDV3* to the local chromatin state

Two observations have led to the conclusion that the CU-RNA-T0 and CU-RNA-ΔT0 have distinct activities. The first one is that only CU-RNA-T0, as part of the *CDV3a* transcript, can induce H3K27me3 at the *lac*O array when immobilized there. The second one is that EZH2 inhibition changes the ratio of the two transcript variants *CDV3a* and *CDV3b*. The latter suggests that different splicing regulators are recruited depending on the H3K27me3 state of the chromatin locus. Based on these two observations, it was speculated that CU-RNA-T0 and CU-RNA-ΔT0 interact with distinct protein partners that function together with the transcripts to perform their function. RNA affinity purification in combination with mass spectrometry was therefore carried out to map the CU-RNA-T0 and CU-RNA-ΔT0 protein interaction partners. The majority of proteins pulled down by either of the two and the control RNAs were RNA-binding proteins (Figure 35 and Appendix 2). This confirms that the affinity purification protocol specifically enriched for RNA-binding proteins at the different sequences. Furthermore, the fraction of proteins that bound CU-RNA-T0 and CU-RNA-ΔT0 largely overlapped. Since both RNA fragments are derived from the 3'-UTR of *CDV3* and are identical in 170 out of 250 nucleotides this is not surprising. Nevertheless, several proteins were identified that could be responsible for the

H3K27me3-promoting activity of CU-RNA-T0 and its involvement in regulating alternative splicing (Table 15 and Figure 36) and are discussed in the following.

U2AF2

One of the most enriched proteins at CU-RNA-T0 in comparison to CU-RNA-ΔT0 was U2AF2, with an enrichment factor of 4.7 fold (Table 15). Binding to CU-RNA-T0 was confirmed by western blotting whereas no binding to the CU-RNA-ΔT0 was detected (Figure 36). Furthermore, both mass spectrometry and western blotting indicated that RepA binds U2AF2 even stronger than the CU-RNA-T0 (Appendix 2 and Figure 36). Whether this binding event also plays a functional role awaits further analysis. However, it needs to be considered that mass spectrometry is not quantitative *per se* due to varying detector response and differing ionization yields of the different substrates. To receive a quantitative readout that can be compared more closely to the western blot analysis (Figure 36) SILAC (stable isotope labeling by amino acids in cell culture) could be carried out (Ong et al, 2002). This method uses the fact that peak ratios from isotopic analogs are very accurate because there are no chemical differences between different species. Furthermore, alternative RNA affinity purification methods could be applied. For example, streptavidin-binding RNA aptamers could be used that were suggested to perform even better than the use of biotin-labeled transcripts in terms of labeling efficiency and background reduction (Butter et al, 2009; Leppek & Stoecklin, 2014).

Although it needs further confirmation, the differential binding to CU-RNA-T0 and CU-RNA-ΔT0 could be relevant for the regulation of the *CDV3* splicing. U2AF2 has been reported to have the capacity to define functional 3' splice sites (Valcarcel et al, 1996; Zamore et al, 1992) in 88% of the human genome (Shao et al, 2014). The 5' splice site in the *CDV3a* and *CDV3b* transcript are identical. Thus, the 3' splice site sequence provides the essential region for differentiating between the two (Figure 25). Potentially, U2AF2 is the splicing regulator that is needed to allow assembly of the spliceosome complex at the proper site in line with previous findings on its function (Ruskin et al, 1988).

HNRNPC

HNRNPC has been suggested to compete with U2AF2 binding to protect the transcriptome from the exonization of Alu elements (Zarnack et al, 2013). In the context of this thesis, HNRNPC is interesting because the mass spectrometry data suggest that it binds to CU-RNA-ΔT0 but not to CU-RNA-T0. Confirming previous observations, here this protein also interacted with the RepA transcript here (McHugh et al, 2015). However, the peptide counts with only 2 and 5 hits respectively, were rather low in both cases. This might explain why the differential enrichment of HNRNPC at the CU-RNA-ΔT0 could not be confirmed by western blot, where CU-RNA-T0, CU-RNA-ΔT0, CU-RNA-T2 and RepA demonstrated approximately equal binding affinities (Figure 36). Further quantitative analyses are needed as described above (see under U2AF2). Based on the reported involvement of HNRNPC in the intronization of certain sequences, it is nonetheless considered as potentially relevant in the regulation of *CDV3* splicing.

PTBP1

Interestingly, also the well-studied splicing regulator PTBP1 was found highly enriched at the CU-RNA-T0 as demonstrated by both mass spectrometry and the western blot (Table 15 and Figure 36). Similarly to U2AF2, both data sets also suggest that RepA and CU-RNA-T2 bind this factor most strongly. As stated above, the functional relevance of these observations remains to be investigated in more detail and with more quantitative methods (see under U2AF2). Nevertheless, the differential binding of PTBP1 to CU-RNA-T0 in comparison to CU-RNA-ΔT0 is potentially relevant for the underlying splicing mechanism. There are several publications suggesting that PTBP1 regulates alternative RNA processing by different activities (Spellman & Smith, 2006). For example it binds to both exonic and intronic CU-rich regions to obstruct exon definition (Izquierdo et al, 2005), to obstruct intron definition (Chou et al, 2000; Sharma et al, 2005) or to prevent the transition from exon to intron definition (Sharma et al, 2008). These findings are substantiated on a genome wide level in a study by Xue and colleagues revealing a positional effect of PTBP1 on the regulation of splicing

(Xue et al, 2009). There are also reports that suggest competitive binding of U2AF2 and PTBP1 making it unlikely that they both bind CU-RNA-T0 simultaneously to regulate splicing of the *CDV3* pre-mRNA (Lin & Patton, 1995; Sauliere et al, 2006; Singh et al, 1995). However, given the apparently broad role of PTBP1 in splicing regulation, involvement of both PTBP1 and U2AF2 in the *CDV3a* splicing is conceivable. Shao and colleagues (Shao et al, 2014) suggest also that there is an additional role of U2AF2 when it binds in intronic regions where it interferes with recognition of the downstream 3'-UTR splice site. With respect to splicing regulation of the *CDV3* transcript, it remains unclear how exactly the inclusion of the CU-RNA-T0 containing exon is controlled by U2AF2 and PTBP1 and where exactly in the transcript they bind. However, based on the current knowledge about the many mechanistic roles of the two proteins it appears feasible that the concerted action of the two proteins results in a splicing event favoring the *CDV3a* variant.

MRG15

PTBP1 binding to CU-RNA-T0 is of interest for the present thesis because it has been shown to bind MRG15 (Gonzalez et al, 2015; Luco et al, 2010). *In vitro* binding assays of MRG15 have indicated that its chromo-domain binds methylated lysine 36 but not lysine 4, lysine 9, or lysine 27 of histone H3 (Zhang et al, 2006). According to Gonzalez and colleagues, the PTBP1-MRG15 complex forms an adaptor that can regulate alternative splicing depending on the presence of H3K36me3 and absence of H3K27me3. MRG15 could not be detected in the mass spectrometry data likely because it is no direct interactor of CU-RNA-T0 but interacts on indirectly via PTBP1. In this thesis, a knockdown of MRG15 led to the increase in expression of both transcript variants *CDV3a* and *CDV3b* (Figure 37). This demonstrates its effect on processing the *CDV3* pre-mRNA but it remains elusive why the MRG15 knockdown affects both splicing transcripts equally. However, it has to be noted that the knockdown was not very efficient and a substantial level of MRG15 remained in the cells. Thus, further

experiments are needed to make strong conclusions about the contribution of MRG15 with respect to *CDV3* splicing.

HNRNPK

In addition to the interaction with MRG15, the splicing regulatory protein PTBP1 has been suggested to interact with HNRNPK (King et al, 2014), which is also involved in alternative splicing (Kumar et al, 2014). HNRNPK has been identified in two independent publications as a RepA-interacting protein (Chu et al, 2015; Minajigi et al, 2015). Here, HNRNPK binding to RepA was confirmed (Appendix 2). Minajigi and colleagues showed by knockdown studies that it is necessary for XIST-mediated chromatin modifications, i.e. H3K27me3, and polycomb targeting in mouse cells. It remains elusive how HNRNPK interacts with the polycomb complex to perform this function. Yet, the observation that HNRNPK induces H3K27me3 presents an important building block for the mechanism suggested in this thesis. It links the splicing machinery at the *CDV3* transcript to the H3K27me3-promoting activity of the CU-RNA-T0 that has been observed in living cells (Figure 23). In line with this hypothesis, it was also slightly enriched at the CU-RNA-T0 in comparison to CU-RNA-ΔT0 in the RNA affinity purification assay as determined by mass spectrometry and western blotting (Table 15 and Figure 36) (see also section 3.3 of the discussion). This observation could explain the unique H3K27me3-promoting activity of CU-RNA-T0. Moreover, after HNRNPK knockdown the ratio of the two *CDV3* splice variants was shifted in the same direction as after EZH2 inhibition and knockdown (Figure 37). This also suggests that it is involved in the regulation of alternative splicing at the *CDV3* locus. It is noted that only the *CDV3b* variant was influenced by the HNRNPK knockdown while the *CDV3a* transcript variant level remained unchanged. One could speculate that while the *CDV3a* transcript variant does not need H3K27me3 to be expressed, the HNRNPK knockdown mainly affects *CDV3b* expression because this isoform is dependent on HNRNPK that putatively influences the local H3K27me3 level. Furthermore, although the knockdown is almost complete, the effect on the *CDV3b*/*CDV3a* ratio is relatively

small, which might point to the involvement of additional factors in regulating the local H3K27me3.

DDX17

Finally, the helicase DDX17 was identified to uniquely bind the CU-RNA-T0 and not the CU-RNA-ΔT0 (Figure 35 and Appendix 2). A link between splicing regulation and chromatin state has not been made before for DDX17. It is noted that DDX17 interact with the H3K27 deacetylase HDAC1 (Wilson et al, 2004). This activity could thus contribute indirectly to lysine 27 methylation by EZH2 by removing a counteracting acetylation modification (Reynolds et al, 2011).

5 Model of how *CDV3* balances the abundance of its splice variants by modifying the local chromatin signature

The suggested complexes of PTBP1-MRG15 and HNRNPK-PTBP1, and also putatively DDX17-HDAC1 represent suitable candidates for linking the splicing machinery of the *CDV3* transcript variants to reading and writing of the local chromatin state at the *CDV3* locus. Together with the direct competition between U2AF2 and HNRNPC, this network of RNA and chromatin binding proteins can be integrated in a model of how CU-RNA-T0 balances the abundance of the *CDV3* splice isoforms by modulating the local H3K27me3 state (Figure 40). The following cycling mechanism is suggested:

1. The active *CDV3* locus carries the H3K36me3 modification as a hallmark of transcribed genes but lacks H3K27me3.
2. *CDV3* pre-mRNA is transcribed.
3. The splicing regulators U2AF2 and PTBP1 in complex with MRG15 and HNRNPK are recruited co-transcriptionally.
4. The recruitment of the PTBP1-MRG15 complex requires the H3K36me3 mediated targeting of MRG15 to chromatin.
5. *CDV3* pre-mRNA is spliced in favor of the *CDV3a* transcript containing the CU-RNA-T0 element. Unless PRC2 is recruited, *CDV3* pre-mRNA transcription starts again in the same cycle.
6. Alternatively, the spliced but still chromatin associated CU-RNA-T0 recruits the PRC2 complex.
7. H3K27me3 is set by PRC2 inducing the simultaneous presences of H3K27me3 and H3K36me3, which inhibits the MRG15-H3K36me3 interaction.
8. *CDV3* pre-mRNA is transcribed from the *CDV3* locus that now carries H3K27me3 and H3K36me3.
9. Due to the changed chromatin state, HNRNPC is recruited instead of U2AF2 and the MRG15-PTBP1-HNRNPK complex.
10. *CDV3* pre-mRNA is spliced in favor of *CDV3b* containing the CU-RNA-ΔT0 element and CDV3 pre-mRNA transcription starts again.
11. This splice form lacks the H3K27me3-promoting activity recruited via the CU-RNA-ΔT0 element. Accordingly, the transcription cycle will switch to producing CDV3a again, if H3K27me3 is lost spontaneously or removed actively.

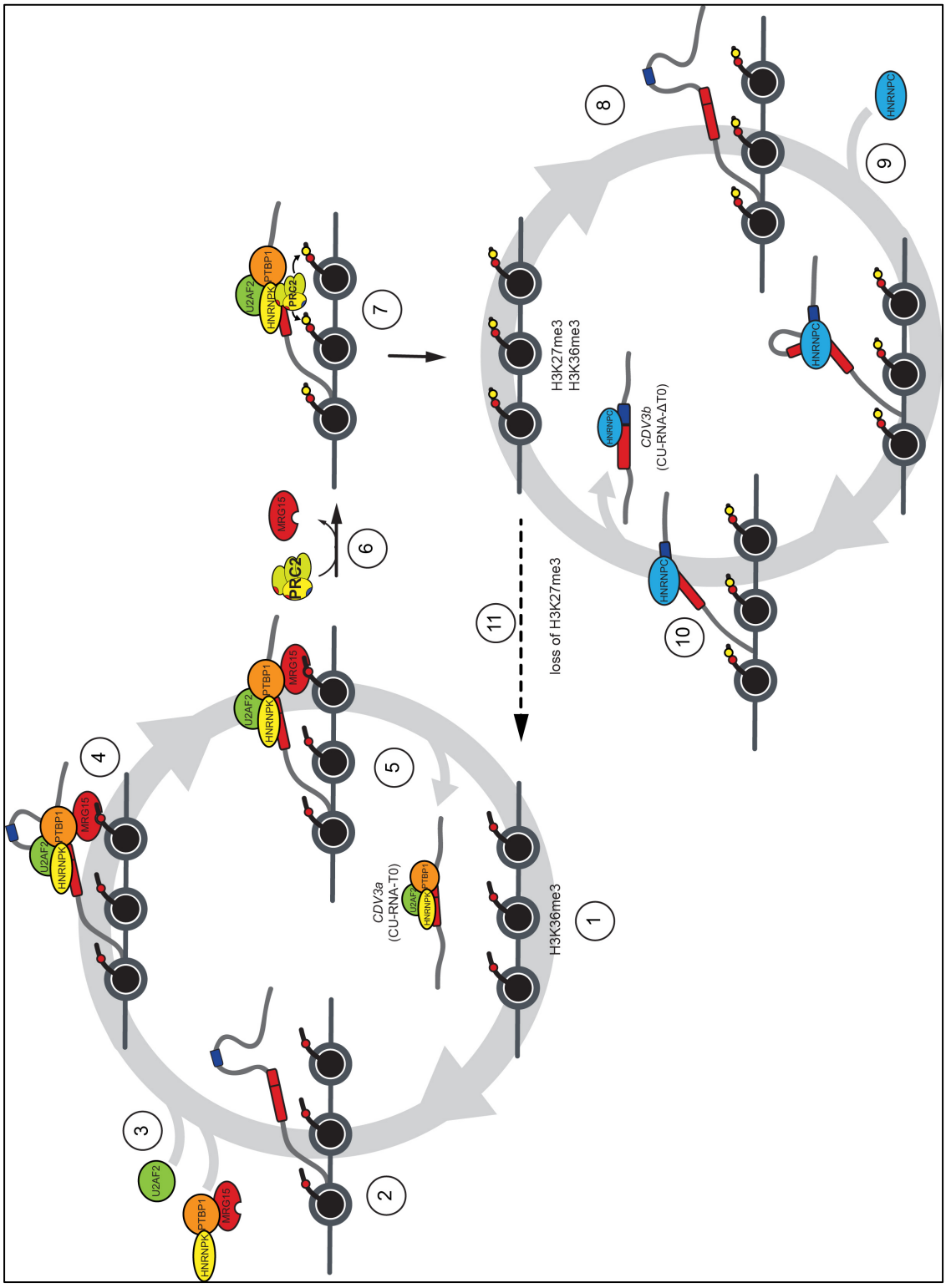


Figure 40: Model of how alternative splicing at the *CDV3* locus is regulated. A complex of PTBP1 with the CU-RNA-T0 containing transcript (*CDV3a*) promotes H3K27me3 via HNRNPK. This state favors formation of the CU-RNA-ΔT0 containing splicing variant (*CDV3b*), which lacks the ability to induce H3K27me3. Thereby a feedback loop is established linking 3'-UTR directed H3K27me3 modification to splicing.

Many of the steps in this model are substantiated by the experimental data provided in this thesis and further supported by publications from other groups: First, it is known from RNA-FISH, northern blot and RT-PCR analyses that the *CDV3* gene is expressed in HeLa cells (Figure 18). These data also show that the two transcript variants, *CDV3a* and *CDV3b*, are expressed at approximately equal levels. Whole transcriptome data from Caudron-Herger and colleagues further support this finding (Caudron-Herger et al, 2011). Accordingly, there must be a mechanism to adjust the balance between the two splicing variants as suggested in the model.

Second, it has been demonstrated that CU-RNA-T0 as part of the *CDV3a* transcript has the capability to recruit EZH2 and to induce H3K27me3 when tethered to the *lac*O arrays (Figure 23 and Figure 24). This is translated into the model by the fact that *CDV3a*, which is still chromatin tethered during the process of transcription, could induce H3K27me3 of the locus. It is noted that also the *CDV3b* contained CU-RNA-ΔT0 could recruit EZH2. The model suggests, however, that the interaction with HNRNPK is necessary for an EZH2 state that is capable to set the H3K27me3 modification (Chu et al, 2015; Minajigi et al, 2015). It could be shown here that HNRNPK binds CU-RNA-T0 stronger than to CU-RNA-ΔT0 (Table 15 and Figure 36) and that knocking down HNRNPK leads to a decrease in the abundance of *CDV3b* in comparison to *CDV3a* (Figure 37). Therefore the H3K27me3-promoting activity is depicted as a unique feature of CU-RNA-T0.

Third, it was demonstrated that reducing the level of H3K27me3 at the *CDV3* locus by EZH2 inhibition leads to a shift in the ratio of *CDV3a* and *CDV3b* in favor of the *CDV3a* variant (Figure 30, Figure 31 and Figure 34). This suggests that when the *CDV3* locus is not methylated at H3K27 the *CDV3a* transcript is produced.

Fourth, the coordination of the splicing event seems to be mediated by distinct binding of a complex of PTBP1-MRG15 to CU-RNA-T0 and the competition between U2AF2 and HNRNPC as RNA affinity purification, mass spectrometry, and western blotting suggested. In particular, the influence of the chromatin

reader MRG15 has been demonstrated by the fact that its knockdown leads to a change in the expression levels of both *CDV3b* and *CDV3a*. Moreover, as discussed in detail in section 4.3, all of these proteins have previously been shown to contribute to alternative splicing decisions in other systems and for other transcripts. The fact that pre-mRNA splicing occurs mostly co-transcriptionally, i.e. while the transcript is still associated to chromatin, is in agreement with this (Listerman et al, 2006; Tilgner et al, 2012).

A number of points depicted in the model shown in Figure 40 remain speculative and thus require further validation. As such, additional experiments, for example knockdown studies showing the involvement of the splicing factors U2AF2, HNRNPC and PTBP1 in the splicing decision, would be insightful. The contribution of the H3K36me3 reader MRG15 and the H3K27me3 promoting protein HNRNPK also need to be further validated. HNRNPK is suggested here to be a crucial factor for the formation of an active EZH2 state in complex with RNA. It thus needs to be further investigated if HNRNPK affects the enzymatic activity of EZH2 when in complex with different RNAs. This could for example be accomplished by repeating the *in vitro* methylation assay conducted in this thesis in the presence of recombinant HNRNPK or by examining the presence of HNRNPK at the *lac*O arrays when recruiting CU-RNA-T0.

An alternative candidate for modulating EZH2 activity is JARID2. Although the link between alternative splicing and JARID2 has not been made previously, results from other groups have shown that *in vitro* JARID2 can partly rescue the inhibitory effect of RNA on the methylation activity of EZH2 (Cifuentes-Rojas et al, 2014). Additionally, RNA pull-down experiments conducted in this study suggest that JARID2 is bound by CU-RNA-T0 (Figure 28) implying that it is potentially also involved in modulating the function of the EZH2-CU-RNA-T0 complex. Further studies in living cells will help to elucidate whether JARID2 is also involved in the regulation of H3K27me3 at the *CDV3* locus.

Lastly, the transitions to and from the transcription cycle leading to the expression of *CDV3b* remains poorly characterized. One unanswered questions is for example whether there is a specific reader of the H3K27me3 replacing the

MRG15 as a reader for H3K36me3. Furthermore, it is unclear how H3K27me3 is removed before the splicing machinery switches back to producing the *CDV3a* transcript. Potentially, H3K27me3 could be actively removed for example by the KDM6 family of lysine demethylases (Hong et al, 2007) or it could be lost spontaneously, e.g. during replication where the newly incorporated histone H3 lacks a H3K27me3 modification (Scharf et al, 2009). Since both methylation and demethylation are rather slow processes (Zee et al, 2010), this also raises the question of how quickly the switch between *CDV3a* and *CDV3b* splicing can occur for example in response to external stimuli.

In summary, the model of how alternative splicing at the *CDV3* locus could be regulated by the H3K27me3-promoting activity of CU-RNA-T0 provides new insights into the connection between chromatin and splicing. It includes both the regulation of alternative splicing events by the local chromatin modifications as well as the influence of splicing on the chromatin structure. It thereby proposes a feedback loop that serves to maintain the balance between alternative transcript variants of *CDV3*. Although not addressed in this thesis, it is conceivable that external stimuli altering any of the involved protein and RNA factors could shift the balance to render the expression ratio of the two transcript variants to a favorable outcome for the cell's current state. Generally, such a mechanism could serve to adapt the cell's needs by regulating the availability of transcripts that are distinct with respect to their 3'-UTR sequence-dependent cytoplasmic activities. These could for example include stability of the transcripts or potential binding sites for miRNAs that are distinct for the splicing variants with different 3'-UTRs (Figure 41). Future genome and transcriptome wide studies will show if the mechanism proposed here for the *CDV3* locus is a mechanism that is generally used to regulate expression of alternative transcripts.

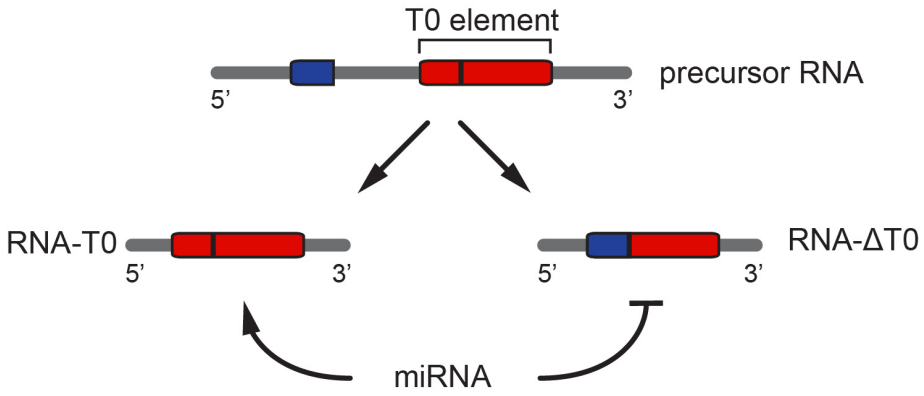


Figure 41: Shifting the balance of transcript isoforms to adapt the cell's needs by regulating the availability of transcripts with distinct cytoplasmic functions. For example transcript stability or binding sites for miRNAs could be distinct within the different isoforms.

6 Conclusion and perspectives

This work investigated a previously unknown nuclear function of the 3'-UTR of the *CDV3* transcript. A model is proposed in which a specific sequence element from this 3'-UTR affects the local level of H3K27me3 and thereby regulates alternative splicing of the *CDV3* transcript variants (Figure 40). This chromatin state-dependent feedback loop might regulate the 3'-UTR-dependent cellular stability of the *CDV3* transcript. Moreover, it might provide a mechanism to stably maintain a given ratio between transcript isoforms in the cell.

To further strengthen the speculative points in this model discussed above, future experiments could comprise the following: By systematically perturbing each of the involved splicing factors, their role in the hypothesized model could be elucidated. Furthermore, investigating the role of HNRNPK in rescuing EZH2 activity both *in vitro* and *in vivo* would be insightful. Finally, it would be interesting to gain insight into the dynamics of the process to be able to assess, on which time scale cells can adjust their splicing decision in response to external cues. While these steps of the model await further analyses, the model presents a new function of a 3'-UTR. It demonstrated a link between the previously described involvement of RNA in the recruitment and activity of the PCR2 complex and the cross talk between alternative splicing and the local chromatin signature. It is proposed that the regulation of the pathway from DNA to protein via chromatin modifications, transcription, and alternative splicing is not restricted to specific

groups of RNAs as introduced at the beginning (Figure 1). Instead, a cross talk between the RNA-mediated regulatory mechanisms is inferred from the results obtained here. Genome- and transcriptome-wide studies will show whether the proposed mechanism is also used by other transcripts, which can vary in their last exons to modify their 3'-UTR-dependent cellular functions.

Understanding the complex network of alternative splicing regulation, EZH2 activity, H3K27me3 distribution, and the influence of RNA and in particular 3'-UTRs has a number of implications. It is important for the development of novel drugs for cancer therapies for the following reasons: First, the tight control of alternative splicing is deregulated in many cancers so that aberrant proteins resulting in aggressive and invasive cancer phenotypes are produced (Oltean & Bates, 2014).

Second, EZH2 itself has been shown to be mutated or deregulated in many cancer types. Interestingly, mutations leading to the depletion of EZH2 activity (Nikoloski et al, 2010) as well as an increase in EZH2 activity by mutations in the catalytically active domain (Sneeringer et al, 2010) or its overexpression (Asangani et al, 2013; Kleer et al, 2003; Varambally et al, 2008; Visser et al, 2001) have been identified. For this reason, drugs targeting EZH2 enzymatic activity are currently being developed and have been shown to be effective in influencing the global levels of H3K27me3 (McCabe et al, 2012). Future studies will show whether a potentially positive treatment outcome can only be attributed to changes in the H3K27me3 levels or whether the influence of inhibiting EZH2 is more far-reaching by also targeting alternative splicing as suggested here in this thesis.

Third, a mutation in the gene encoding for the histone H3 variant H3.3 resulting in an amino acid change at position 27 from a lysine to a methionine (H3.3K27M mutation) was identified in glioblastoma multiforme (Khuong-Quang et al, 2012; Schwartzentruber et al, 2012; Sturm et al, 2012; Wu et al, 2012). This finding is of particular interest here because tumors with this mutation show lower overall H3K27me3 levels due to the fact that H3.3K27M inhibits the enzymatic activity of EZH2 (Bender et al, 2013; Bjerke et al, 2013; Chan et al, 2013; Lewis et al, 2013).

By this mutation, another way of deregulating H3K27me3 levels exists in tumors in addition to mutations in EZH2.

Finally, the regulation of chromatin modifications that alter alternative splicing decisions might not be restricted to EZH2 itself but could potentially be expanded to further chromatin modifying enzymes. Genome-sequencing studies have reported on a large variety of mutations in genes encoding for chromatin-modifying proteins (Plass et al, 2013). Therefore, chromatin-modifying proteins are emerging as targets in cancer therapeutics and the first epigenetic-based therapies have already been approved, which involve the inhibition of histone deacetylases (Helin & Dhanak, 2013; Khan & La Thangue, 2012; Wagner et al, 2010). Most studies that investigate the potential of targeting chromatin-modifying proteins disregard the effect of those drugs on alternative splicing. However, it could turn out that their positive influence on cancer progression can, at least in parts, be attributed to their effect on alternative splicing.

In conclusion, further unraveling the connections between 3'-UTRs, alternative splicing and RNA-mediated chromatin modifications will have far-reaching implications for our understanding of how cells manage to stably control the abundance of certain transcript variants. Thus, understanding the mechanisms that link alternative splicing and alterations in chromatin modifications in healthy and cancer cells will significantly broaden the possibilities for therapeutic intervention when those processes are deregulated.

References

Abuli A, Fernandez-Rozadilla C, Giraldez MD, Munoz J, Gonzalo V, Bessa X, Bujanda L, Rene JM, Lanas A, Garcia AM, Salo J, Arguello L, Vilella A, Carreno R, Jover R, Xicola RM, Llor X, Carvajal-Carmona L, Tomlinson IP, Kerr DJ, Houlston RS, Pique JM, Carracedo A, Castells A, Andreu M, Ruiz-Ponte C, Castellvi-Bel S, Gastrointestinal Oncology Group of the Spanish Gastroenterological A (2011) A two-phase case-control study for colorectal cancer genetic susceptibility: candidate genes from chromosomal regions 9q22 and 3q22. *Br J Cancer* **105:** 870-875

Ainger K, Avossa D, Morgan F, Hill SJ, Barry C, Barbarese E, Carson JH (1993) Transport and localization of exogenous myelin basic protein mRNA microinjected into oligodendrocytes. *J Cell Biol* **123:** 431-441

Alló M, Buggiano V, Fededa JP, Petrillo E, Schor I, de la Mata M, Agirre E, Plass M, Eyras E, Elela SA, Klinck R, Chabot B, Kornblihtt AR (2009) Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat Struct Mol Biol* **16:** 717-724

Andersson R, Enroth S, Rada-Iglesias A, Wadelius C, Komorowski J (2009) Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res* **19:** 1732-1741

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

Arnoult N, Van Beneden A, Decottignies A (2012) Telomere length regulates TERRA levels through increased trimethylation of telomeric H3K9 and HP1alpha. *Nat Struct Mol Biol* **19:** 948-956

Arora A, Sunbul M, Jaschke A (2015) Dual-colour imaging of RNAs using quencher- and fluorophore-binding aptamers. *Nucleic Acids Res*

Asangani IA, Ateeq B, Cao Q, Dodson L, Pandhi M, Kunju LP, Mehra R, Lonigro RJ, Siddiqui J, Palanisamy N, Wu YM, Cao X, Kim JH, Zhao M, Qin ZS, Iyer MK, Maher CA, Kumar-Sinha C, Varambally S, Chinnaiyan AM (2013) Characterization of the EZH2-MMSET histone methyltransferase regulatory axis in cancer. *Mol Cell* **49:** 80-93

Audas TE, Jacob MD, Lee S (2012) Immobilization of Proteins in the Nucleolus by Ribosomal Intergenic Spacer Noncoding RNA. *Molecular Cell* **45:** 147-157

Baltimore D (2001) Our genome unveiled. *Nature* **409:** 814-816

Baltz AG, Munschauer M, Schwanhausser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, Wyler E, Bonneau R, Selbach M, Dieterich C, Landthaler M (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell* **46:** 674-690

Barak O, Lazzaro MA, Cooch NS, Picketts DJ, Shiekhattar R (2004) A tissue-specific, naturally occurring human SNF2L variant inactivates chromatin remodeling. *Journal of Biological Chemistry* **279:** 45130-45138

Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, Blencowe BJ, Frey BJ (2010) Deciphering the splicing code. *Nature* **465:** 53-59

Baron-Benhamou J, Gehring NH, Kulozik AE, Hentze MW (2004) Using the lambdaN peptide to tether proteins to RNAs. *Methods Mol Biol* **257:** 135-154

Barrett LW, Fletcher S, Wilton SD (2012) Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell Mol Life Sci* **69:** 3613-3634

Bashirullah A, Cooperstock RL, Lipshitz HD (2001) Spatial and temporal control of RNA stability. *Proc Natl Acad Sci U S A* **98:** 7025-7028

Batista PJ, Chang HY (2013) Long noncoding RNAs: cellular address codes in development and disease. *Cell* **152:** 1298-1307

Belmont AS, Straight AF (1998) In vivo visualization of chromosomes using lac operator-repressor binding. *Trends in Cell Biology* **8:** 121-124

Bender S, Tang Y, Lindroth AM, Hovestadt V, Jones DT, Kool M, Zapatka M, Northcott PA, Sturm D, Wang W, Radlwimmer B, Hojfeldt JW, Truffaux N, Castel D, Schubert S, Ryzhova M, Seker-Cin H, Gronych J, Johann PD, Stark S, Meyer J, Milde T, Schuhmann M, Ebinger M, Monoranu CM, Ponnuswami A, Chen S, Jones C, Witt O, Collins VP, von Deimling A, Jabado N, Puget S, Grill J, Helin K, Korshunov A, Lichter P, Monje M, Plass C, Cho YJ, Pfister SM (2013) Reduced H3K27me3 and DNA hypomethylation are major drivers of gene expression in K27M mutant pediatric high-grade gliomas. *Cancer Cell* **24:** 660-672

Bergmann JH, Spector DL (2014) Long non-coding RNAs: modulators of nuclear structure and function. *Curr Opin Cell Biol* **26:** 10-18

Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM (1998) Localization of ASH1 mRNA particles in living yeast. *Molecular cell* **2:** 437-445

Bjerke L, Mackay A, Nandhabalan M, Burford A, Jury A, Popov S, Bax DA, Carvalho D, Taylor KR, Vinci M, Bajrami I, McGonnell IM, Lord CJ, Reis RM, Hargrave D, Ashworth A, Workman P, Jones C (2013) Histone H3.3. mutations drive pediatric glioblastoma through upregulation of MYCN. *Cancer Discov* **3:** 512-519

Black DL (2000) Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. *Cell* **103:** 367-370

Black DL (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72:** 291-336

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

Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK, Bell GW, Otte AP, Vidal M, Gifford DK, Young RA, Jaenisch R (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441:** 349-353

Bracken AP, Dietrich N, Pasini D, Hansen KH, Helin K (2006) Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev* **20:** 1123-1136

Brockdorff N (2013) Noncoding RNA and Polycomb recruitment. *RNA* **19:** 429-442

Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, Swift S, Rastan S (1992) The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **71:** 515-526

Busch H, Muramatsu M, Adams H, Steele WJ, Liau MC, Smetana K (1963) Isolation of Nucleoli. *Exp Cell Res* **24:** SUPPL9:150-163

Butter F, Scheibe M, Morl M, Mann M (2009) Unbiased RNA-protein interaction screen by quantitative proteomics. *Proc Natl Acad Sci U S A* **106:** 10626-10631

Cao R, Wang H, He J, Erdjument-Bromage H, Tempst P, Zhang Y (2008) Role of hPHF1 in H3K27 methylation and Hox gene silencing. *Mol Cell Biol* **28:** 1862-1872

Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298:** 1039-1043

Cao R, Zhang Y (2004) SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol Cell* **15:** 57-67

Casanova M, Preissner T, Cerase A, Poot R, Yamada D, Li X, Appanah R, Bezstarosti K, Demmers J, Koseki H, Brockdorff N (2011) Polycomblike 2 facilitates the recruitment of PRC2 Polycomb group complexes to the inactive X chromosome and to target loci in embryonic stem cells. *Development* **138:** 1471-1482

Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, Krijgsveld J, Hentze MW (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149:** 1393-1406

Caudron-Herger M, Muller-Ott K, Mallm JP, Marth C, Schmidt U, Fejes-Toth K, Rippe K (2011) Coding RNAs with a non-coding function: maintenance of open chromatin structure. *Nucleus* **2:** 410-424

Caudron-Herger M, Pankert T, Seiler J, Nemeth A, Voit R, Grummt I, Rippe K (2015) Alu element-containing RNAs maintain nucleolar structure and function. *EMBO J*

Caudron-Herger M, Rippe K (2012) Nuclear architecture by RNA. *Curr Opin Genet Dev* **22:** 179-187

Celotto AM, Graveley BR (2001) Alternative splicing of the Drosophila Dscam pre-mRNA is both temporally and spatially regulated. *Genetics* **159:** 599-608

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

Chandra T, Kirschner K, Thuret JY, Pope BD, Ryba T, Newman S, Ahmed K, Samarajiwa SA, Salama R, Carroll T, Stark R, Janky R, Narita M, Xue L, Chicas A, Nunez S, Janknecht R, Hayashi-Takanaka Y, Wilson MD, Marshall A, Odom DT, Babu MM, Bazett-Jones DP, Tavare S, Edwards PA, Lowe SW, Kimura H, Gilbert DM, Narita M (2012) Independence of repressive histone marks and chromatin compaction during senescent heterochromatic layer formation. *Mol Cell* **47:** 203-214

Chase A, Cross NC (2011) Aberrations of EZH2 in cancer. *Clin Cancer Res* **17:** 2613-2618

Chekanova JA, Belostotsky DA (2003) Evidence that poly(A) binding protein has an evolutionarily conserved function in facilitating mRNA biogenesis and export. *RNA* **9:** 1476-1490

Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, Park J, Blackburn EH, Weissman JS, Qi LS, Huang B (2013) Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155:** 1479-1491

Chen M, Manley JL (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol* **10:** 741-754

Chou MY, Underwood JG, Nikolic J, Luu MH, Black DL (2000) Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing. *Mol Cell* **5:** 949-957

Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, Magnuson T, Heard E, Chang HY (2015) Systematic discovery of Xist RNA binding proteins. *Cell* **161:** 404-416

Chuang C-H, Belmont AS (2007) Moving chromatin within the interphase nucleus-controlled transitions? *Seminars in cell & developmental biology* **18:** 698-706

Chung I, Leonhardt H, Rippe K (2011) De novo assembly of a PML nuclear subcompartment occurs through multiple pathways and induces telomere elongation. *J Cell Sci* **124:** 3603-3618

Cifuentes-Rojas C, Hernandez AJ, Sarma K, Lee JT (2014) Regulatory interactions between RNA and polycomb repressive complex 2. *Mol Cell* **55:** 171-185

Cloos PA, Christensen J, Agger K, Helin K (2008) Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. *Genes Dev* **22:** 1115-1140

Coller J, Wickens M (2002) Tethered function assays using 3' untranslated regions. *Methods* **26:** 142-150

da Rocha ST, Boeva V, Escamilla-Del-Arenal M, Ancelin K, Granier C, Matias NR, Sanulli S, Chow J, Schulz E, Picard C, Kaneko S, Helin K, Reinberg D, Stewart AF, Wutz A, Margueron R, Heard E (2014) Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Mol Cell* **53:** 301-316

Daigle N, Ellenberg J (2007) LambdaN-GFP: an RNA reporter system for live-cell imaging. *Nature methods* **4:** 633-636

David CJ, Manley JL (2010) Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev* **24:** 2343-2364

Davidovich C, Goodrich KJ, Gooding AR, Cech TR (2014) A dimeric state for PRC2. *Nucleic Acids Res* **42:** 9236-9248

Davidovich C, Wang X, Cifuentes-Rojas C, Goodrich KJ, Gooding AR, Lee JT, Cech TR (2015) Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. *Mol Cell* **57:** 552-558

Davidovich C, Zheng L, Goodrich KJ, Cech TR (2013) Promiscuous RNA binding by Polycomb repressive complex 2. *Nat Struct Mol Biol* **20:** 1250-1257

Dawson MA, Kouzarides T (2012) Cancer epigenetics: from mechanism to therapy. *Cell* **150:** 12-27

De Gregorio E, Preiss T, Hentze MW (1999) Translation driven by an eIF4G core domain in vivo. *EMBO J* **18:** 4865-4874

de la Mata M, Alonso CR, Kadener S, Fededa JP, Blaustein M, Pelisch F, Cramer P, Bentley D, Kornblihtt AR (2003) A slow RNA polymerase II affects alternative splicing in vivo. *Molecular Cell* **12:** 525-532

Deng Z, Norseen J, Wiedmer A, Riethman H, Lieberman PM (2009) TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Mol Cell* **35:** 403-413

Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4:** P3

Du H, Rosbash M (2002) The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. *Nature* **419:** 86-90

Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander ES, Plath K, Guttman M (2013) The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341:** 1237973

Farazi TA, Spitzer JI, Morozov P, Tuschl T (2011) miRNAs in human cancer. *The Journal of pathology* **223:** 102-115

Faustino NA, Cooper TA (2003) Pre-mRNA splicing and human disease. *Genes Dev* **17:** 419-437

Fischle W, Tseng BS, Dormann HL, Ueberheide BM, Garcia BA, Shabanowitz J, Hunt DF, Funabiki H, Allis CD (2005) Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* **438:** 1116-1122

Fisher AG (2002) Cellular identity and lineage choice. *Nat Rev Immunol* **2:** 977-982

Francis NJ, Kingston RE, Woodcock CL (2004) Chromatin compaction by a polycomb group protein complex. *Science* **306:** 1574-1577

Fraser P, Bickmore W (2007) Nuclear organization of the genome and the potential for gene regulation. *Nature* **447:** 413-417

Friedman RC, Farh KK, Burge CB, Bartel DP (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19:** 92-105

Fu XD (2004) Towards a splicing code. *Cell* **119:** 736-738

Fu XD, Ares M, Jr. (2014) Context-dependent control of alternative splicing by RNA-binding proteins. *Nat Rev Genet* **15:** 689-701

Fukumaru S, Horiuchi M, Kobayashi K, Jalil MA, Iijima M, Masuda M, Begum L, Higashi M, Wakana S, Kanzaki T, Saheki T (2002) Novel mRNA molecules are induced in hypertrophied ventricles of carnitine-deficient mice and belong to a family of up-regulated gene in cells overexpressing c-erbB-2. *Biochim Biophys Acta* **1577:** 437-444

Furuno M, Pang KC, Ninomiya N, Fukuda S, Frith MC, Bult C, Kai C, Kawai J, Carninci P, Hayashizaki Y, Mattick JS, Suzuki H (2006) Clusters of internally primed transcripts reveal novel long noncoding RNAs. *PLoS Genet* **2:** e37

Gebauer F, Hentze MW (2004) Molecular mechanisms of translational control. *Nat Rev Mol Cell Biol* **5:** 827-835

Gonzalez I, Munita R, Agirre E, Dittmer TA, Gysling K, Misteli T, Luco RF (2015) A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature. *Nat Struct Mol Biol* **22:** 370-376

Goodier JL, Mandal PK, Zhang L, Kazazian HH, Jr. (2010) Discrete subcellular partitioning of human retrotransposon RNAs despite a common mechanism of genome insertion. *Hum Mol Genet* **19:** 1712-1725

Görisch SM, Wachsmuth M, Fejes Tóth K, Lichter P, Rippe K (2005) Histone acetylation increases chromatin accessibility. *Journal of Cell Science* **118:** 5825-5834

Granick D (1975) Nucleolar necklaces in chick embryo fibroblast cells. II. Microscope observations of the effect of adenosine analogues on nucleolar necklace formation. *J Cell Biol* **65:** 418-427

Graveley BR (2001) Alternative splicing: increasing diversity in the proteomic world. *Trends Genet* **17:** 100-107

Grewal SI, Jia S (2007) Heterochromatin revisited. *Nat Rev Genet* **8:** 35-46

Grzybowski AT, Chen Z, Ruthenburg AJ (2015) Calibrating ChIP-Seq with Nucleosomal Internal Standards to Measure Histone Modification Density Genome Wide. *Mol Cell* **58:** 886-899

Guil S, Soler M, Portela A, Carrere J, Fonalleras E, Gomez A, Villanueva A, Esteller M (2012) Intronic RNAs mediate EZH2 regulation of epigenetic targets. *Nat Struct Mol Biol* **19:** 664-670

Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, Sukumar S, Chang HY (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464:** 1071-1076

Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, Yang X, Amit I, Meissner A, Regev A, Rinn JL, Root DE, Lander ES (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477:** 295-300

Haaf T, Ward DC (1996) Inhibition of RNA polymerase II transcription causes chromatin decondensation, loss of nucleolar structure, and dispersion of chromosomal domains. *Exp Cell Res* **224:** 163-173

Haque N, Oberdoerffer S (2014) Chromatin and splicing. *Methods Mol Biol* **1126:** 97-113

Helin K, Dhanak D (2013) Chromatin proteins and modifications as drug targets. *Nature* **502:** 480-488

Hong S, Cho YW, Yu LR, Yu H, Veenstra TD, Ge K (2007) Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases. *Proc Natl Acad Sci U S A* **104:** 18439-18444

Hsieh J, Gage FH (2004) Epigenetic control of neural stem cell fate. *Curr Opin Genet Dev* **14:** 461-469

Huang H, Sabari BR, Garcia BA, Allis CD, Zhao Y (2014) SnapShot: histone modifications. *Cell* **159:** 458-458 e451

Huang RC, Bonner J (1965) Histone-bound RNA, a component of native nucleohistone. *Proc Natl Acad Sci USA* **54:** 960-967

Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, Khalil AM, Zuk O, Amit I, Rabani M, Attardi LD, Regev A, Lander ES, Jacks T, Rinn JL (2010) A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142:** 409-419

Hughes TA (2006) Regulation of gene expression by alternative untranslated regions. *Trends Genet* **22:** 119-122

Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, Horlings HM, Shah N, Umbricht C, Wang P, Wang Y, Kong B, Langerød A, Børresen-Dale A-L, Kim SK, van de Vijver M, Sukumar S, Whitfield ML, Kellis M, Xiong Y, Wong DJ, Chang HY (2011) Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet* **43:** 621-629

International Human Genome Sequencing C, Adekoya E, Ait-Zahra M, Allen N, Anderson M, Anderson S, Anufriev F, Ambruster J, Ayele K, Baker J, Baldwin J, Barna N, Bastien V, Batzoglou S, Beckerly R, Beda F, Bernard J, Birren B, Blumensteil B, Boguslavsky L et al (2001) Initial sequencing and analysis of the human genome. *Nature* **409:** 860-921

Izquierdo JM, Majos N, Bonnal S, Martinez C, Castelo R, Guigo R, Bilbao D, Valcarcel J (2005) Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. *Mol Cell* **19:** 475-484

Jacob MD, Audas TE, Mullineux S-T, Lee S (2012) Where no RNA polymerase has gone before: Novel functional transcripts derived from the ribosomal intergenic spacer. *Nucleus* **3:** 315-319

Jacob MD, Audas TE, Uniacke J, Trinkle-Mulcahy L, Lee S (2013) Environmental Cues Induce a Long Noncoding RNA-dependent Remodeling of the Nucleolus. *Mol Biol Cell* **24:** 2943-2953

Jacob Y, Feng S, LeBlanc CA, Bernatavichute YV, Stroud H, Cokus S, Johnson LM, Pellegrini M, Jacobsen SE, Michaels SD (2009) ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. *Nat Struct Mol Biol* **16:** 763-768

Jacobs SA, Khorasanizadeh S (2002) Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295:** 2080-2083

Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33 Suppl:** 245-254

Jansen RP (2001) mRNA localization: message on the move. *Nat Rev Mol Cell Biol* **2:** 247-256

Jegou T, Chung I, Heuvelman G, Wachsmuth M, Gorisch SM, Greulich-Bode KM, Boukamp P, Lichter P, Rippe K (2009) Dynamics of telomeres and promyelocytic leukemia nuclear bodies in a telomerase-negative human cell line. *Mol Biol Cell* **20:** 2070-2082

Jenuwein T, Allis CD (2001) Translating the histone code. *Science* **293:** 1074-1080.

Jeon Y, Lee JT (2011) YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* **146:** 119-133

Johansson HE, Dertinger D, LeCuyer KA, Behlen LS, Greef CH, Uhlenbeck OC (1998) A thermodynamic analysis of the sequence-specific binding of RNA by bacteriophage MS2 coat protein. *Proc Natl Acad Sci U S A* **95:** 9244-9249

Kalb R, Latwiel S, Baymaz HI, Jansen PW, Muller CW, Vermeulen M, Muller J (2014) Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat Struct Mol Biol* **21:** 569-571

Kaneko S, Bonasio R, Saldana-Meyer R, Yoshida T, Son J, Nishino K, Umezawa A, Reinberg D (2014a) Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Mol Cell* **53:** 290-300

Kaneko S, Li G, Son J, Xu CF, Margueron R, Neubert TA, Reinberg D (2010) Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. *Genes & Development* **24:** 2615-2620

Kaneko S, Son J, Bonasio R, Shen SS, Reinberg D (2014b) Nascent RNA interaction keeps PRC2 activity poised and in check. *Genes Dev* **28:** 1983-1988

Kaneko S, Son J, Shen SS, Reinberg D, Bonasio R (2013) PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat Struct Mol Biol* **20:** 1258-1264

Kanhere A, Viiri K, Araújo CC, Rasaiyaah J, Bouwman RD, Whyte WA, Pereira CF, Brookes E, Walker K, Bell GW, Pombo A, Fisher AG, Young RA, Jenner RG (2010) Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Molecular Cell* **38:** 675-688

Karapetyan AR, Buiting C, Kuiper RA, Coolen MW (2013) Regulatory Roles for Long ncRNA and mRNA. *Cancers (Basel)* **5:** 462-490

Keryer-Bibens C, Barreau C, Osborne HB (2008) Tethering of proteins to RNAs by bacteriophage proteins. *Biol Cell* **100:** 125-138

Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, Regev A, Lander ES, Rinn JL (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences of the USA* **106:** 11667-11672

Khan O, La Thangue NB (2012) HDAC inhibitors in cancer biology: emerging mechanisms and clinical applications. *Immunol Cell Biol* **90:** 85-94

Khare SP, Habib F, Sharma R, Gadewal N, Gupta S, Galande S (2012) HIstome--a relational knowledgebase of human histone proteins and histone modifying enzymes. *Nucleic Acids Res* **40:** D337-342

Khuong-Quang D-A, Buczkowicz P, Rakopoulos P, Liu X-Y, Fontebasso AM, Bouffet E, Bartels U, Albrecht S, Schwartzentruber J, Letourneau L, Bourgey M, Bourque G, Montpetit A, Bourret G, Lepage P, Fleming A, Lichter P, Kool M, von Deimling A, Sturm D, Korshunov A, Faury D, Jones DT, Majewski J, Pfister SM, Jabado N, Hawkins C (2012) K27M mutation in histone H3.3 defines clinically and biologically distinct subgroups of pediatric diffuse intrinsic pontine gliomas. *Acta neuropathologica* **124:** 439-447

Kim H, Kang K, Kim J (2009) AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. *Nucleic Acids Res* **37:** 2940-2950

King HA, Cobbold LC, Pichon X, Poyry T, Wilson LA, Booden H, Jukes-Jones R, Cain K, Lilley KS, Bushell M, Willis AE (2014) Remodelling of a polypyrimidine tract-binding protein complex during apoptosis activates cellular IRESs. *Cell Death Differ* **21:** 161-171

Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP (2010) Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci Signal* **3:** ra8

Klattenhoff CA, Scheuermann JC, Surface LE, Bradley RK, Fields PA, Steinhauser ML, Ding H, Butty VL, Torrey L, Haas S, Abo R, Tabebordbar M, Lee RT, Burge CB, Boyer LA (2013) Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell* **152:** 570-583

Kleer CG, Cao Q, Varambally S, Shen R, Ota I, Tomlins SA, Ghosh D, Sewalt RG, Otte AP, Hayes DF, Sabel MS, Livant D, Weiss SJ, Rubin MA, Chinnaiyan AM (2003) EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci U S A* **100:** 11606-11611

Konig J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* **17:** 909-915

Kornblihtt AR, Schor IE, Allo M, Dujardin G, Petrillo E, Munoz MJ (2013) Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol* **14:** 153-165

Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M, Xiong Y (2011) Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* **30:** 1956-1962

Kouzarides T (2007) Chromatin modifications and their function. *Cell* **128:** 693-705

Kretz M, Meister G (2014) RNA binding of PRC2: promiscuous or well ordered? *Mol Cell* **55:** 157-158

Kumar PP, Franklin S, Emechebe U, Hu H, Moore B, Lehman C, Yandell M, Moon AM (2014) TBX3 regulates splicing in vivo: a novel molecular mechanism for Ulnar-mammary syndrome. *PLoS Genet* **10:** e1004247

Kuzmichev A, Jenuwein T, Tempst P, Reinberg D (2004) Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. *Mol Cell* **14:** 183-193

Lallena MJ, Chalmers KJ, Llamazares S, Lamond AI, Valcarcel J (2002) Splicing regulation at the second catalytic step by Sex-lethal involves 3' splice site recognition by SPF45. *Cell* **109:** 285-296

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25

Lazzaro MA, Todd MA, Lavigne P, Vallee D, De Maria A, Picketts DJ (2008) Characterization of novel isoforms and evaluation of SNF2L/SMARCA1 as a candidate gene for X-linked mental retardation in 12 families linked to Xq25-26. *BMC Med Genet* **9:** 11

Le Guiner C, Lejeune F, Galiana D, Kister L, Breathnach R, Stevenin J, Del Gatto-Konczak F (2001) TIA-1 and TIAR activate splicing of alternative exons with weak 5' splice sites followed by a U-rich stretch on their own pre-mRNAs. *Journal of Biological Chemistry* **276:** 40638-40646

Lee N, Moss WN, Yario TA, Steitz JA (2015) EBV noncoding RNA binds nascent RNA to drive host PAX5 to viral DNA. *Cell* **160:** 607-618

Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K, Koseki H, Fuchikami T, Abe K, Murray HL, Zucker JP, Yuan B, Bell GW, Herbolsheimer E, Hannett NM, Sun K, Odom DT, Otte AP, Volkert TL, Bartel DP, Melton DA, Gifford DK, Jaenisch R, Young RA (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125:** 301-313

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

Leppek K, Stoecklin G (2014) An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins. *Nucleic Acids Res* **42:** e13

Lewis PW, Muller MM, Koletsky MS, Cordero F, Lin S, Banaszynski LA, Garcia BA, Muir TW, Becher OJ, Allis CD (2013) Inhibition of PRC2 Activity by a Gain-of-Function H3 Mutation Found in Pediatric Glioblastoma. *Science*

Li G, Margueron R, Ku M, Chambon P, Bernstein BE, Reinberg D (2010) Jarid2 and PRC2, partners in regulating gene expression. *Genes & Development* **24:** 368-380

Liang L, Astruc D (2011) The copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC) “click” reaction and its applications. An overview. *Coordination Chemistry Reviews* **255:** 2933-2945

Licatalosi DD, Yano M, Fak JJ, Mele A, Grabinski SE, Zhang C, Darnell RB (2012) Ptbp2 represses adult-specific splicing to regulate the generation of neuronal precursors in the embryonic brain. *Genes Dev* **26:** 1626-1642

Lim F, Downey TP, Peabody DS (2001) Translational repression and specific RNA binding by the coat protein of the Pseudomonas phage PP7. *Journal of Biological Chemistry* **276:** 22507-22513

Lin CH, Patton JG (1995) Regulation of alternative 3' splice site selection by constitutive splicing factors. *RNA* **1:** 234-245

Listerman I, Sapra AK, Neugebauer KM (2006) Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat Struct Mol Biol* **13:** 815-822

Liu X, Jiang Q, Mansfield SG, Puttaraju M, Zhang Y, Zhou W, Cohn JA, Garcia-Blanco MA, Mitchell LG, Engelhardt JF (2002) Partial correction of endogenous DeltaF508 CFTR in human cystic fibrosis airway epithelia by spliceosome-mediated RNA trans-splicing. *Nat Biotechnol* **20:** 47-52

Liu ZR (2002) p68 RNA helicase is an essential human splicing factor that acts at the U1 snRNA-5' splice site duplex. *Mol Cell Biol* **22:** 5443-5450

Lleres D, James J, Swift S, Norman DG, Lamond AI (2009) Quantitative analysis of chromatin compaction in living cells using FLIM-FRET. *The Journal of cell biology* **187:** 481-496

Luco RF, Misteli T (2011) More than a splicing code: integrating the role of RNA, chromatin and non-coding RNA in alternative splicing regulation. *Curr Opin Genet Dev* **21:** 366-372

Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T (2010) Regulation of alternative splicing by histone modifications. *Science* **327:** 996-1000

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389:** 251-260

Luijsterburg MS, Lindh M, Acs K, Vrouwe MG, Pines A, van Attikum H, Mullenders LH, Dantuma NP (2012) DDB2 promotes chromatin decondensation at UV-induced DNA damage. *J Cell Biol* **197:** 267-281

Luo MJ, Reed R (1999) Splicing is required for rapid and efficient mRNA export in metazoans. *Proc Natl Acad Sci U S A* **96:** 14937-14942

Mandel CR, Bai Y, Tong L (2008) Protein factors in pre-mRNA 3'-end processing. *Cell Mol Life Sci* **65:** 1099-1122

Maniatis T, Reed R (1987) The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. *Nature* **325:** 673-678

Mao YS, Sunwoo H, Zhang B, Spector DL (2011) Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nature Cell Biology* **13:** 95-101

Margueron R, Justin N, Ohno K, Sharpe ML, Son J, Drury WJ, 3rd, Voigt P, Martin SR, Taylor WR, De Marco V, Pirrotta V, Reinberg D, Gamblin SJ (2009) Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* **461:** 762-767

Margueron R, Li G, Sarma K, Blais A, Zavadil J, Woodcock CL, Dynlacht BD, Reinberg D (2008) Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol Cell* **32:** 503-518

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

Marino-Ramirez L, Levine KM, Morales M, Zhang S, Moreland RT, Baxevanis AD, Landsman D (2011) The Histone Database: an integrated resource for histones and histone fold-containing proteins. *Database (Oxford)* **2011:** bar048

Mateescu B, England P, Halgand F, Yaniv M, Muchardt C (2004) Tethering of HP1 proteins to chromatin is relieved by phosphoacetylation of histone H3. *EMBO reports* **5:** 490-496

Matoulkova E, Michalova E, Vojtesek B, Hrstka R (2012) The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. *RNA Biol* **9:** 563-576

McCabe MT, Ott HM, Ganji G, Korenchuk S, Thompson C, Van Aller GS, Liu Y, Graves AP, Della Pietra A, 3rd, Diaz E, LaFrance LV, Mellinger M, Duquenne C, Tian X, Kruger RG, McHugh CF, Brandt M, Miller WH, Dhanak D, Verma SK, Tummino PJ, Creasy CL (2012) EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* **492:** 108-112

McHugh CA, Chen CK, Chow A, Surka CF, Tran C, McDonel P, Pandya-Jones A, Blanco M, Burghard C, Moradian A, Sweredoski MJ, Shishkin AA, Su J, Lander ES, Hess S, Plath K, Guttman M (2015) The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521:** 232-236

Mercer TR, Mattick JS (2013) Understanding the regulatory and transcriptional complexity of the genome through structure. *Genome Res* **23:** 1081-1088

Mercer TR, Wilhelm D, Dinger ME, Solda G, Korbie DJ, Glazov EA, Truong V, Schwenke M, Simons C, Matthaei KI, Saint R, Koopman P, Mattick JS (2011) Expression of distinct RNAs from 3' untranslated regions. *Nucleic Acids Res* **39:** 2393-2403

Mignone F, Gissi C, Liuni S, Pesole G (2002) Untranslated regions of mRNAs. *Genome Biol* **3:** REVIEWS0004

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448:** 553-560

Minajigi A, Froberg JE, Wei C, Sunwoo H, Kesner B, Colognori D, Lessing D, Payer B, Boukhali M, Haas W, Lee JT (2015) Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349**

Miyanari Y, Ziegler-Birling C, Torres-Padilla ME (2013) Live visualization of chromatin dynamics with fluorescent TALEs. *Nature structural & molecular biology* **20:** 1321-1324

Molitor J, Mallm J-P, Rippe K, Erdel F (2015) Dissecting genome-wide chromatin patterns from deep sequencing data with correlation functions. *Mol Syst Biol***:** submitted

Muller-Ott K, Erdel F, Matveeva A, Mallm JP, Rademacher A, Hahn M, Bauer C, Zhang Q, Kaltofen S, Schotta G, Hofer T, Rippe K (2014) Specificity, propagation, and memory of pericentric heterochromatin. *Mol Syst Biol* **10:** 746

Naughton C, Sproul D, Hamilton C, Gilbert N (2010) Analysis of active and inactive X chromosome architecture reveals the independent organization of 30 nm and large-scale chromatin structures. *Mol Cell* **40:** 397-409

Nayak V, Xu C, Min J (2011) Composition, recruitment and regulation of the PRC2 complex. *Nucleus* **2:** 277-282

Niazi F, Valadkhan S (2012) Computational analysis of functional long noncoding RNAs reveals lack of peptide-coding capacity and parallels with 3' UTRs. *RNA* **18:** 825-843

Nikoloski G, Langemeijer SM, Kuiper RP, Knops R, Massop M, Tonnissen ER, van der Heijden A, Scheele TN, Vandenberghe P, de Witte T, van der Reijden BA, Jansen JH (2010) Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. *Nat Genet* **42:** 665-667

Nozawa RS, Nagao K, Igami KT, Shibata S, Shirai N, Nozaki N, Sado T, Kimura H, Obuse C (2013) Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. *Nat Struct Mol Biol* **20:** 566-573

Oh JJ, Grosshans DR, Wong SG, Slamon DJ (1999) Identification of differentially expressed genes associated with HER-2/neu overexpression in human breast cancer cells. *Nucleic Acids Res* **27:** 4008-4017

Oltean S, Bates DO (2014) Hallmarks of alternative splicing in cancer. *Oncogene* **33:** 5311-5318

Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1:** 376-386

Ozawa T, Natori Y, Sato M, Umezawa Y (2007) Imaging dynamics of endogenous mitochondrial RNA in single living cells. *Nat Methods* **4:** 413-419

Paige JS, Wu KY, Jaffrey SR (2011) RNA mimics of green fluorescent protein. *Science* **333:** 642-646

Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* **32:** 232-246

Pasini D, Cloos PA, Walfridsson J, Olsson L, Bukowski JP, Johansen JV, Bak M, Tommerup N, Rappsilber J, Helin K (2010) JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* **464:** 306-310

Paul J, Duerksen JD (1975) Chromatin-associated RNA content of heterochromatin and euchromatin. *Molecular and Cellular Biochemistry* **9:** 9-16

Pederson T, Bhorjee JS (1979) Evidence for a role of RNA in eukaryotic chromosome structure. Metabolically stable, small nuclear RNA species are covalently linked to chromosomal DNA in HeLa cells. *J MOL BIOL* **128:** 451-480

Peng JC, Valouev A, Swigut T, Zhang J, Zhao Y, Sidow A, Wysocka J (2009) Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* **139:** 1290-1302

Pichon X, Wilson LA, Stoneley M, Bastide A, King HA, Somers J, Willis AE (2012) RNA binding protein/RNA element interactions and the control of translation. *Current protein & peptide science* **13:** 294-304

Plass C, Pfister SM, Lindroth AM, Bogatyrova O, Claus R, Lichter P (2013) Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. *Nature Reviews Genetics* **14:** 765-780

Proudnikov D, Mirzabekov A (1996) Chemical methods of DNA and RNA fluorescent labeling. *Nucleic Acids Res* **24:** 4535-4542

Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, Farrell CM, Hart J, Landrum MJ, McGarvey KM, Murphy MR, O'Leary NA, Pujar S, Rajput B, Rangwala SH, Riddick LD, Shkeda A, Sun H, Tamez P, Tully RE, Wallin C, Webb D, Weber J, Wu W, DiCuccio M, Kitts P, Maglott DR, Murphy TD, Ostell JM (2014) RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res* **42:** D756-763

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26:** 841-842

Rego A, Sinclair PB, Tao W, Kireev I, Belmont AS (2008) The facultative heterochromatin of the inactive X chromosome has a distinctive condensed ultrastructure. *J Cell Sci* **121:** 1119-1127

Reynolds N, Salmon-Divon M, Dvinge H, Hynes-Allen A, Balasooriya G, Leaford D, Behrens A, Bertone P, Hendrich B (2011) NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. *The EMBO Journal* **31:** 593-605

Riising EM, Comet I, Leblanc B, Wu X, Johansen JV, Helin K (2014) Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Mol Cell* **55:** 347-360

Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. *Annu Rev Biochem* **81:** 145-166

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129:** 1311-1323

Rippe K (2012) *Genome organization and function in the cell nucleus*, Weinheim: Wiley-VCH.

Rippe K, Mazurkiewicz J, Kepper N (2008) Interactions of histones with DNA: nucleosome assembly, stability and dynamics. In *DNA interactions with polymers and surfactants*, Dias RS, Lindman B (eds), pp 135-172. London: Wiley

Robinett CC, Straight A, Li G, Willhelm C, Sudlow G, Murray A, Belmont AS (1996) In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J Cell Biol* **135:** 1685-1700

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. *Nat Biotechnol* **29:** 24-26

Rossow KL, Janknecht R (2003) Synergism between p68 RNA helicase and the transcriptional coactivators CBP and p300. *Oncogene* **22:** 151-156

Rougeulle C, Chaumeil J, Sarma K, Allis CD, Reinberg D, Avner P, Heard E (2004) Differential histone H3 Lys-9 and Lys-27 methylation profiles on the X chromosome. *Mol Cell Biol* **24:** 5475-5484

Ruskin B, Zamore PD, Green MR (1988) A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* **52:** 207-219

Sarma K, Cifuentes-Rojas C, Ergun A, del Rosario A, Jeon Y, White F, Sadreyev R, Lee JT (2014) ATRX Directs Binding of PRC2 to Xist RNA and Polycomb Targets. *Cell* **159:** 869-883

Sarma K, Margueron R, Ivanov A, Pirrotta V, Reinberg D (2008) Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. *Mol Cell Biol* **28:** 2718-2731

Sato M, Koriyama M, Watanabe S, Ohtsuka M, Sakurai T, Inada E, Saitoh I, Nakamura S, Miyoshi K (2015) Direct Injection of CRISPR/Cas9-Related mRNA into Cytoplasm of Parthenogenetically Activated Porcine Oocytes Causes Frequent Mosaicism for Indel Mutations. *Int J Mol Sci* **16:** 17838-17856

Sauliere J, Sureau A, Expert-Bezancon A, Marie J (2006) The polypyrimidine tract binding protein (PTB) represses splicing of exon 6B from the beta-tropomyosin pre-mRNA by directly interfering with the binding of the U2AF65 subunit. *Mol Cell Biol* **26:** 8755-8769

Scharf AND, Barth TK, Imhof A (2009) Establishment of histone modifications after chromatin assembly. *Nucleic Acids Research* **37:** 5032-5040

Schmidt U, Basyuk E, Robert MC, Yoshida M, Villemin JP, Auboeuf D, Aitken S, Bertrand E (2011) Real-time imaging of cotranscriptional splicing reveals a kinetic model that reduces noise: implications for alternative splicing regulation. *J Cell Biol* **193:** 819-829

Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE, Zipursky SL (2000) Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101:** 671-684

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9:** 671-675

Schor IE, Fiszbein A, Petrillo E, Kornblihtt AR (2013) Intragenic epigenetic changes modulate NCAM alternative splicing in neuronal differentiation. *The EMBO Journal* **32:** 2264-2274

Schwartz S, Ast G (2010) Chromatin density and splicing destiny: on the cross-talk between chromatin structure and splicing. *EMBO J* **29:** 1629-1636

Schwartzentruber J, Korshunov A, Liu XY, Jones DT, Pfaff E, Jacob K, Sturm D, Fontebasso AM, Quang DA, Tonjes M, Hovestadt V, Albrecht S, Kool M, Nantel A, Konermann C, Lindroth A, Jager N, Rausch T, Ryzhova M, Korbel JO et al (2012) Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* **482:** 226-231

Shao C, Yang B, Wu T, Huang J, Tang P, Zhou Y, Zhou J, Qiu J, Jiang L, Li H, Chen G, Sun H, Zhang Y, Denise A, Zhang DE, Fu XD (2014) Mechanisms for U2AF to define 3' splice sites and regulate alternative splicing in the human genome. *Nat Struct Mol Biol* **21:** 997-1005

Sharma S, Falick AM, Black DL (2005) Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal E complex. *Mol Cell* **19:** 485-496

Sharma S, Kohlstaedt LA, Damianov A, Rio DC, Black DL (2008) Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nat Struct Mol Biol* **15:** 183-191

Shechter D, Dormann HL, Allis CD, Hake SB (2007) Extraction, purification and analysis of histones. *Nat Protoc* **2:** 1445-1457

Shen X, Kim W, Fujiwara Y, Simon MD, Liu Y, Mysliwiec MR, Yuan G-C, Lee Y, Orkin SH (2009) Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. *Cell* **139:** 1303-1314

Shevtsov SP, Dundr M (2011) Nucleation of nuclear bodies by RNA. *Nat Cell Biol* **13:** 167-173

Shogren-Knaak M, Ishii H, Sun J-M, Pazin MJ, Davie JR, Peterson CL (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311:** 844-847

Shukla S, Oberdoerffer S (2012) Co-transcriptional regulation of alternative pre-mRNA splicing. *Biochim Biophys Acta* **1819:** 673-683

Singh R, Valcarcel J, Green MR (1995) Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* **268:** 1173-1176

Smits AH, Jansen PW, Poser I, Hyman AA, Vermeulen M (2013) Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Res* **41:** e28

Sneeringer CJ, Scott MP, Kuntz KW, Knutson SK, Pollock RM, Richon VM, Copeland RA (2010) Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. *Proc Natl Acad Sci U S A* **107:** 20980-20985

Son J, Shen SS, Margueron R, Reinberg D (2013) Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. *Genes & Development* **27:** 2663-2677

Song JJ, Garlick JD, Kingston RE (2008) Structural basis of histone H4 recognition by p55. *Genes Dev* **22:** 1313-1318

Spellman R, Smith CW (2006) Novel modes of splicing repression by PTB. *Trends Biochem Sci* **31:** 73-76

Spies N, Nielsen CB, Padgett RA, Burge CB (2009) Biased chromatin signatures around polyadenylation sites and exons. *Mol Cell* **36:** 245-254

Spitale RC, Tsai MC, Chang HY (2011) RNA templating the epigenome: long noncoding RNAs as molecular scaffolds. *Epigenetics* **6:** 539-543

Stoecklin G, Muhlemann O (2013) RNA decay mechanisms: specificity through diversity. *Biochim Biophys Acta* **1829:** 487-490

Straight AF, Belmont AS, Robinett CC, Murray AW (1996) GFP tagging of budding yeast chromosomes reveals that protein-protein interactions can mediate sister chromatid cohesion. *Current Biology* **6:** 1599-1608

Sturm D, Witt H, Hovestadt V, Khuong-Quang DA, Jones DT, Konermann C, Pfaff E, Tonjes M, Sill M, Bender S, Kool M, Zapatka M, Becker N, Zucknick M, Hielscher T, Liu XY, Fontebasso AM, Ryzhova M, Albrecht S, Jacob K et al (2012) Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell* **22:** 425-437

Sullivan GJ, Bridger JM, Cuthbert AP, Newbold RF, Bickmore WA, McStay B (2001) Human acrocentric chromosomes with transcriptionally silent nucleolar organizer regions associate with nucleoli. *The EMBO Journal* **20:** 2867-2874

Suzuki T, Fujikura K, Higashiyama T, Takata K (1997) DNA staining for fluorescence and laser confocal microscopy. *J Histochem Cytochem* **45:** 49-53

Tachibana M, Sugimoto K, Fukushima T, Shinkai Y (2001) Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *The Journal of Biological Chemistry* **276:** 25309-25317

Tajrishi MM, Tuteja R, Tuteja N (2011) Nucleolin: The most abundant multifunctional phosphoprotein of nucleolus. *Communicative & integrative biology* **4:** 267-275

Tarn WY, Steitz JA (1997) Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem Sci* **22:** 132-137

Theurkauf WE, Hazelrigg TI (1998) In vivo analyses of cytoplasmic transport and cytoskeletal organization during Drosophila oogenesis: characterization of a multi-step anterior localization pathway. *Development* **125:** 3655-3666

Tilgner H, Knowles DG, Johnson R, Davis CA, Chakrabortty S, Djebali S, Curado J, Snyder M, Gingeras TR, Guigo R (2012) Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res* **22:** 1616-1625

Toth KF, Knoch TA, Wachsmuth M, Frank-Stohr M, Stohr M, Bacher CP, Muller G, Rippe K (2004) Trichostatin A-induced histone acetylation causes decondensation of interphase chromatin. *J Cell Sci* **117:** 4277-4287

Trojer P, Reinberg D (2007) Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell* **28:** 1-13

Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329:** 689-693

Tsuchiya K, Kawano Y, Kojima T, Nagata K, Takao T, Okada M, Shinohara H, Maki K, Toyama-Sorimachi N, Miyasaka N, Watanabe M, Karasuyama H (2003) Molecular cloning and characterization of TPP36 and its isoform TPP32, novel substrates of Abl tyrosine kinase. *FEBS Lett* **537:** 203-209

Tumbar T, Sudlow G, Belmont AS (1999) Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain. *Journal of Cell Biology* **145:** 1341-1354.

Tyagi S (2009) Imaging intracellular RNA distribution and dynamics in living cells. *Nat Methods* **6:** 331-338

Vakoc CR, Letting DL, Gheldof N, Sawado T, Bender MA, Groudine M, Weiss MJ, Dekker J, Blobel GA (2005) Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Molecular Cell* **17:** 453-462

Valcarcel J, Gaur RK, Singh R, Green MR (1996) Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science* **273:** 1706-1709

Valencia-Burton M, Broude NE (2007) Visualization of RNA using fluorescence complementation triggered by aptamer-protein interactions (RFAP) in live bacterial cells. *Curr Protoc Cell Biol* **Chapter 17:** Unit 17 11

Valencia-Burton M, McCullough RM, Cantor CR, Broude NE (2007) RNA visualization in live bacterial cells using fluorescent protein complementation. *Nat Methods* **4:** 421-427

Valencia-Sanchez MA, Liu J, Hannon GJ, Parker R (2006) Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes & Development* **20:** 515-524

Valk-Lingbeek ME, Bruggeman SW, van Lohuizen M (2004) Stem cells and cancer; the polycomb connection. *Cell* **118:** 409-418

van der Velden AW, Thomas AA (1999) The role of the 5' untranslated region of an mRNA in translation regulation during development. *International Journal of Biochemistry & Cell Biology* **31:** 87-106

van Holde KE (1989) *Chromatin*, Heidelberg: Springer.

Varambally S, Cao Q, Mani RS, Shankar S, Wang X, Ateeq B, Laxman B, Cao X, Jing X, Ramnarayanan K, Brenner JC, Yu J, Kim JH, Han B, Tan P, Kumar-Sinha C, Lonigro RJ, Palanisamy N, Maher CA, Chinnaiyan AM (2008) Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. *Science* **322:** 1695-1699

Verma SK, Tian X, LaFrance LV, Duquenne C, Suarez DP, Newlander KA, Romeril SP, Burgess JL, Grant SW, Brackley JA, Graves AP, Scherzer DA, Shu A, Thompson C, Ott HM, Aller GS, Machutta CA, Diaz E, Jiang Y, Johnson NW, Knight SD, Kruger RG, McCabe MT, Dhanak D, Tummino PJ, Creasy CL, Miller WH (2012) Identification of Potent, Selective, Cell-Active Inhibitors of the Histone Lysine Methyltransferase EZH2. *ACS Med Chem Lett* **3:** 1091-1096

Villa T, Pleiss JA, Guthrie C (2002) Spliceosomal snRNAs: Mg(2+)-dependent chemistry at the catalytic core? *Cell* **109:** 149-152

Villar-Garea A, Israel L, Imhof A (2008) Analysis of histone modifications by mass spectrometry. *Curr Protoc Protein Sci* **Chapter 14:** Unit 14 10

Visser HP, Gunster MJ, Kluin-Nelemans HC, Manders EM, Raaphorst FM, Meijer CJ, Willemze R, Otte AP (2001) The Polycomb group protein EZH2 is upregulated in proliferating, cultured human mantle cell lymphoma. *Br J Haematol* **112:** 950-958

Wagner JM, Hackanson B, Lubbert M, Jung M (2010) Histone deacetylase (HDAC) inhibitors in recent clinical trials for cancer therapy. *Clin Epigenetics* **1:** 117-136

Wahl MC, Will CL, Luhrmann R (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell* **136:** 701-718

Wakiyama M, Kaitsu Y, Muramatsu R, Takimoto K, Yokoyama S (2012) Tethering of proteins to RNAs using the bovine immunodeficiency virus-Tat peptide and BIV-TAR RNA. *Anal Biochem* **427:** 130-132

Wang X, Hayes JJ (2006) Physical methods used to study core histone tail structures and interactions in solution. *Biochem Cell Biol* **84:** 578-588

Wang Z, Kayikci M, Briese M, Zarnack K, Luscombe NM, Rot G, Zupan B, Curk T, Ule J (2010) iCLIP predicts the dual splicing effects of TIA-RNA interactions. *PLoS Biol* **8:** e1000530

Wilkie GS, Davis I (2001) Drosophila wingless and pair-rule transcripts localize apically by dynein-mediated transport of RNA particles. *Cell* **105:** 209-219

Wilson BJ, Bates GJ, Nicol SM, Gregory DJ, Perkins ND, Fuller-Pace FV (2004) The p68 and p72 DEAD box RNA helicases interact with HDAC1 and repress transcription in a promoter-specific manner. *BMC Mol Biol* **5:** 11

Wong LH, Brettingham-Moore KH, Chan L, Quach JM, Anderson MA, Northrop EL, Hannan R, Saffery R, Shaw ML, Williams E, Choo KH (2007) Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. *Genome Res* **17:** 1146-1160

Wu G, Broniscer A, McEachron TA, Lu C, Paugh BS, Becksfort J, Qu C, Ding L, Huether R, Parker M, Zhang J, Gajjar A, Dyer MA, Mullighan CG, Gilbertson RJ, Mardis ER, Wilson RK, Downing JR, Ellison DW, Zhang J, Baker SJ, St. Jude Children's Research Hospital-Washington University Pediatric Cancer Genome P (2012) Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nat Genet* **44:** 251-253

Wu H, Chen X, Xiong J, Li Y, Li H, Ding X, Liu S, Chen S, Gao S, Zhu B (2011) Histone methyltransferase G9a contributes to H3K27 methylation in vivo. *Cell Res* **21:** 365-367

Wu L, Murat P, Matak-Vinkovic D, Murrell A, Balasubramanian S (2013) The binding interaction between long non-coding RNA HOTAIR and PRC2 proteins. *Biochemistry*

Xu N, Chen CY, Shyu AB (1997) Modulation of the fate of cytoplasmic mRNA by AU-rich elements: key sequence features controlling mRNA deadenylation and decay. *Mol Cell Biol* **17:** 4611-4621

Xue Y, Zhou Y, Wu T, Zhu T, Ji X, Kwon YS, Zhang C, Yeo G, Black DL, Sun H, Fu XD, Zhang Y (2009) Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol Cell* **36:** 996-1006

Yang L, Lin C, Jin C, Yang JC, Tanasa B, Li W, Merkurjev D, Ohgi KA, Meng D, Zhang J, Evans CP, Rosenfeld MG (2013) lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. *Nature* **500:** 598-602

Yang L, Lin C, Liu W, Zhang J, Ohgi KA, Grinstein JD, Dorrestein PC, Rosenfeld MG (2011) ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell* **147:** 773-788

Yao H, Brick K, Evrard Y, Xiao T, Camerini-Otero RD, Felsenfeld G (2010) Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev* **24:** 2543-2555

Yap KL, Li S, Munoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, Gil J, Walsh MJ, Zhou MM (2010) Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell* **38:** 662-674

Yearim A, Gelfman S, Shayevitch R, Melcer S, Glaich O, Mallm JP, Nissim-Rafinia M, Cohen AH, Rippe K, Meshorer E, Ast G (2015) HP1 is involved in regulating the global impact of DNA methylation on alternative splicing. *Cell Rep* **10:** 1122-1134

Yuan S, Sun Z (2009) Microinjection of mRNA and morpholino antisense oligonucleotides in zebrafish embryos. *J Vis Exp*

Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JH, Noller HF (2001) Crystal structure of the ribosome at 5.5 A resolution. *Science* **292:** 883-896

Zamore PD, Patton JG, Green MR (1992) Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355:** 609-614

Zang C, Schones DE, Zeng C, Cui K, Zhao K, Peng W (2009) A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25:** 1952-1958

Zappulla DC, Cech TR (2006) RNA as a flexible scaffold for proteins: yeast telomerase and beyond. *Cold Spring Harb Symp Quant Biol* **71:** 217-224

Zarnack K, Konig J, Tajnik M, Martincorena I, Eustermann S, Stevant I, Reyes A, Anders S, Luscombe NM, Ule J (2013) Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* **152:** 453-466

Zee BM, Levin RS, Xu B, LeRoy G, Wingreen NS, Garcia BA (2010) In vivo residue-specific histone methylation dynamics. *The Journal of Biological Chemistry* **285:** 3341-3350

Zhang J, Manley JL (2013) Misregulation of pre-mRNA alternative splicing in cancer. *Cancer Discov* **3:** 1228-1237

Zhang P, Du J, Sun B, Dong X, Xu G, Zhou J, Huang Q, Liu Q, Hao Q, Ding J (2006) Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. *Nucleic Acids Res* **34:** 6621-6628

Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song JJ, Kingston RE, Borowsky M, Lee JT (2010) Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell* **40:** 939-953

Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322:** 750-756

Zhao W, Blagev D, Pollack JL, Erle DJ (2011) Toward a systematic understanding of mRNA 3' untranslated regions. *Proc Am Thorac Soc* **8:** 163-166

Zheng C, Hayes JJ (2003) Intra- and inter-nucleosomal protein-DNA interactions of the core histone tail domains in a model system. *Journal of Biological Chemistry* **278:** 24217-24224

Appendix

Appendix 1: ChIP-seq of DMSO- and GSK343-treated HeLa cells. The amount of reads obtained from each reaction, the number of reads that mapped to the human genome, the mouse genome, and the number of reads that could not be mapped are depicted.

Treatment	Replicate	ChIP	# total reads	# mapped human	# mapped mouse	# unmapped
DMSO	1	H3K27me3	113911968	74787230	19442463	19682275
		H3	102522508	71012639	12856658	18653211
		IgG	72747831	49018820	11404975	12324036
		input	116369053	86354416	10626027	19388610
	2	H3K27me3	79910282	47626991	18050085	14233206
		H3	97098140	65915419	13171632	18011089
		IgG	97297313	58871263	20771057	17654993
		input	88937213	66512430	7773056	14651727
GSK343	1	H3K27me3	72326818	47343747	12307152	12675919
		H3	86630951	62341154	8316458	15973339
		IgG	75508616	51284152	11085523	13138941
		input	89329647	68812742	5500450	15016455
	2	H3K27me3	84596062	57854730	11956507	14784825
		H3	94716321	70051851	7322333	17342137
		IgG	86459351	60414415	11046479	14998457
		input	94243869	72269396	5580627	16393846

Appendix 2: RNA affinity purification and mass spectrometry of *alu*RNA$_R$, CU-RNA-T0, CU-RNA-T2, RepA and CU-RNA-ΔT0. The peptide counts for all proteins that were identified are depicted.

Protein	Description	beads	*alu*RNA$_R$	T0	T2	RepA	ΔT0
NONO	Non-POU domain-containing octamer-binding protein	9	17	84	117	91	65
PTBP1	Polypyrimidine tract-binding protein 1	5	6	43	105	108	24
ROA1	Heterogeneous nuclear ribonucleoprotein A1	1	19	60	100	54	75
VIGLN	Vigilin	0	12	65	69	81	70
FUBP1	Far upstream element-binding protein 1	1	5	85	84	77	42
SND1	Staphylococcal nuclease domain-containing protein 1	0	0	58	109	64	21
HNRNPK	Heterogeneous nuclear ribonucleoprotein K	1	3	57	46	71	45
FUBP2	Far upstream element-binding protein 2	1	16	61	70	61	35
SFPQ	Splicing factor, proline- and glutamine-rich	13	21	37	65	53	32
HNRH1	Heterogeneous nuclear ribonucleoprotein H	3	37	28	49	45	17
ROA2	Heterogeneous nuclear ribonucleoproteins A2/B1	0	19	55	54	28	60
K2C1	Keratin, type II cytoskeletal 1	31	47	12	20	39	23
G3BP1	Ras GTPase-activating protein-binding protein 1	0	3	35	60	52	23
HNRNPQ	Heterogeneous nuclear ribonucleoprotein Q	1	10	36	56	41	35
K1C10	Keratin, type I cytoskeletal 10	28	30	8	21	28	17
LPPRC	Leucine-rich PPR motif-containing protein, mitochondrial	0	0	16	45	62	14
PABP1	Polyadenylate-binding protein 1	9	9	21	43	34	24
U2AF2	Splicing factor U2AF 65 kDa subunit	0	0	28	47	46	6
HNRNPD	Heterogeneous nuclear ribonucleoprotein D0	0	3	16	38	43	19
HNRNPU	Heterogeneous nuclear ribonucleoprotein U	0	11	30	31	22	27
NUCL	Nucleolin	6	19	14	21	37	12
HNRNPM	Heterogeneous nuclear ribonucleoprotein M	0	0	6	43	67	5

Protein	Description	beads	*alu*RNA$_R$	T0	T2	RepA	ΔT0
K1C9	Keratin, type I cytoskeletal 9	20	31	5	8	27	14
HNRNPL	Heterogeneous nuclear ribonucleoprotein L	0	0	18	40	12	16
K22E	Keratin, type II cytoskeletal 2 epidermal	21	32	6	12	26	15
CSDE1	Cold shock domain-containing protein E1	0	0	22	36	11	25
ROAA	Heterogeneous nuclear ribonucleoprotein A/B	0	4	22	22	27	22
HNRNPF	Heterogeneous nuclear ribonucleoprotein F	0	25	20	32	18	11
STRAP	Serine-threonine kinase receptor-associated protein	0	0	19	34	16	21
FUS	RNA-binding protein FUS	0	0	14	32	27	14
DAZP1	DAZ-associated protein 1	0	0	14	26	18	17
CAPR1	Caprin-1	0	2	16	22	23	18
PCBP2	Poly(rC)-binding protein 2	2	0	10	11	35	11
TADBP	TAR DNA-binding protein 43	0	4	14	17	19	8
RS3	40S ribosomal protein S3	0	1	11	18	20	16
IF2B3	Insulin-like growth factor 2 mRNA-binding protein 3	0	1	13	18	21	7
ELAV1	ELAV-like protein 1	0	0	9	17	27	9
TIAR	Nucleolysin TIAR	0	0	13	22	22	9
PUF60	Poly(U)-binding-splicing factor PUF60	0	0	5	23	28	3
CELF1	CUGBP Elav-like family member 1	0	0	1	38	22	3
RENT1	Regulator of nonsense transcripts 1	0	2	4	25	23	3
HNRH2	Heterogeneous nuclear ribonucleoprotein H2	0	22	19	34	28	13
DHX9	ATP-dependent RNA helicase A	0	11	8	16	11	7
ANXA2	Annexin A2	2	1	7	16	20	11
FUBP3	Far upstream element-binding protein 3	0	1	17	37	22	15
ROA0	Heterogeneous nuclear ribonucleoprotein A0	0	0	15	18	14	14
DDX5	Probable ATP-dependent RNA helicase	1	3	8	14	11	13
KHDR1	KH domain-containing, RNA-binding, sial transduction-associated protein 1	0	2	12	14	3	9
HNRDL	Heterogeneous nuclear ribonucleoprotein D-like	0	2	13	22	22	16
G3BP2	Ras GTPase-activating protein-binding protein 2	0	2	8	26	22	8
IF2G	Eukaryotic translation initiation factor 2 subunit 3	0	1	10	12	6	9
ROA3	Heterogeneous nuclear ribonucleoprotein A3	0	2	10	18	7	16
IF2A	Eukaryotic translation initiation factor 2 subunit 1	0	1	13	10	6	12
CCAR1	Cell division cycle and apoptosis regulator protein 1	0	0	2	11	20	4
HNRH3	Heterogeneous nuclear ribonucleoprotein H3	0	2	4	13	13	7
U2AF1	Splicing factor U2AF 35 kDa subunit	0	0	7	15	19	5
MATR3	Matrin-3	0	1	2	16	21	2
UBP2L	Ubiquitin-associated protein 2-like	0	0	9	14	13	8
CPSF7	Cleavage and polyadenylation specificity factor subunit 7	0	0	3	20	10	11
CPSF5	Cleavage and polyadenylation specificity factor subunit 5	0	0	2	18	11	14
BUB3	Mitotic checkpoint protein BUB3	0	0	8	11	12	8
EF1A1 (+1)	Elongation factor 1-alpha 1	0	2	4	14	11	8
CSTF3	Cleavage stimulation factor subunit 3	0	0	0	19	15	0
HNRNPR	Heterogeneous nuclear ribonucleoprotein R	0	0	16	33	20	16
RU17	U1 small nuclear ribonucleoprotein 70 kDa	0	0	10	4	0	12
GRSF1	G-rich sequence factor 1	0	8	2	10	9	1
SNRPA	U1 small nuclear ribonucleoprotein A	0	0	16	6	2	11
PABP4	Polyadenylate-binding protein 4	5	8	13	26	22	12
CSTF1	Cleavage stimulation factor subunit 1	0	0	0	18	17	0
IF2B1	Insulin-like growth factor 2 mRNA-binding protein 1	0	0	12	12	11	8
SF01	Splicing factor 1	1	2	5	8	5	2
MBNL1	Muscleblind-like protein 1	0	0	6	5	7	5
PAIRB	Plasminogen activator inhibitor 1 RNA-binding protein	0	0	4	11	6	6
ILF2	Interleukin enhancer-binding factor 2	0	0	0	11	11	3
PSPC1	Paraspeckle component 1	0	0	6	10	7	2
ADT3	ADP/ATP translocase 3	0	1	2	9	9	7
DDX17	Probable ATP-dependent RNA helicase DDX17	0	4	6	19	15	0
CN166	UPF0568 protein C14orf166	0	0	0	8	8	10
DDX3X	ATP-dependent RNA helicase DDX3X	0	2	3	9	5	5
AQR	Intron-binding protein aquarius	0	2	3	13	5	1
SMD2	Small nuclear ribonucleoprotein Sm D2	0	0	8	6	3	10
RTCB	tRNA-splicing ligase RtcB homolog	0	0	5	4	3	5
UBP10	Ubiquitin carboxyl-terminal hydrolase 10	0	0	2	9	6	4
SRSF1	Serine/arginine-rich splicing factor 1	0	0	5	3	8	6
SRSF2	Serine/arginine-rich splicing factor 2	0	1	4	5	8	4
RS10	40S ribosomal protein S10	0	0	2	5	3	4
ILF3	Interleukin enhancer-binding factor 3	0	0	1	10	8	0
EWS	RNA-binding protein EWS	0	1	4	8	5	3
RSMB	Small nuclear ribonucleoprotein-associated proteins B and B'	0	0	6	2	1	8
SMD1	Small nuclear ribonucleoprotein Sm D1	0	0	6	6	3	11
RBM14	RNA-binding protein 14	0	0	3	8	10	2
RLA0	60S acidic ribosomal protein P0	0	0	3	10	5	4
IF2B	Eukaryotic translation initiation factor 2 subunit 2	0	0	4	6	1	5
RS2	40S ribosomal protein S2	0	0	4	4	6	5
RL10A	60S ribosomal protein L10a	0	0	2	7	3	3
SMD3	Small nuclear ribonucleoprotein Sm D3	0	0	11	4	3	7

Protein	Description	beads	*alu*RNA$_R$	T0	T2	RepA	ΔT0
RS20	40S ribosomal protein S20	0	0	4	5	4	4
GEMI5	Gem-associated protein 5	0	0	0	6	12	0
DRG1	Developmentally-regulated GTP-binding protein 1	0	0	6	6	2	2
PTBP3	Polypyrimidine tract-binding protein 3	0	0	0	9	17	0
RS12	40S ribosomal protein S12	0	0	2	7	8	7
GTPB8	GTP-binding protein 8	0	0	2	4	6	5
RBM3	Putative RNA-binding protein 3	0	0	2	8	5	2
IF4G1	Eukaryotic translation initiation factor 4 gamma 1	0	0	1	9	3	5
CIRBP	Cold-inducible RNA-binding protein	0	0	2	6	6	5
PCBP1	Poly(rC)-binding protein 1	0	0	0	7	25	7
RUXE	Small nuclear ribonucleoprotein E	0	0	6	3	3	6
NEP1	Ribosomal RNA small subunit methyltransferase NEP1	0	0	1	4	6	7
DDX1	ATP-dependent RNA helicase DDX1	0	0	2	10	6	0
ACTG	Actin, cytoplasmic 2	1	4	1	4	3	3
HSP7C	Heat shock coate 71 kDa protein	0	0	0	6	6	2
RBM4	RNA-binding protein 4	0	0	0	0	15	0
RS5	40S ribosomal protein S5	0	0	4	4	4	2
SLIRP	SRA stem-loop-interacting RNA-binding protein, mitochondrial	0	0	1	4	5	3
CSTF2	Cleavage stimulation factor subunit 2	0	0	0	6	6	0
SRSF3	Serine/arginine-rich splicing factor 3	0	0	1	5	5	3
SRS11	Serine/arginine-rich splicing factor 11	0	0	1	6	7	0
SYF1	Pre-mRNA-splicing factor SYF1	0	0	0	6	1	0
NCOA2	Nuclear receptor coactivator 2	0	2	0	0	0	0
K2C8	Keratin, type II cytoskeletal 8	0	5	0	4	6	2
ODB2	Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial	1	4	0	0	0	2
SF3A1	Splicing factor 3A subunit 1	2	5	0	0	1	0
PHB	Prohibitin	0	0	0	4	1	5
IQGA1	Ras GTPase-activating-like protein IQGAP1	0	3	0	4	0	0
RL22	60S ribosomal protein L22	0	0	4	4	3	2
RS9	40S ribosomal protein S9	0	0	0	2	2	3
RBM39	RNA-binding protein 39	0	0	1	4	5	0
MANF	Mesencephalic astrocyte-derived neurotrophic factor	1	1	0	1	2	2
RS19	40S ribosomal protein S19	0	0	1	3	3	1
RU2A	U2 small nuclear ribonucleoprotein A'	0	0	1	3	3	4
MSI2H	RNA-binding protein Musashi homolog 2	0	0	1	2	0	7
RL12	60S ribosomal protein L12	1	1	1	3	3	3
TIA1	Nucleolysin TIA-1 isoform p40	0	0	7	10	8	0
TOP1	DNA topoisomerase 1	0	0	1	4	1	1
ZN207	BUB3-interacting and GLEBS motif-containing protein ZNF207	0	0	1	3	3	2
RAN	GTP-binding nuclear protein Ran	0	0	2	3	3	2
PPIB	Peptidyl-prolyl cis-trans isomerase B	0	0	1	2	3	2
K2C5	Keratin, type II cytoskeletal 5	7	9	0	0	4	0
MOES	Moesin	0	0	1	3	2	0
RBM10	RNA-binding protein 10	0	0	0	1	6	1
RS16	40S ribosomal protein S16	0	0	0	3	3	3
SREK1	Splicing regulatory glutamine/lysine-rich protein 1	0	0	0	5	4	0
SDOS	Protein syndesmos	0	1	0	2	1	1
RCC2	Protein RCC2	0	0	1	2	2	0
K2C6A	Keratin, type II cytoskeletal 6A	6	9	0	0	5	0
K1C14	Keratin, type I cytoskeletal 14	8	6	0	0	0	3
CPSF6	Cleavage and polyadenylation specificity factor subunit 6	0	0	0	9	0	0
HNRNPC	Heterogeneous nuclear ribonucleoproteins C1/C2	0	0	1	1	5	2
RUXF	Small nuclear ribonucleoprotein F	0	0	3	2	1	4
PHB2	Prohibitin-2	0	0	0	2	2	2
LRC59	Leucine-rich repeat-containing protein 59	0	0	0	2	1	3
NUFP2	Nuclear fragile X mental retardation-interacting protein 2	0	0	0	2	3	1
RBMX	RNA-binding motif protein, X chromosome	0	0	0	3	0	3
CNBP	Cellular nucleic acid-binding protein CNBP	0	0	2	0	0	5
NCBP1	Nuclear cap-binding protein subunit 1	0	0	0	4	0	0
PLOD1	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1	0	7	0	0	0	0
ALBU	Serum albumin	2	4	0	0	0	0
F120C	Constitutive coactivator of PPAR-gamma-like protein 2	0	0	0	0	1	2
LC7L3	Luc7-like protein 3	0	0	1	1	0	7
IF4B	Eukaryotic translation initiation factor 4B	0	0	0	4	1	0
RL11	60S ribosomal protein L11	0	0	2	2	2	2
HNRL1	Heterogeneous nuclear ribonucleoprotein U-like protein 1	0	2	0	2	1	0
RS4X	40S ribosomal protein S4, X isoform	0	0	0	2	1	0
LYRIC	Protein LYRIC MTDH	0	0	0	6	0	0
RS7	40S ribosomal protein S7	0	0	1	1	1	2
MCM5	DNA replication licensing factor	0	1	0	2	1	0
TCP4	Activated RNA polymerase II transcriptional coactivator p15	0	0	0	2	2	2

Protein	Description	beads	*alu*RNA$_R$	T0	T2	RepA	ΔT0
RBM47	RNA-binding protein 47	0	0	1	3	2	0
RS15A	40S ribosomal protein S15a	0	0	0	1	2	2
GRP78	78 kDa glucose-regulated protein	0	0	0	3	2	1
RS27A	Ubiquitin-40S ribosomal protein S27a	0	0	0	2	1	0
DHX15	Putative pre-mRNA-splicing factor ATP-dependent RNA helicase	0	0	0	2	3	0
RL9	60S ribosomal protein L9	0	0	0	2	2	0
RUXG	Small nuclear ribonucleoprotein G	0	0	1	0	0	3
HS71A	Heat shock 70 kDa protein 1A	0	0	0	5	4	0
ZCCHV	Zinc finger CCCH-type antiviral protein 1	0	0	0	1	3	0
PURB	Transcriptional activator protein Pur-beta	0	0	0	0	0	3
RAB10	Ras-related protein Rab-10	0	0	1	2	1	0
SC61B	Protein transport protein Sec61 subunit beta	0	0	0	0	2	1
RS28	40S ribosomal protein S28	0	0	0	1	0	2
DHX36	ATP-dependent RNA helicase DHX36	0	0	0	2	3	0
ZCRB1	Zinc finger CCHC-type and RNA-binding motif-containing protein 1	0	0	0	0	3	0
ADT2	ADP/ATP translocase 2	0	0	0	8	10	6
IMA1	Importin subunit alpha-1	0	0	0	1	2	0
ZC3HF	Zinc finger CCCH domain-containing protein 15	0	0	1	3	0	0
IF2P	Eukaryotic translation initiation factor 5B	0	0	1	3	0	0
MCM3	DNA replication licensing factor	0	0	0	2	0	0
K1C16	Keratin, type I cytoskeletal 16	10	0	0	0	0	0
SYFB	Phenylalanine-tRNA ligase beta subunit	0	0	0	1	2	0
SSBP	Single-stranded DNA-binding protein, mitochondrial	0	0	0	3	0	0
PUM2	Pumilio homolog 2	0	0	0	2	0	0

Acknowledgements

My sincere thanks to…

PD Dr. Karsten Rippe for providing me the opportunity to perform my PhD work in his laboratory, for his supervision and support throughout the project, for his optimistic view on new results, and for his motivating words.

PD Dr. Georg Stöcklin and Dr. Sylvia Erhardt for taking the time to read and listen to my work, for the fruitful discussions during the TAC meetings, Georg for being my second referee and Sylvia for being part of my examination committee.

Prof. Dr. G. Elisabeth Pollerberg for her interest in my work, and for chairing my examination committee.

Dr. Maïwen Caudron-Herger for always being there to discuss and comment on my work, for always defending my interests, for sharing her advice on scientific as well as career-related issues, for being my taxi driver to the EMBL, and for critical reading of my thesis.

Katha for sharing the ups and downs of the last four years, for always having an open ear, and for being a superb event manager.

Sabrina for her excellent assistance and patience with experiments particularly over the last few months.

All present and former members of the “Organization & Fun” group, Katharina, Maïwen, Fabian, Benni, Caro, Sabrina, JanaMo, JanaHe, Sarah, Sebastian, Anne, Katha, Inn, Philipp, Ronja, Michael, Delia, Lara, Laura, Kathi, Nick, Verena, Sofie, Daria, Vladimir, and Karsten for our coffee breaks, Kubb sessions, the Friday-madness, the great atmosphere and for your readiness to always help me and to discuss my work.

The Cusanuswerk for supporting me financially and for their outstanding educational program, which continuously widened my horizon beyond my PhD work.

My friends from Heidelberg and abroad for a great time, and for successfully taking my mind off my experiments every now and then.

My parents and my sister for their continuous support on so many levels and despite all the little obstacles on my way to where I am now. Without you I would not be there!

Joschka for proofreading my thesis, for never getting tired of discussing science even over dinner, but mostly for always being there for me, and for making my time spent outside the lab so wonderful.