

Aus dem Institut für Molekularbiologie und Tumorforschung  
Geschäftsführender Direktor: Prof. Dr. Alexander Brehm

des Fachbereichs Medizin der Philipps-Universität Marburg

## **Characterization of RNA interactions of dMi-2**



Inaugural-Dissertation

zur Erlangung des Doktorgrades  
der Naturwissenschaften  
(Dr. rer. Nat.)

dem Fachbereich Medizin  
der Philipps Universität Marburg  
vorgelegt von

**Ikram Ullah**  
aus Srinagar, Jammu & Kashmir

Marburg, 2021

Angenommen vom Fachbereich Medizin der Philipps-Universität Marburg am:  
12.03.2021

Gedruckt mit Genehmigung des Fachbereichs

Dekanin: Frau Prof. Dr. D. Hilfiker-Kleiner  
Referent: Prof. Dr. A. Brehm  
1. Korreferent: Frau Prof. Dr. S. Hake

To,

*My Parents and my Sisters!*

عالمَن یکتا، علیم جُحکھ اُگر  
پیش نظر تھاو حکیمَن کلام  
کتاب خوانو! مَد تَمکر سُنڈ عالم  
اِدِ ہو کرکھ جَہلس سلام

"There is the unity of knowledge as the All-Knowing is their [only] source;  
so keep the Word of the All-Wise in front of you.  
O the reader of [His] Book! Study His universe [also],  
then you can bid farewell to ignorance."

(Dr. Muied-uz-Zaffar)

**Table of Content**

**LIST OF FIGURES .....IX**

**LIST OF TABLES .....X**

**LIST OF ACRONYMS AND ABBREVIATIONS.....XI**

**1. INTRODUCTION..... 1**

1.1 CHROMATIN..... 1

1.1.1 Histones..... 4

1.1.2 Histone variants ..... 5

1.2 CHROMATIN MODIFICATIONS AND GENE EXPRESSION REGULATION..... 7

1.2.1 Histone modifications ..... 8

1.2.1.1 Histone acetylation..... 9

1.2.1.2 Histone Phosphorylation ..... 11

1.2.1.3 Histone methylation..... 12

1.2.2 ATP-dependent chromatin remodeling enzymes ..... 15

1.2.2.1 ISWI family ..... 15

1.2.2.2 SWI/SNF family ..... 18

1.2.2.3 INO80 family..... 19

1.2.2.4 CHD family ..... 20

1.2.3 Nucleosome remodeling and deacetylase (NuRD) complex..... 25

1.3 RNA BINDING PROTEINS..... 29

1.3.1 Chromatin regulators as RNA binding proteins ..... 30

1.3.1.1 MOF ..... 30

1.3.1.2 PRC2 complex ..... 31

1.3.1.3 dMi-2/CHD4..... 32

1.4 HOW DO PROTEINS BIND RNA?..... 33

1.4.1 Molecular Interactions..... 33

1.4.1.1 Hydrogen bonds and Van der Waals interactions..... 34

1.4.1.2 Hydrophobic and  $\pi$  ( $\pi$ ) stacking interactions ..... 35

1.4.2 RNA binding domains ..... 35

1.4.2.1 RNA recognition motif ..... 36

1.4.2.2 K homology domain ..... 36

1.4.2.3 Zinc Finger domain ..... 37

1.4.2.4 Cold shock domain .....	37
1.4.3 Intrinsically disordered regions.....	38
1.5 OBJECTIVES.....	39
<b>2 MATERIALS AND METHODS .....</b>	<b>43</b>
2.1 MATERIALS .....	43
2.1.1 Cell lines .....	43
2.1.2 Cell Culture .....	43
2.1.3 Baculovirus .....	44
2.1.4 Plasmids .....	44
2.1.5 Inhibitors .....	46
2.1.6 Enzymes .....	47
2.1.7 Oligonucleotides .....	47
2.1.7.1 Oligonucleotides for generating templates for <i>in vitro</i> transcription.....	47
2.1.7.2 Oligonucleotides for cloning in pFastBac1 vector to generate Baculovirus .....	48
2.1.7.3 Oligonucleotides for sequencing.....	49
2.1.7.4 Oligonucleotides for ChIP and qPCR .....	49
2.1.8 Ladders and dyes .....	50
2.1.9 Chemicals .....	50
2.1.10 Kits.....	52
2.1.11 Consumables.....	52
2.1.12 Devices .....	53
2.1.13 Software.....	54
2.1.14 RNAs.....	54
2.1.15 Antibodies and antisera.....	57
2.1.15.1 Primary antibodies .....	57
2.1.15.2 Secondary antibodies.....	58
2.2 METHODS.....	59
2.2.1 Cell Culture .....	59
2.2.1.1 Culturing Sf9 cells .....	59
2.2.1.2 Culturing S2 cells .....	59
2.2.1.3 Cell counting.....	59
2.2.1.4 Freezing and thawing cells.....	59
2.2.1.5 Infection of Sf9 cells.....	60
2.2.2 Biochemical Methods .....	61

2.2.2.1 Baculovirus protein extraction .....	61
2.2.2.2 FLAG-immunoprecipitation (FLAG-IP).....	61
2.2.2.3 Protein concentration estimation.....	62
2.2.2.4 Sodium dodecyl sulfate Polyacrylamide gel electrophoresis (SDS- PAGE) .....	63
2.2.2.5 Coomassie staining and imaging.....	64
2.2.2.6 RNase A treatment and Cell Fractionation .....	65
2.2.2.7 Western Blot .....	66
2.2.2.8 Agarose gel electrophoresis.....	67
2.2.2.9 Electrophoretic mobility shift assay (EMSA) .....	68
2.2.2.10 Native Polyacrylamide gel electrophoresis .....	69
2.2.2.11 Denaturing polyacrylamide gel electrophoresis (denaturing PAGE)...	69
2.2.2.12 Circular dichroism (CD) spectroscopy .....	69
2.2.2.13 <i>in vitro</i> transcription.....	70
2.2.2.14 Nucleosome assembly .....	71
2.2.2.15 Nucleosome remodelling assay.....	72
2.2.2.16 Reverse transcript- quantitative polymerase chain reaction (RT-qPCR) .....	72
2.2.2.17 Chromatin immunoprecipitation-quantitative polymerase chain reaction (ChIP-qPCR).....	73
2.2.2.18 RNase A and Immunofluorescence .....	77
2.2.2.19 Gibson assembly.....	77
2.2.2.20 individual nucleotide resolution cross-linking and immunoprecipitation-2 (iCLIP2).....	79
2.2.3 Bioinformatics Pipeline.....	86
<b>3 RESULTS.....</b>	<b>90</b>
3.1 CHAPTER 1 .....	90
3.1.1 dMi-2 is an RNA-binding protein.....	90
3.1.2 dMi-2 binds RNA promiscuously.....	91
3.1.3 dMi-2 binds some nucleic acids with higher affinity than others <i>in vitro</i> .....	95
3.1.4 dMi-2 binds RNA in high salt and detergent conditions .....	98
3.1.5 Mapping the RNA binding region of dMi-2.....	101
3.1.6 dMi-2 binding to homopolymer RNAs.....	106
3.1.7 dMi-2 binds to G-rich RNA.....	109
3.2 CHAPTER 2 .....	113

3.2.1 dMi-2 Binds RNA <i>in vivo</i> .....	113
3.2.2 dMi-2 iCLIP.....	115
3.2.3 dMi-2-GFP promiscuously binds to different types of RNAs <i>in vivo</i> .....	118
3.2.4 dMi-2 GFP binds 3'-ends of mature mRNAs .....	120
3.3 CHAPTER 3 .....	126
3.3.1 RNA binding inhibits dMi-2 remodelling activity.....	126
3.3.2 RNAs that bind efficiently to dMi-2 inhibit remodelling.....	129
3.3.3 Loss of RNA causes increased association of dMi-2 with chromatin.....	131
3.3.4 Binding to RNA prevents dMi-2 from interaction with chromatin .....	134
<b>4 DISCUSSION .....</b>	<b>138</b>
<b>5 SUMMARY.....</b>	<b>148</b>
5.1 SUMMARY.....	148
5.2 ZUSAMMENFASSUNG .....	150
<b>6 REFERENCES .....</b>	<b>152</b>
<b>7 APPENDIX .....</b>	<b>170</b>
SUPPLEMENTARY DATA.....	170
PHD PORTFOLIO - MGK.....	203
CURRICULUM VITAE .....	206
LIST OF ACADEMIC TEACHERS.....	208
ACKNOWLEDGEMENT:.....	209
EHRENWÖRTLICHE ERKLÄRUNG.....	210

## List of Figures

List of Figures.....	2
Figure 1.1 The structure of nucleosome .....	2
Figure 1.2: Core histones and their variants .....	7
Figure 1.3: Histone post-translational modifications.....	9
Figure 1.4: Subfamilies of ATP-dependent chromatin remodelers and their roles. ....	17
Figure 1.5: Protein-RNA interactions. ....	30
Figure 1.6: Protein-RNA interactions .....	34
Figure 3.1.1: dMi-2 binds RNA <i>in vitro</i> .....	91
Figure 3.1.2: dMi-2 binds RNA promiscuously .....	94
Figure 3.1.3: dMi-2 binds some nucleic acids with more affinity than others.....	98
Figure 3.1.4: dMi-2 binds RNA in high salt and detergent conditions.....	100
Figure 3.1.5: Mapping the RNA binding region of dMi-2.....	106
Figure 3.1.6: dMi-2 binding to homopolymer RNAs.....	109
Figure 3.1.7: dMi-2 binds to G-rich RNAs.....	111
Figure 3.2.1: dMi-2 Binds RNA <i>in vivo</i> .....	114
Figure 3.2.2: dMi-2 iCLIP-seq membranes.....	116
Figure 3.2.3: iCLIP library preparation .....	118
Figure 3.2.4: dMi-2 binds to different types of RNAs <i>in vivo</i> .....	120
Figure 3.2.5: dMi-2 binds 3'-ends of mature mRNAs .....	124
Figure 3.3.1: RNA binding inhibits dMi-2 remodelling.....	128
Figure 3.3.2: RNAs that bind efficiently to dMi-2 inhibit remodelling .....	130
Figure 3.3.3: Loss of RNA causes increased association of dMi-2 with chromatin...	133
Figure 3.3.4: Binding to RNA prevents dMi-2 from interaction with chromatin .....	136
Figure 4.0: Model.....	146
Supplementary Figure S3.1.2 .....	171
Supplementary Figure S3.1.3: .....	173
Supplementary Figure S3.1.5: .....	176
Supplementary Figure S3.2.5: .....	177
Supplementary figure S3.3.1: RNA binding inhibits dMi-2 remodelling activity.....	179
Supplementary figure S3.3.2: RNAs that bind efficiently to dMi-2 inhibit remodelling .....	182
Supplementary figure S3.3.3: Loss of RNA causes increased association of dMi-2 with chromatin .....	183



## List of Tables

Table 1: Dissociation constants for binding of various nucleotides to dMi-2.....	96
Table 2: RNA sequences.....	114
Table 3: dMi-2-GFP iCLIP-seq list of enriched mRNAs with Fold change > 4 and ( $p < 0.05$ ).....	179

## List of acronyms and abbreviations

<b>Acronym</b>	<b>Abbreviation</b>
Å	Angstrom
ac	Acetylation
ACF	ATP-dependent chromatin assembly and remodeling factor
Acronyms	Abbreviation
ADP	Adenosine diphosphate
ANRIL	Antisense ncRNA in the INK4 locus
ARMs	Arginine-rich motifs
ATP	Adenosine triphosphate
AutoN	Autoinhibitory N-terminal
BAP	Brahma associated factors
bp	Basepair
BRK	Bhrama/Kismet
BRM	Brahma
C-terminal	Carboxy-terminal
CAF1	Chromatin assembly factor 1
CBP	CREB binding protein
CCC	Chromatin conformation capture
CD	Chromodomains
CENP-A	Centromeric protein A
CHD	Chromodomain helicase DNA binding
CHRAC	Chromatin accessibility complex
CoA	Co-enzyme A
COMPASS	Complex protein associated with set1
CpG	5'-C phosphate G-3'
cryo-EM	Chyogenic electron micriscopy
CSD	Cold shock domain
CSPs	Cold shocj proteins
CTCF	CCCCTC-binding factor
CTD	C-terminal domain
DNA	Deoxyribo nucleic acid

## Acronyms and Abbreviations

Dom	Domino
EED	Embryonic ectoderm development
EZH2	Enhancer of zeste 2
FAD	Flavin adenine dinucleotide
FMRP	Fragile-X mental retardation protein
FOXO1	Forkhead box protein O1
G-4	G-quadruplex
GATAD2A/B	GATA Zinc finger domain containing 2A/B
GCN5	General control non-derepressible 5
GNAT	GCN5-related N-acetyltransferase
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HDACi	HDAC inhibitor
HDX-MS	Hydrogen deuterium exchange-Mass spectroscopy
Hi-C	High throughput-Chromatin conformation capture
HIRA	Histone regulator A
HMT	Histone methyltransferase
hnRNP-U	human nuclear ribonucleoprotein-U
hnRNPK	Heterogeneous nuclear ribonucleoprotein K
HOTAIR	HOX antisense intergenic RNA
HOX	Homeobox
HP1 $\alpha$	Heterochromatin protein 1 alpha
HS	Heat shock
Hsp	Heat shock protein
HSS	HAND, SANT, SLIDE
IDRs	Intrinsically disordered regions
INO80	Inositol requiring 80
iPSCs	induced pluripotent stem cells
ISCs	Intestinal stem cells
ISWI	Imitation switch
K	Lysine
kb	Kilo-bases
kcal/mol	kilo-calories per mole
kDa	kilo-Daltons
KH domain	K-homology domain

## Acronyms and Abbreviations

KIS	Kismet
LB-A	Luria broth-agar
LincRNAs	Long intergenic non-coding RNAs
LSD1	Lysine specific demethylase 1
MALAT1	Metastasis associated lung adenocarcinoma virus transcript 1
MBD2/3	Methyl-CpG DNA binding domains 2/3
MBNL-1	Muscleblind-like 1
MCPH1	Microcephalin 1
me	Methylation
Mec	Mep-1 containing complex
Mep-1	mog interacting ectopic P granulocyte 1
mESCs	mouse embryonic stem cells
Mi-2	Mitchell-2
mM	Milimolar
MOF	Males absent on the first
MSK1	Mitogen- and stress-activated protein kinase 1
MSL	Male-specific lethal
MTA1/2/3	Metastasis associated proteins-1/2/3
MYST	MOZ, Ybf2, Sas2 and Tip60
N-terminal	Amino-terminal
NAD	Nicotine amine adenine dinucleotide
ncRNA	non-coding RNA
NegC	Negative regulator of coupling
NFRs	Nucleosome free regions
nm	Nanometer
NMR	Nuclear magnetic resonance
nt	nucleotides
NuA4	Nucleosome acetyltransferase 4
NuRD	Nucleosome remodelling and deacetylase
NURF	Nucleosome remodeling factor
OB-folds	Ologonucleotide binding folds
PAD4	Protein Arginine deiminase 4
PAF1	Polymerase associated factor 1
PANDA	p21-associated ncRNA DANN damage activated
PAPAS	Promoter and pre-RNA antisense

## Acronyms and Abbreviations

PAR	Poly-ADP ribose
PAX3	Paired box gene 3
PBAP	Polybromo-associated Brahma
ph	Phosphorylation
PHD	Plant homeodomain
Pho	Pleiohomeotic
PINC	Pregnancy induced long ncRNA
Pol II	Polymerase II
PP1	Protein phosphatase 1
PRC2	Polycomb repressive complex 2
PRE	Polycomb responsive elements
PRMTs	Protein Arginine methyltransferases
PTMs	Post-translational modifications
R	Arginine
RBBP 4/7	Retinoblastoma binding protein 4/7
RBD	RNA binding domain
RBPs	RNA binding proteins
rDNA	Ribosomal DNA
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RRM	RNA recognition motif
S-phase	Synthesis phase
S2	Schneider 2
SAGA	Spt-Ada-GCN5-acetyltransferase
SAM	S-adenosylmethionine
SET	Su(var) 3-9, Enhancer of zeste and Trithorax
SHL	Super helix location
Snf2	Sucrose non-fermenting 2
SRC3	Steroid receptor cofactor 3
SRSF-1	Serine/Arginine-repeat splicing factor 1
STORM	Stochastic optical reconstruction microscopy
SUZ12	Suppressor of zeste 12
SWI/SNF	Switch/sucrose non-fermenting
TADs	Topologically associated domains
TAF1	Transcription initiation factor TFIID subunit 1

## Acronyms and Abbreviations

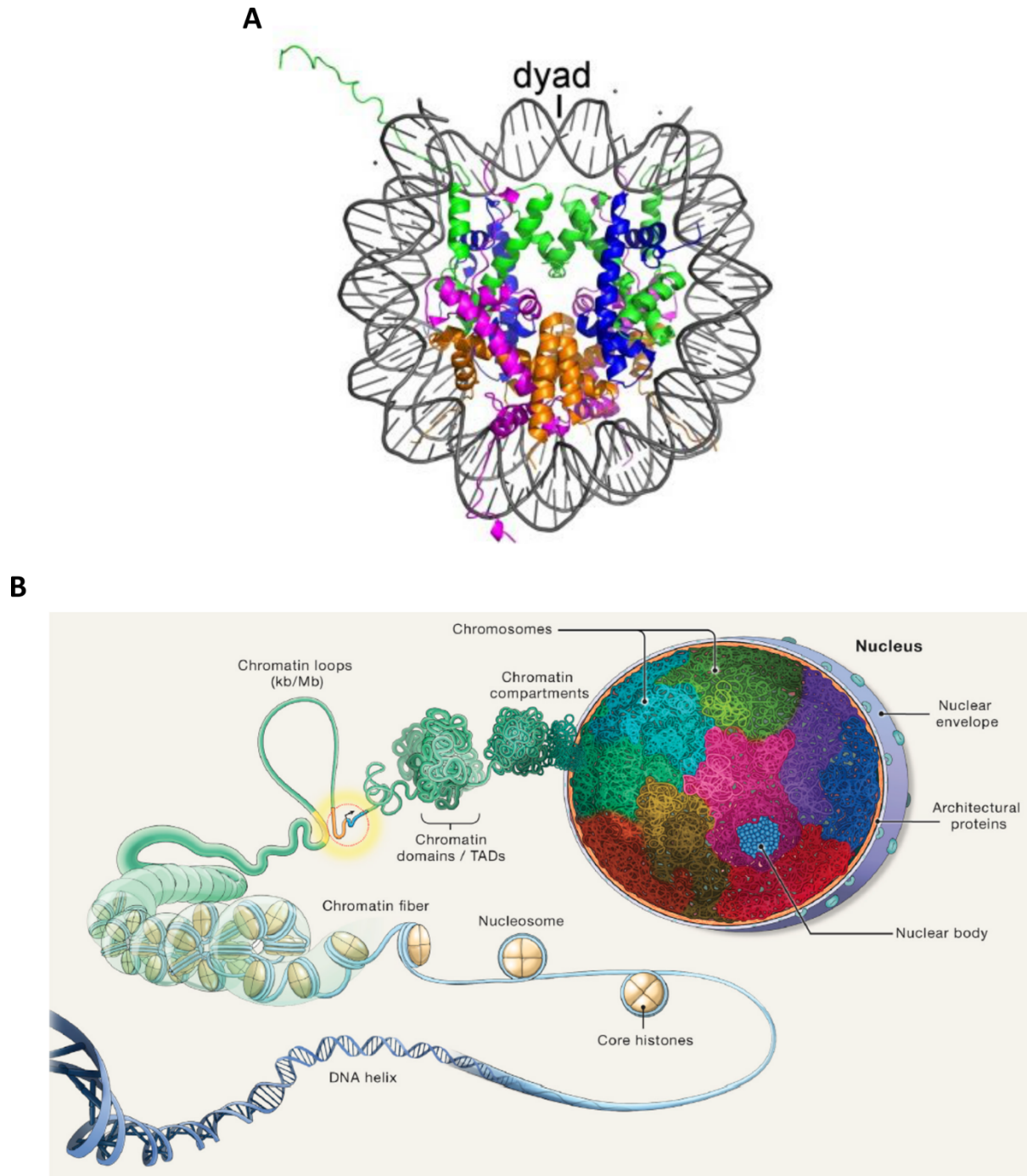
TAR	Trans-activation response element
TAT	Trans-activator of transcription
TFIIH	Transcription factor II Human
TSS	Transcription start site
ub	Ubiquitination
UTR	Untranslated region
VdW	Van der Waals
WDR5	WD-repeat containing protein 5
ZnF	Zinc finger
μM	Micrometer

# 1. Introduction

## 1.1 Chromatin

Almost each molecular aspect of life is controlled by a flow of information encoded in DNA to protein through an intermediary molecule called RNA. Some aspects of life are regulated by the RNA itself. This flow of coded information, from DNA to proteins, is called the 'central dogma of molecular biology'(Crick, 1970). First reported as 'invisible factor' his work in heredity, Gregor John Mendel, discovered the basic unit of DNA that codes for a protein, now called a gene (Mendel, 1901). DNA, thus began to be called the genetic material (Griffith, 1928). Later, it was demonstrated that RNA as well could serve as the genetic material in some viruses (Gierer and Schramm, 1956). The genetic material of a human diploid cell is about 6 billion base pairs (bp) of DNA which theoretically amounts to about 2 meters in length (Annunziato, 2008). How does the DNA of such length fit in the nucleus of a cell that is only few micrometres ( $\mu\text{M}$ ) in diameter? The answer lies in the manner in which DNA is packaged with proteins called histones that allow super-folding of DNA such that it fits into the microscopically small nucleus of the cell. Histones are highly positively charged proteins that attract the negatively charged phosphodiester backbone in the DNA to form a histone-DNA complex. The DNA is wrapped, approximately 1.65 times, around the histone core octamer (two copies each of - H2A, H2B, H3 and H4) forming a structure that encompasses 146 or 147 bp of DNA and is called a nucleosome (**Figure 1.1A**).

The nucleosome forms the fundamental building block of the chromatin and it has been crystallized to give important structural details that help us in understanding how the chromatin is built (**Figure 1.1A**). The DNA in a cell is wrapped around millions of histone octamers separated by 10 to 90 bp linker DNA, giving rise to numerous nucleosomes that appear as 'beads on a string' of 10 nanometer (nm) diameter. This provides the first order of compaction of the DNA (Woodcock, 2006). Another histone protein, histone H1, binds the linker DNA at its entry and exit sites on the histone core. This causes the DNA to bend and allows further compaction of the chromatin (Xiao et al., 2012). Previous hypothesis had suggested that at the physiological state with 100-150 millimolar (mM) salt and 2-5 mM  $\text{Mg}^{2+}$ , the nucleosome-nucleosome and nucleosome-DNA interactions allow the formation of secondary structures and further folding of the chromatin, leading to the formation of a 30-nm fibre or solenoid. The fibre-fibre interactions were hypothesized to allow the formation of tertiary chromatin structures (Szerlong and Hansen, 2011).



**Figure 1.1 The structure of nucleosome**

**(A)** The 2.8 Å (angstrom) crystal structure of the nucleosome core particle with histone proteins H2A (Magenta), H2B (Orange), H3 (Green) and H4 (Blue) in a left-handed superhelix. The 146 bp double-helix DNA is depicted in grey. The centre of the nucleosomal DNA – dyad, is indicated. Adapted from (Luger et al., 1997). **(B)** Schematic representation of chromatin organization depicting double helical DNA wrapped around histone octamers creating chromatin fibre that appear as ‘beads on a string’ with linker DNA. The chromatin fibers interact to form clutches on nucleosomes that are compacted by looping. Loops interact to



form TADs which in turn lead to the formation of compartments in the nucleus. Adapted from (Misteli, 2020).

However, the technological advancements over the last decade have helped to improve our understanding of the chromatin architecture. In fact, the existence of the 30-nm fibre and the higher order structures that were stipulated to form from fibre-fibre interactions have been challenged. The stochastic optical reconstruction microscopy (STORM) experiments show the presence of nucleosomes in 'clutches' instead. The clutches are several nucleosomes in a group and can have varying density of nucleosomes depending on the context and cell type. For example, in the transcriptionally silenced regions of chromatin, the clutches have a high density of nucleosomes while those clutches in actively transcribed regions have less nucleosomes. Similarly, the pluripotent stem cells have less density of nucleosomes in their clutches allowing access for several transcriptional factors to their binding targets. The differentiated cells, like fibroblasts, have higher density of nucleosomes in their clutches (Ricci et al., 2015). Another major technological advancement that improved our understanding of chromatin organization is the chromatin conformation capture (CCC) assay and its most recent iteration which couples it with high throughput sequencing (Hi-C). This technique makes use of chemical crosslinking of chromatin combined with sequencing to extract the information on the architecture of chromatin and has revolutionized our understanding of chromatin organization. Recent findings strongly support the existence of loops and domains. Chromatin forms loops and it has been proposed that these loops help to bring the enhancers in close proximity to the promoters that might otherwise be distant. Loops can be of different sizes, ranging from few kilo-bases to hundreds of kilo-bases and have been shown to form by a process called loop extrusion (Dekker and Misteli, 2015; Misteli, 2020). Loop extrusion involves a chromatin fibre looping out through the ring-shaped chromatin architectural protein – cohesin, helped by CTCF, among other proteins. CTCF has also been involved in chromatin domains, whereby, the regions of the genome that interact with each other are separated into a domain. Such domains that interact mostly within, but rarely with the surrounding regions, are called topologically associated domains (TADs). The TADs are usually several hundred kilo-bases in size and form by loop extrusion, wherein, once a loop is extruded it folds onto itself and forms distinct domains (Dekker and Mirny, 2016). The genes within a TAD are regulated from within the TAD. It has been shown that disturbing a TAD boundary could affect the regulation of the genes within that TAD. Depending on whether the genes within a domain are actively transcribed or silenced, the TADs can interact to form higher-order chromatin compartments - A or B, which correspond

to euchromatin and heterochromatin, respectively. The compartment - A comprises of actively transcribed genes whereas silenced genes remain in compartment - B (Figure 1.1B) (Misteli, 2020).

The major units of genome organization – chromosomes, exist in the form of chromosome territories. During interphase, each chromosome appears as a relatively compact and spatially restricted territory in the nucleus. Chromosome territories in mammalian cells are about 2-3  $\mu\text{M}$  in diameter. The localization of chromosomes in the nuclear space also is a level of chromatin organization. The regions of chromosomes harbouring actively transcribed genes locate towards the inner nuclear space which corresponds to euchromatin or compartment – A. The parts of the chromosomes with silenced genes are arranged towards the nuclear periphery or the heterochromatin or compartment – B (Misteli, 2020). Recent studies indicate that phase separation, that is, distinct protein populations that preferentially interacts among themselves than with other types of biomolecules, upon reaching a threshold of saturation, separate into distinct and dynamic phases, play an important role in genome organization. For example, the protein involved in heterochromatin formation, heterochromatin protein HP1 $\alpha$ , phase separates both in vitro as well as in vivo leading to the belief that heterochromatin formation may be driven by phase separation (Banani et al., 2017; Misteli, 2020).

### **1.1.1 Histones**

The core histone octamer consists of two copies each of histones H2A, H2B, H3 and H4. Histones contain a three-helix core domain, called the histone fold. These domains interact with each other in a ‘handshake’ arrangement that leads to the formation of H2A-H2B and H3-H4 dimers (Arents and Moudrianakis, 1993; Khorasanizadeh, 2004). These canonical histones are produced from highly regulated gene clusters during the S-phase of the cell cycle and are subsequently incorporated into the chromatin during the DNA replication (DeLisle et al., 1983). The core histone proteins have similar structural features and are highly conserved among the eukaryotes (Malik and Henikoff, 2003). Apart from the highly structured central region, each histone contains unstructured N-terminal and C-terminal tails which protrude from the nucleosome. The histone tails are post-translationally modified by histone modifying enzymes and these modifications are important for the formation of higher order chromatin structures that influence the dynamic nature of the chromatin. The core histones

can be replaced by histone variants, that along with the post-translational modifications of histone tails, contribute in regulating the gene expression (Fischle et al., 2003; Lawrence et al., 2016).

### **1.1.2 Histone variants**

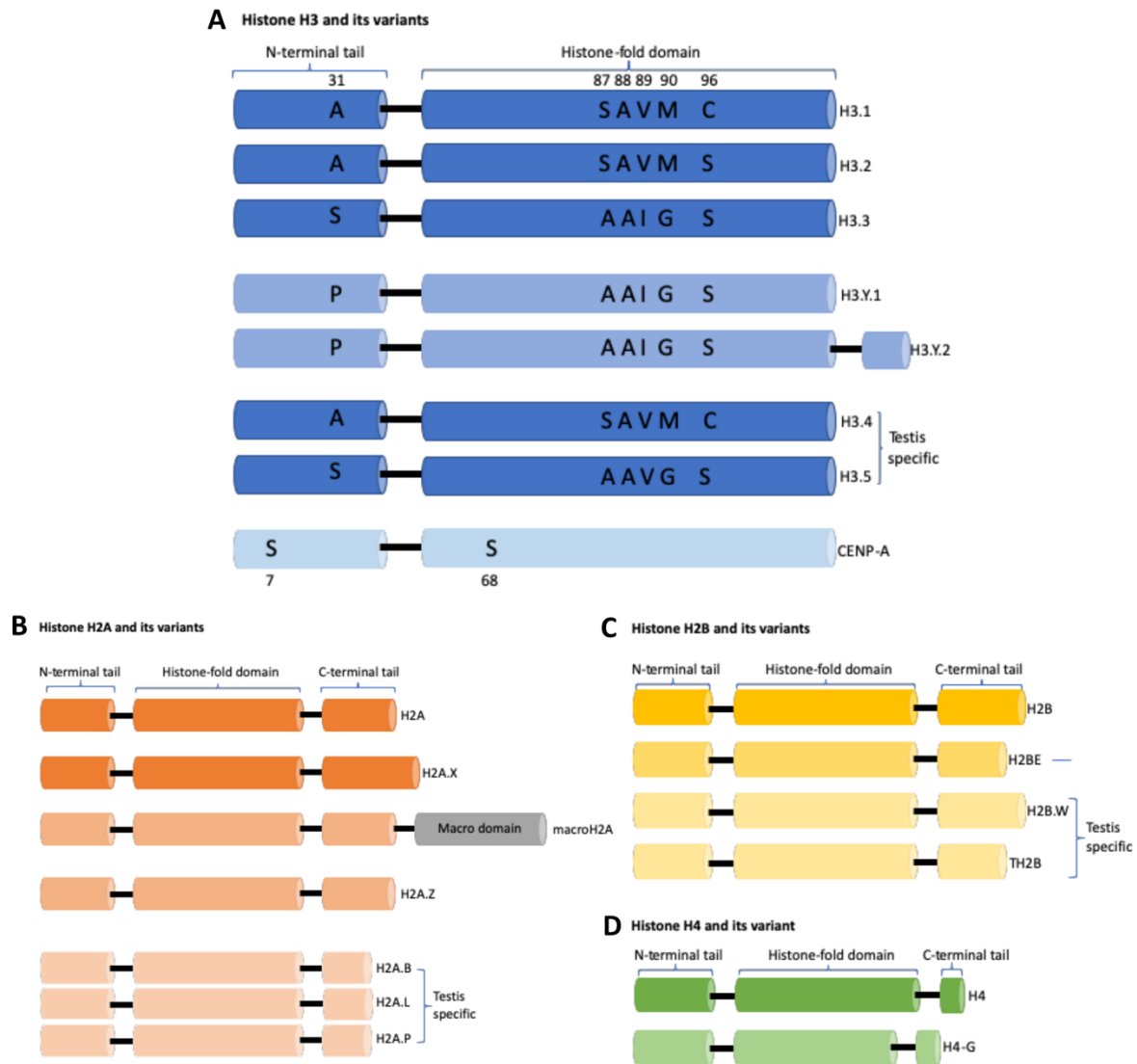
The variants of histones play an important role in the diversification of the nucleosome structure and the epigenetic regulation of the gene expression. Histone variants could attract variant-specific interacting proteins, thereby imparting certain genomic regions with unique characteristics to regulate the gene expression. For example, centromeric protein A (CENP-A), a variant of histone H3, is specific to centromeres, and epigenetically marks them distinctly (Dunleavy et al., 2009; Foltz et al., 2009; Howman et al., 2000; Martire and Banaszynski, 2020). While the canonical histone genes have no introns and are transcribed from multiple histone gene clusters, the variant histones are transcribed from regions outside of the canonical histone clusters and often have introns that give rise to several splice isoforms. From amongst the histone variants, some are uniformly expressed while others are expressed in a tissue specific manner, particularly, in testis and brain (Maehara et al., 2015; Talbert and Henikoff, 2010).

The variants of histone H3 include H3.1, H3.2, H3.3, H3.Y.1, H3.Y.2, H3.4, H3.5 and CENP-A. Of these, H3.1, H3.2 and H3.3 are the most abundant variants of histone H3 while H3.4 and H3.5 are testis specific variants. H3 variants - Y.1 and Y.2 are specific to primates (Gambogi and Black, 2019; Tachiwana et al., 2011; Wiedemann et al., 2010).

The variants of histone H3 have subtle amino acid sequence differences (**Figure 1.2A**), yet their incorporation into the chromatin profoundly impacts the molecular processes. For example, there is a high degree of structural similarity between the nucleosomes containing H3.1 and H3.3 histones (Martire and Banaszynski, 2020). Nevertheless, nucleosomes containing canonical H3.1 are more stable than those nucleosomes that contain H3.3 variant. The few amino acid differences also impact the histone chaperones that interact with histones. For example, H3.1 histones interact specifically with chromatin assembly factor 1 (CAF1) whereas H3.3 interacts with histone regulator A (HIRA) chaperone. The H3.1 histones which are only expressed during S-phase are incorporated during DNA replication and such nucleosomes are largely found at repressed chromatin regions that are enriched in histone

and DNA methylation. In contrast, the variant H3.3 are expressed throughout the cell cycle and nucleosomes containing H3.3 histones are largely distributed to active chromatin elements like enhancer, promoters and gene bodies (Conn et al., 2013; Martire and Banaszynski, 2020; Stroud et al., 2012). H3.3 histones are thus typically associated with the euchromatin. Similarly, the histone post-translational modification, H3K4me3, is more enriched among nucleosomes containing H3.3 variants than among histone H3.1 containing nucleosomes, further illustrating the molecularly distinct features that subtle differences in amino acid sequence in histone variants could yield (Maehara et al., 2015; Tagami et al., 2004; Wong et al., 2010; Xiong et al., 2018).

Similar to histone H3, other core histones also have variants that diversify the chromatin contexts and play specific roles in chromatin related processes. All the known histone variants and their domains are illustrated in Figure 2.2, however, detailing all of them would be beyond the scope of this thesis.



**Figure 1.2: Core histones and their variants**

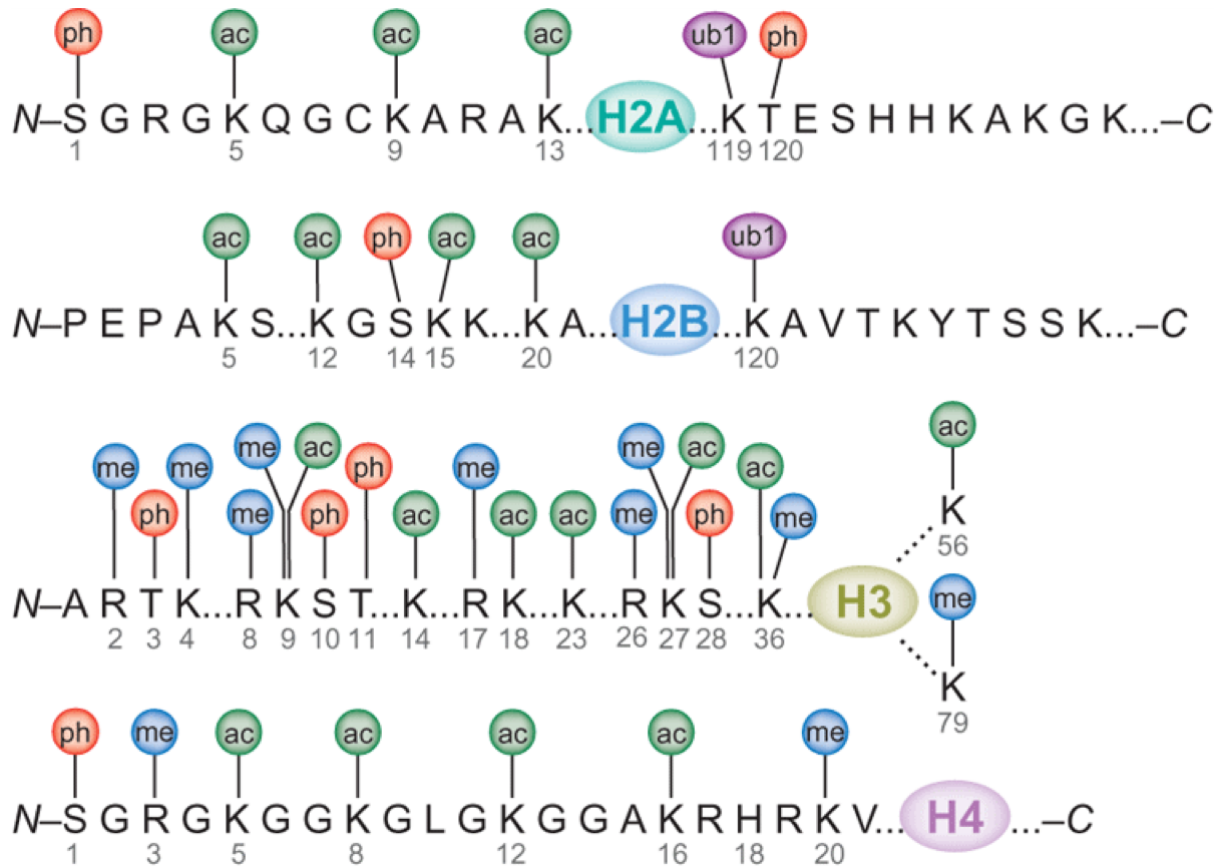
Illustration of core histone proteins and their variants. The histone chaperone or chromatin remodeler complex associated with a particular histone variant is indicated. Colour shades of histone variants depict differences within domains relative to replication-coupled histone.

## 1.2 Chromatin modifications and gene expression regulation

The chromatin is maintained in a dynamic state where its structure is modified such that the gene expression can be regulated. This is achieved by chromatin modifying enzymes that post-translationally modify the histone globular domains or the histone tails that protrude out of the nucleosome; and by ATP-dependent chromatin remodelers. These remodelers use the hydrolysis of ATP to drive the changes in the conformation of the chromatin.

### **1.2.1 Histone modifications**

As mentioned above, histones are modified through covalent post-translational modifications (PTMs) (Alfrey et al., 1964). The highly basic N-terminal tails of the histones that protrude out of the nucleosomes make contacts with other nucleosomes in their vicinity. This influences the nucleosome-nucleosome interactions, and thereby, the chromatin structure. In addition to regulating the chromatin structure, the histone PTMs also influence the recruitment of certain chromatin modifying enzyme complexes including ATP-dependent chromatin remodelers and non-enzymatic proteins. Together, PTMs and chromatin modifying complexes influence the overall dynamic nature of the chromatin, its structure and the regulation of gene expression. Histone PTMs such as acetylation, phosphorylation, methylation and ubiquitination have been well studied (**Figure 1.3**). Other histone modifications that are beginning to be of interest include SUMOylation, deamination,  $\beta$ -N-acetylglucosamine, ADP-ribosylation, and several lysine modifications that have been discovered recently including its propionylation (Huang et al., 2014; Kebede et al., 2017; Sakabe et al., 2010). Such proteins that catalyse the deposition of histone modifications are termed ‘writers’ while the proteins that remove these modifications are called ‘eraser’ proteins. The proteins that recognize and bind to the histone PTMs are called ‘readers’ (Bannister and Kouzarides, 2011; Ruthenburg et al., 2007).



**Figure 1.3: Histone post-translational modifications.**

Histone are acetylated (ac), phosphorylated (ph), methylated (me) and ubiquitinated (ub1), besides other modifications. Positions of the amino acids modified are denoted by numbers. From (Bhaumik et al., 2007).

### 1.2.1.1 Histone acetylation

The  $\epsilon$ -amino group of the lysine side chain in histones are acetylated by the transfer of an acetyl group ( $\text{CO-CH}_3$ ) from the acetyl coenzyme A (acetyl-CoA) co-factor (Wahab et al., 2020; Xhemalce et al., 2011). This process is catalysed by a group of enzymes called histone acetyl transferases (HATs). HATs can be broadly classified into two categories – Type-A and Type-B. The Type-A HATs are mostly nuclear and highly diverse. Based on their amino acid sequence homology and structure, they can be further classified into five families - GNAT, MYST, CBP/p300, general transcription factor HATs (e.g. TAF1) and Nuclear hormone-related HATs (e.g. SRC3) (Hodawadekar and Marmorstein, 2007; Xhemalce et al., 2011). The Type-A HATs are often present as part of multiprotein complexes and the other components of the complex influence their acetylation characteristics. For example, the purified yeast

GCN5 can acetylate free histones but is not able to acetylate those in a nucleosome (Bannister and Kouzarides, 2011). However, GCN5 present in the Spt-Ada-GCN5-Acetyltransferase (SAGA) complex can efficiently acetylate histones in a nucleosome. The HATs not only acetylate the N-terminal histone tails but also the globular core of the histones. For example, the human GCN5 acetylates histone H3 at lysine 56 (H3K56) residue in the globular core (Grant et al., 1997; Miller and Grant, 2013; Tjeertes et al., 2009).

The Type-B HATs, are mostly cytoplasmic, highly conserved and are involved in the acetylation of the newly synthesized histone H4 at lysine residues 5 and 12 (H4K5ac and H4K12ac) before their deposition into the nucleosomes (Parthun, 2007).

The histones are rich in positively charged amino acids like lysine (K) and arginine (R) imparting a high affinity for the negatively charged DNA. The acetylation diminishes the histone-DNA interaction by neutralizing the positive charge inherent to histones, thereby, influencing the chromatin structure and conformation (Yang and Seto, 2007). The conformational changes in chromatin upon histone acetylation allow the transcription machinery access to its targets. Therefore, histone acetylation is generally associated with gene expression. A vivid example of this is provided by genes related to neuronal activity and memory: neural activity is accompanied by increase in acetylation levels of memory related genes so that the expression of certain genes is sustained for long-term memory and synaptic plasticity. Reduction in acetylation levels, therefore, have been proposed to be the underlying causes of several neurodegenerative diseases including Alzheimer's disease and Rubinstein-Taybi syndrome (Gräff and Tsai, 2013; Kandel, 2001).

The acetylation by HATs is reversed by another set of enzymes called histone deacetylases (HDACs) which restore the positive charge on the histones and thereby, the affinity toward the DNA, tightening the chromatin. Thus, HDACs are generally repressors of transcription. In some cases of neurodegenerative diseases generated by the reduction of histone acetylation levels as mentioned above, a gain-of-function of HDACs has been found to be one of the underlying causes. Therefore, HDACs are therapeutically targeted by small molecule inhibitors called HDAC inhibitors (HDACi), to restore the acetylation levels necessary for memory and other neural activity. HDACs can be classified into four classes. The Class I, II and IV HDACs, which have a zinc metal ion dependent mode of action that does not involve the use of a cofactor (De Ruijter et al., 2003). On the other hand, the class III HDACs which are termed sirtuins, employ a NAD<sup>+</sup>-dependent catalytic mechanism (Saunders and Verdin,



2007). HDACs generally have a low substrate specificity, therefore, one HDAC enzyme has the capability to deacetylate several sites in histones. Often one HDAC enzyme is found within multiple distinct complexes, sometimes, working in cohort with other HDAC enzymes. For example, HDAC1, along with HDAC2, is present in NuRD, Sin3a and Co-REST complexes (Yang and Seto, 2007, 2008).

HATs and HDACs have an opposite effect on the acetylation levels in the genome and thereby, on the transcription of multiple genes related to signal transduction, cell growth and cell death. Therefore, a fine balance between the two enzyme families is important for normal acetylation levels and genome stability. A loss in the balance has been associated with several cancers (Ceccacci and Minucci, 2016). Therefore, over the last few years both these enzyme families have been targeted in cancer therapies. However, the efficacy of HDACi on solid tumors has been challenged. Therefore, new strategies involve use of combinational therapies, that is, HAT or HDAC inhibitors in combination with other anti-cancer drugs (Francesco Fiorentino and Rotili, 2020; Park and Kim, 2020).

### **1.2.1.2 Histone Phosphorylation**

Histone phosphorylation was first reported nearly 40 years ago. The phosphorylation takes place majorly on the serine, threonine and tyrosine residues in the N-terminal histone tails (Shoemaker and Chalkley, 1978). The levels of phosphorylation are regulated by two sets of enzymes – kinases and phosphatases. The kinases are responsible for phosphorylation while the phosphatases remove this mark (Oki et al., 2007). During phosphorylation, a phosphate group from an ATP is transferred to the hydroxyl group of the substrate amino acid side chain. This leads to the addition of negative charge to the histone. Thereby, just like acetylation, phosphorylation weakens the histone-DNA contacts and influences the local chromatin structure (Rossetto et al., 2012). Importantly, phosphorylation provides a recruitment scaffold for other proteins including chromatin remodeling proteins or proteins that recruit chromatin remodelers, to help alter the chromatin conformation based on the cellular context. For example, during DNA damage response, the histone H2A.X is phosphorylated at its residues 139 and 142. These phosphorylation marks are recognized by microcephalin 1 (MCPH1) and is recruited to the damage sites. MCPH1, in turn recruits SWI/SNF chromatin remodeling complex that helps in DNA repair (Sawicka and Seiser, 2014).

Further, the phosphorylation of histones has been reported to be essential for mitosis, apoptosis, DNA repair, replication and transcription (Xhemalce et al., 2011). The most extensively studied histone phosphorylation is on serine 10 of histone H3 (H3S10) which is catalysed by the Aurora B kinase. This mark becomes detectable first during the late G2 phase in the pericentric heterochromatin. Later, it spreads across the entire chromosome arm (Hsu et al., 2000). Several studies have demonstrated the direct correlation of chromatin compaction, during mitosis and meiosis, and H3S10 phosphorylation (H3S10ph). This makes H3S10ph a mark for these cellular processes (Hans and Dimitrov, 2001). During anaphase of the cell cycle, protein phosphatase 1 (PP1) dephosphorylates H3S10 and this is of critical importance for proper chromosome segregation (Hsu et al., 2000). While the H3S10ph plays pivotal roles during cell division in chromosome compaction and segregation, it also is important for transcriptional activation and cell growth. Therefore, mis-regulation of this mark (H3S10ph) has been identified in several cancers and associated with poor prognosis. In fact, higher levels of H3S10ph have been associated with invasive breast cancers, gastric carcinoma and glioblastoma among others (Khan, 2015; Komar and Juszczynski, 2020). The kinases that regulate H3S10 phosphorylation including aurora B and mitogen- and stress-activated protein kinase 1 (MSK1), have therefore, been targeted to treat cancers (Bavetsias and Linardopoulos, 2015; Khan et al., 2016).

### **1.2.1.3 Histone methylation**

The histone methylations have been reported to occur on the side chains of arginine, lysine and histidine residues. One major difference between this modification and the earlier mentioned acetylation or phosphorylation modifications, is that, the methylation does not change the charge of the histones (Lan and Shi, 2009; Ng et al., 2009). The histone methylation marks are deposited by enzymes called histone methyltransferases (HMTs). HMTs catalyse the transfer of one, two or three methyl-groups from S-Adenosylmethionine (SAM) to the lysine residues in histones. Arginine residues as well can be mono- or dimethylated and such methylations can be symmetrical or asymmetrical (Bedford and Clarke, 2009). In the core histones, there are over twenty possible histone methylation sites and given the multiple methylation levels, this modification is the most complex of all the histone PTMs (Miller and Grant, 2013). The HMTs that methylate lysine residues (HKMTs) largely deposit the mark on N-terminal histone tails. Most HKMTs possess a SET domain which harbours the catalytic activity. One example for a HKMT that lacks a SET domain is the Dot1 enzyme

that methylates the histone H3 globular core at lysine 79 (H3K79) (Bannister and Kouzarides, 2011). Of the several histone methylation sites, the methylation of Histone H3 at lysine 4 (H3K4) is associated with active transcription (Wang et al., 2008). At enhancers, the mono-methylation of H3K4 is highly enriched. At the body of actively transcribed genes, di-methylation of H3K4 is enhanced. The tri-methylation of H3K4 has been reported to be enriched in the promoters, near the transcription start site (TSS) of the actively expressed genes (Pokholok et al., 2005; Santos-Rosa et al., 2002; Wang et al., 2005). The recruitment of some HMTs that methylate H3K4 is driven by RNA polymerase II. The RNA pol II is phosphorylated by TFIIH at serine-5 of its C-terminal domain (CTD) poising the RNA Pol II for active transcription. It is this poised RNA Pol II that recruits the Set1- containing H3K4 HMT, COMPASS, via RNA Pol II bound Paf1 complex (Bae et al., 2020; Ng et al., 2003; Phatnani and Greenleaf, 2006; Wood et al., 2003). Set1 is released when RNA Pol II is phosphorylated at serine 2 of its CTD during transcription (Martin and Zhang, 2005). Another histone methylation mark, H3K36me, is associated with active transcription. However, in contrast to H3K4 methylation, this mark (H3K36me) is enriched at the 3'- end of genes. It is catalysed by Set2 HMT which is recruited by serine-2 phosphorylated CTD as RNA Pol II approaches the 3' end of genes (Fong et al., 2017; Sun et al., 2020)).

Unlike, methylation marks at H3- K4, K36 and K79, the methylation mark at K27 is a repressive mark. It is catalysed by the Set-domain containing Enhancer of Zeste 2 (EZH2) component of the Polycomb repressive complex 2 (PRC2) (Diehl et al., 2019). The transcriptional repression mediated by the methylation of H3K27 is crucial for X-chromosome inactivation (Inoue et al., 2017), suppression of homeotic genes and imprinting of genes at pericentric heterochromatin (Martin and Zhang, 2005; Wiles and Selker, 2017).

The methylation at H3K9 is the most extensively studied PTM. Mono-methylation at H3K9 (H3K9me1) is catalysed by HMT1C/G9a and is found at the 5'-UTRs of genes (Miller and Grant, 2013). Although, H3K9me1 has been implicated in the establishment of functional heterochromatin, not much is known about its role (Rivera et al., 2015). H3K9me2/3, on the other hand, have been extensively studied and are enriched at pericentromeric, sub-telomeric and gene-desert regions. H3K9me2 co-localizes with Lamin B1 and is largely associated with inactive chromatin (Rosenfeld et al., 2009). Similarly, HP1 $\alpha$  and HP1 $\beta$  interact with H3K9me3 through their chromodomains to establish pericentromeric heterochromatin indicating the importance of histone modifications in chromatin organization (Hyun et al., 2017).

Arginine residues are methylated by enzymes that can be classified into two types, type-I and type-II HMTs or Protein Arginine Methyltransferases (PRMTs). The type-I PRMTs are responsible for arginine mono-methylation (Rme1) and arginine asymmetric di-methylation (Rme2as). On the other hand, the type-II PRMTs lead to Rme1 and symmetrical di-methylation (Rme2s). These two PRMT families comprise a total of 11 members. All the PRMTs catalyse the transfer of a methyl group from SAM to the  $\omega$ -guanidino nitrogen atoms of the arginine side chain (Bannister and Kouzarides, 2011; Bedford and Clarke, 2009).

Like other histone PTMs, methylation is maintained in a dynamic state and the methylation marks are reversed by enzymes called histone demethylases. The demethylation by lysine-specific demethylases, for example, LSD1, which catalyses the demethylation of lysine residues using flavin adenine dinucleotide (FAD) as a cofactor is well understood (Shi et al., 2004). However, the demethylation at arginine residues is poorly understood. Two enzymes, peptidyl arginine deiminase 4 (PADI4) and JumonjiC-domain containing 6 (JMJD6) were the only proteins known to be involved in arginine demethylation (Rakow et al., 2020). PADI4 demethylates mono-methylated arginine residues by conversion to citrulline. However, it does not act upon di-methylated arginine residues. The JMJD6, which was initially reported as a possible arginine demethylase, has turned to be a lysine hydroxylase (Chang et al., 2007; Rakow et al., 2020). Well established lysine demethylases, KDM4E and KDM5C, have recently been reported to also catalyse the demethylation of arginine residues *in vitro* (Rakow et al., 2020). However, both these enzymes demethylate arginine residues poorly compared with demethylation of lysine residues *in vitro* and this function has not been established *in vivo* yet (Bonnici et al., 2018; Rakow et al., 2020; Walport et al., 2016). Recently though, another lysine demethylase, KDM3B, was reported to act at the promoters of hematopoietic genes and mediate the demethylation of H4R3me2s and its intermediate H4R3me1 (Li et al., 2018; Rakow et al., 2020).

It is pertinent to mention that besides their *bona fide* functions, histone modifications cross-talk with other histone modifying enzymes to regulate epigenetic processes. For example, the acetyltransferase activity of GCN5 at the tail of H3 histone is stimulated by the prior deposition of phosphorylation marks at H3S10 (Lee et al., 2010).

## **1.2.2 ATP-dependent chromatin remodeling enzymes**

As mentioned in section 2.1, the organization of DNA into chromatin with regularly interspaced nucleosomes allows tight packing of this long molecule into the nucleus. However, it renders the DNA sequences that are important for gene expression, DNA replication, repair and other processes, inaccessible to the regulatory proteins. Therefore, to circumvent this nucleosomal obstruction, the cell possesses specialized proteins that use the energy derived from the hydrolysis of adenosine triphosphate (ATP) to slide, reconstitute or eject the nucleosomes, such that the chromatin is maintained in a dynamic fashion. These proteins are called ATP-dependent chromatin remodelers (Narlikar et al., 2013). Not only do the ATP-dependent chromatin remodelers maintain proper density and spacing of the nucleosomes, but several of these remodelers work together with site-specific transcription factor and histone modification enzymes (Clapier et al., 2017). In recent years, several cancers have been found to be driven by mutations in certain ATP-dependent chromatin remodelers, highlighting the importance of their role in biological processes (Garraway and Lander, 2013; Wilson and Roberts, 2011). Based on their phylogenetic and functional characteristics, the ATP-dependent chromatin remodelers can be classified into four main families: imitation switch (ISWI), switch/sucrose non-fermentable (SWI/SNF), INO80 and chromodomain helicase DNA-binding (CHD) (**Figure 1.4A**). Among these subfamily members, there are common features like general affinity for nucleosomes and the presence of a Snf2-type ATPase motor. However, there are several differing aspects including the domains that confer specificity for the recognition of histone modifications and interactions with various proteins (Manelyte and Langst, 2013).

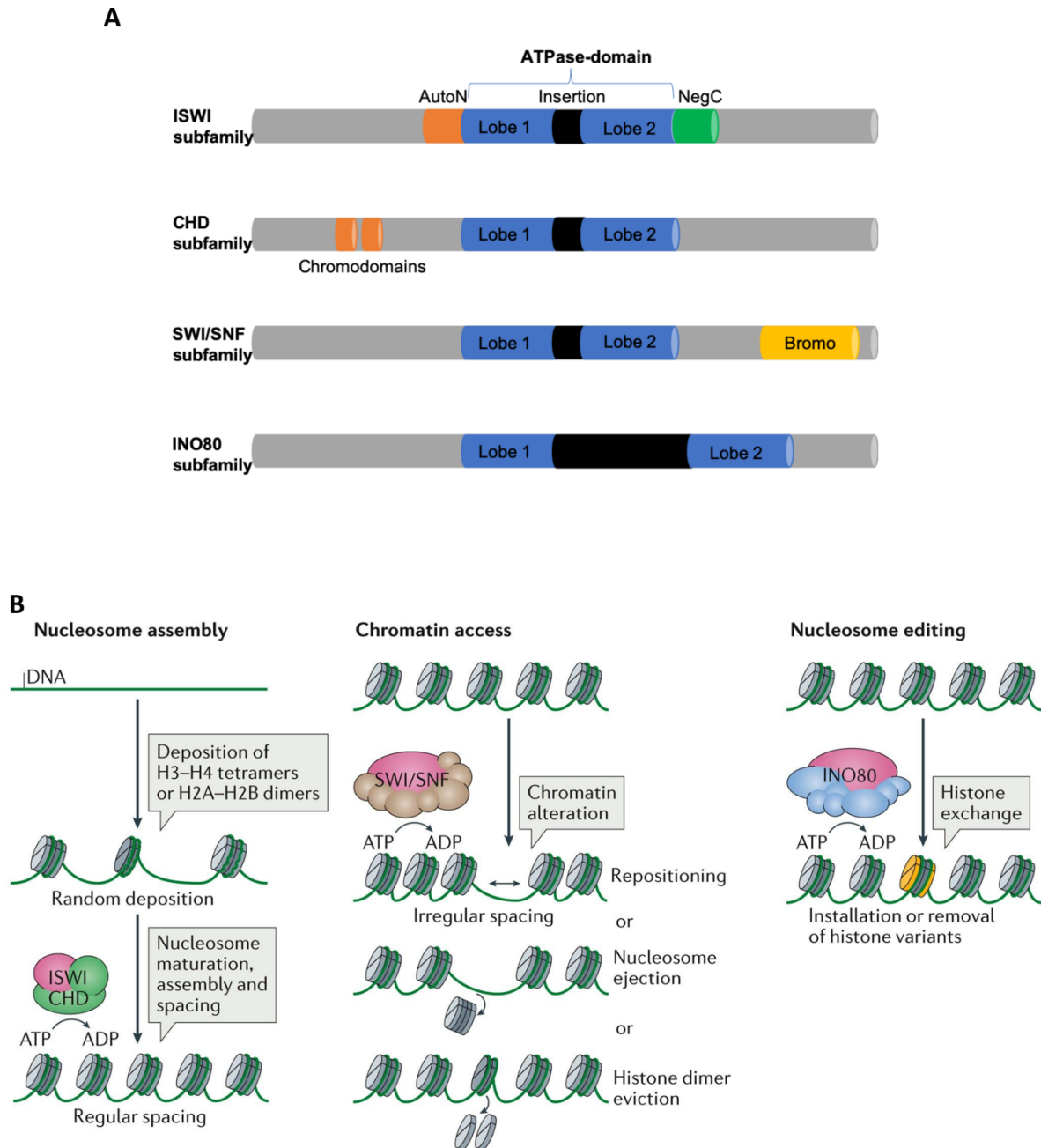
### **1.2.2.1 ISWI family**

The *Drosophila* Imitation Switch (ISWI) ATP-dependent chromatin remodeler was first identified on the basis of its sequence similarity to the *Drosophila* brahma and its mechanism of action has been extensively studied. Structurally, the ISWI remodelers contain two RecA-like lobes in their ATPase domain that are separated by a small insertion sequence. It is this domain that hydrolyses the ATP. The C-terminal region possesses a HAND-SANT-SLIDE (HSS) domain which helps binding to unmodified histone H3 tail and the linker DNA flanking the nucleosome (Grüne et al. 2003; Boyer et al. 2004). The ATPase lobes are flanked on the N-terminal side by AutoN (autoinhibitory N-terminal) and on the C-terminal side by NegC

(negative regulator of coupling) domains that regulate the enzymatic activity of the ATPase domain (Clapier and Cairns, 2012).

In *Drosophila*, ISWI is part of three multi-protein complexes: NURF (Nucleosome Remodeling Factor), ACF (ATP-dependent Chromatin assembly and remodeling Factor) and CHRAC (Chromatin Accessibility Complex) (Ito et al., 1997; Tsukiyama and Wu, 1995; Tsukiyama et al., 1995; Varga-Weisz et al., 1997). Some of the ISWI complexes have been reconstituted *in vitro* and their functional characteristics dissected. During *in vitro* remodeling assays with mononucleosomes that are end-positioned, that is, DNA wrapped around a core histone octamer on one end with another end of the DNA protruding freely, ISWI containing complexes were shown to re-position the histone octamer by sliding it to the centre of the DNA (Zhou et al., 2016). ISWI complexes use this ability of sliding the nucleosomes *in vivo* as well to regulate gene expression. For instance, the ACF complex which contains a heterodimer of ISWI ATPase and Acf1, uses this sliding ability to generate arrays of evenly spaced nucleosomes that give rise to transcriptionally repressed heterochromatin (Ito et al., 1997; Wallrath and Elgin, 1995). Similarly, NURF, a multi-component complex, slides the nucleosomes in the promoter region of genes to make the transcription factor binding sites accessible and facilitate gene expression (Badenhorst et al., 2002; Tsukiyama and Wu, 1995). Furthermore, the CHRAC complex that includes ISWI, ACF1 and the two subunits CHRAC-14 and -15, utilizes the hydrolysis of ATP, to assemble nucleosomes from free DNA and histones *in vitro* and to contribute to repressive chromatin *in vivo* (Scacchetti et al., 2018; Varga-Weisz et al., 1997).

ISWI remodeler containing complexes have been implicated to play multiple roles in transcription, chromosome organization and DNA replication (Corona and Tamkun, 2004; Elfring et al., 1994). A recent high-resolution cryogenic electron microscopy (cryo-EM) structure of yeast ISWI in complex with nucleosomes has shed new light on the mechanism through which ISWI remodelers act. It suggests that in the absence of nucleosomes, ISWI remodeler remains inhibited and once it comes in contact with nucleosomes, there is a conformational change that allows it to proceed with its remodeling activity. In fact, the mechanism of action has revealed that it causes a local DNA distortion and translocation, a mode of action that is similar to snf-2 type remodelers (Yan et al., 2019).



**Figure 1.4: Subfamilies of ATP-dependent chromatin remodelers and their roles.**

**(A)** Representation of domain organization of ATP-dependent chromatin remodeler subfamilies. **(B)** Functional classification of remodelers. Adapted from (Clapier et al., 2017).

### **1.2.2.2 SWI/SNF family**

The SWI/SNF (SWItch/Sucrose Non-Fermenting) ATP-dependent chromatin remodeling complex, was first identified in defective mutants of *S. cerevisiae* that were unable to ferment sucrose (Neigeborn and Carlson, 1984). Later in 1997, (Wu and Winston) reported that this defect is caused by the failure of the Snf2 mutant strains to remove the nucleosomes from the promoter of the SUC2 gene. This failure was attributed to the mutations in their snf2-type ATPase domain impairing their remodeling activity. Therefore, the transcription factor binding sites remained inaccessible and hence the gene responsible for proper sucrose fermentation, SUC2, remained transcriptionally repressed.

Structurally, the SWI/SNF remodeler ATPase also has two RecA-like lobes that are separated by an insertion sequence. In the C-terminal region, there is a bromodomain that is unique to SWI/SNF remodelers and helps them to interact specifically with acetylated histones and hence the transcriptionally active chromatin (Gaspar-Maia and Sevilla, 2016).

SWI/SNF containing complexes are highly conserved from fungi to plants and animals. They have been shown to play important roles in almost every nuclear process including gene expression regulation, DNA replication initiation, and homologous recombination (Euskirchen et al., 2012). The mammalian SWI/SNF complex contains 15 subunits. Some are core subunits that are required for basic remodeling functions and others are accessory subunits that are cell type specific and help in modulating the SWI/SNF functions (Alver et al., 2017). In *Drosophila*, the SWI/SNF-related Brahma (BRM) complex has two sub-types – BAP (Brahma associated factors) and PBAP (Polybromo-associated Brahma). These two sub-complexes share some subunits including the brahma ATPase subunit. However, they also have sub-complex specific components. For example, BAP complex specifically has OSA component, a DNA binding protein that has target gene specificity and helps in the transcriptional activation of its target genes. Similarly, the PBAP sub-complex has two specific subunits – polybromo and BAP170. The sub-type specific components modulate the behaviour and localization of the BRM complexes. For example, on polytene chromosomes, OSA and polybromo immune-staining showed both overlapping as well as distinct localizations suggesting that to regions where distinct localization patterns were observed, BRM complex might be recruited by separate mechanisms. (Collins et al., 1999; Mohrmann et al., 2004). Functionally, OSA-containing BAP complex was shown to specifically regulate



the differentiation of *Drosophila* intestinal stem cells (ISCs) to enterocytes while the PBAP complex specifically regulates the maintenance of *Drosophila* ovarian germline stem cells (He et al., 2014; Zeng et al., 2013).

Like ISWI remodelers, SWI/SNF remodelers also possess the capability to slide mononucleosomes *in vitro*. However, in contrast to the ISWI, SWI/SNF remodelers slide the nucleosomes towards DNA ends and cannot space the nucleosomal arrays evenly (Zhou et al., 2016). SWI/SNF complexes use this remodeling ability in diverse biological processes such as lineage specificity and the genome stability (Alver et al., 2017; Ribeiro-Silva et al., 2019). Therefore, any mis-regulation of SWI/SNF complex leads to widespread effects and may for example, cause cancer. In fact, almost 25% of all cancers have mutations in subunits of SWI/SNF complexes. Therefore, SWI/SNF complexes are currently being targeted for therapeutic interventions in several cancers (Centore et al., 2020; Mittal and Roberts, 2020).

### **1.2.2.3 INO80 family**

The INO80 ATPases were first discovered as transcriptional regulators of inositol responsive genes in *S. cerevisiae* (Ebbert et al., 1999). The two RecA-like lobes in the ATPase domain of INO80 are separated by a large insertion sequence which can vary from about 250 amino acids in yeast to over 1000 amino acids in mammals. In yeast, this insertion sequence has been shown to bind the helicase-related AAA+ type ATP-binding protein chaperones Rvb1p and Rvb2p. These chaperones are components of the INO80 complex and also are essential in the assembly of a functional INO80 complex (Jha and Dutta, 2009; Jónsson et al., 2004; Watanabe et al., 2015).

In yeast there are two INO80 ATPase containing complexes – INO80 complex and SWR1 complex. In *Drosophila*, INO80 exists in INO80 complex and the SWR1 complex homolog Tip60/p400 complex. The dINO80 complex is a multi-subunit complex. The dINO80 complex has also been shown to interact with the DNA binding Pleiohomeotic (Pho) protein. The Pho component of this complex identifies the Polycomb responsive elements (PRE) and recruits the complex to regulate the transcription of homeotic genes that control the proper segmentation and patterning in *Drosophila* (Bhatia et al., 2010). However, dINO80 has been shown to function independent of Pho protein as well. For example, the Pho null mutants in *Drosophila* die at pharate adult stage. On the other hand, dINO80 null mutants die at the late

embryonic stage. This suggests that in addition to cooperative function, these two proteins also have their independent functions (Ghasemi et al., 2015). The *Drosophila* Tip60/p400 complex contains the *Drosophila* homolog of SWR1 protein called Domino. Recent studies have shown that the ATPase containing remodeler Domino (Dom) has two splice isoforms – Dom A and Dom B. The Dom B isoform containing complex functions as the *Drosophila* homolog of SWR1 complex and performs ATP-dependent chromatin remodeling. The Dom A complex functions in an ATP-independent manner as an acetyltransferase, similar to the yeast NuA4 complex (Scacchetti et al., 2020).

Similar to ISWI ATPases, INO80 ATPases can slide the end-positioned mononucleosomes to the central position *in vitro*. In fact, INO80 can remodel a tri-nucleosome array such that the nucleosomes are evenly spaced (Udugama et al., 2011). This nucleosome sliding ability of INO80 has been shown to play crucial roles in gene expression regulation, DNA repair and replication (Eustermann et al., 2018). The actively transcribed genes have nucleosome free regions (NFRs) in their promoters such that the transcription machinery can access the promoters. However, to set up the boundaries of the NFR, there are two nucleosomes that are positioned both upstream as well as downstream of the transcription start site. These nucleosomes are called -1 and +1 nucleosomes, respectively. The correct positioning of these nucleosomes is important for transcription as they regulate the binding of transcription factors as well as the passage of the RNA polymerase II through the gene body. INO80 remodelers have been shown to be crucial for the correct positioning of these canonical -1 and +1 nucleosomes (Brahma et al., 2017; Lai and Pugh, 2017). Besides this, INO80 is critical for the global genome distribution and exchange of histone variant H2A.Z and genome integrity (Papamichos-Chronakis et al., 2011).

### **1.2.2.4 CHD family**

The first member of the Chromodomain-helicase DNA binding (CHD) family of ATP-dependent chromatin remodeling ATPase was identified in mice and named CHD-1. Subsequently, further family members were identified in other organisms including humans and *Drosophila* (Delmas et al., 1993; Platt et al., 2013). Structurally, CHD family members possess tandem chromodomains in their N-terminal part. The RecA-like lobes in the ATPase motor, just like in ISWI and SWI/SNF remodelers, are separated by a short insertion sequence. Based on the presence or absence of additional domains, this family of

remodelers is divided into three subfamilies. In *Drosophila*, the CHD family of proteins is comprised of four members, that is, dCHD1, dCHD3, dMi-2 and Kismet (Bouazoune and Brehm, 2006).

The first CHD subfamily (subfamily 1) has a DNA binding domain in the C-terminal region that preferentially binds to AT-rich DNA sequences in *in vitro* binding assays. This subfamily includes yeast Chd1 (*S. cerevisiae*), yeast Hrp1 and Hrp3 (*S. pombe*) and CHD1 and CHD2 in higher eukaryotes. The chromodomains in the human Chd1 interact with the H3K4me3 mark and may help its recruitment to active genes. However, the *Drosophila* Chd1 does not seem to be recruited through this interaction as deletion of its chromodomains does not perturb its localization on polytene chromosomes (Murawska and Brehm, 2011). In yeasts, the Chd1 has been shown to co-localize with the transcription elongation factors at actively transcribed regions. However, its deletion does not yield a strong phenotype, perhaps because there are other chromatin remodelers that compensate for its loss (Simic et al., 2003). Nevertheless, recent data suggests that the loss of Chd1 in yeasts specifically affects the intron containing and highly transcribed genes. Chd1 loss leads to severe perturbation in the nucleosomal arrangement at highly transcribed genes, particularly the +1 and +2 nucleosomes. This indicates a role of Chd1 in nucleosome organization. Similarly, the loss of Chd1 leads to disruption of the serine-5 phosphorylated form of RNA pol II at highly transcribed genes, even though, the overall RNA Pol II occupancy remained unaffected (Park et al., 2014). Likewise, in *Drosophila*, the loss of Chd1 did not yield any dramatic phenotype such that the flies would die. However, its loss leads to defects in fertility and wing development indicating its non-redundancy in these aspects (McDaniel et al., 2008).

Apart from its role in transcription, Chd1 is essential for the transcription-independent incorporation of H3.3 into the *Drosophila* genome and chromatin organization (Morettini et al., 2011). In fact, Chd1 also plays a role in nucleosome assembly and its loss or overexpression in the salivary gland cells of third instar larvae affects the higher order structure of the polytene chromosomes. The loss of Chd1 leads to increase in the expression of heterochromatin protein HP1a while the overexpression of Chd1 causes the reciprocal effect, leading to alterations in chromosome structure (Bugga et al., 2013).

*In vitro*, Chd1 has been shown to slides the nucleosomes bi-directionally, that is, both towards, as well as, away from the DNA ends. It can also evenly space an array of nucleosomes. To investigate the details of the Chd1 remodeling process, its crystal structure

was solved which showed that in the absence of nucleosomes, the chromodomains inhibit the catalytic activity of Chd1 by contacting and thereby, occluding both of its ATPase lobes. Also, the DNA binding surface within the ATPase domain remains inaccessible as an acidic-helix, the so-called chromo-wedge domain, packs against it. In presence of nucleosomes, a conformational switch takes places that releases the chromodomains from the ATPase lobes, thereby, making the ATPase domain accessible and ready for efficient hydrolysis of ATP and remodeling reaction (Hauk et al., 2010).

The second subfamily of CHD proteins (subfamily 2) includes CHD3 and CHD4 (Mi-2 $\alpha$  and Mi-2 $\beta$ , respectively). Unlike the subfamily 1 proteins, they do not possess a DNA binding domain, instead they have PHD (Plant Homeodomain) fingers in the N-terminal region. These PHD fingers are the readers of the histone PTMs, particularly, binding to the methylated lysine in histone H3 (H3K4). They also bind to the unmodified histone H3 tails (Jain et al., 2020). The CHD5 protein may also be classified in this subfamily. Although, there is a discrepancy in its classification in either subfamily 2 or 3 because like CHD3 and CHD4, it possesses PHD fingers. However, unlike CHD3/4 and like subfamily 3 proteins, it possesses a SANT domain in its C-terminal region (Sims and Wade, 2011; Tyagi et al., 2016). Irrespectively, all these subfamily 2 members are alternative components of the Nucleosome Remodeling and Deacetylase (NuRD complex, see section 2.2.3) (Tencer et al., 2017). The two PHD fingers present in CHD3 protein have been shown to interact with methylated and acetylated forms of histone H3 tails at lysine 9 residue (H3K9). This ability of CHD3 to sense both methylated as well as acetylated histone tails gives the NuRD complex the ability to function in a dynamic fashion. It allows NuRD operation at H3K9me<sub>3</sub>-containing transcriptionally suppressed chromatin and H3K9ac-containing transcriptionally active chromatin. Perhaps, at the H3K9ac loci, the CHD3 mediated NuRD recruitment allows deacetylation through its HDAC components to control gene expression. Also, CHD3 driven recruitment of NuRD at H3K9me<sub>3</sub> loci might mediate its role in heterochromatin formation (Tencer et al., 2017). The *Arabidopsis thaliana* homolog of CHD3 (PICKLE) associates with H3K27me<sub>3</sub> enriched genes and contributes to the H3K29me<sub>3</sub>-related processes (Zhang et al., 2012). In *Drosophila*, dCHD3 has been shown to act as a monomer and associate with the actively transcribed genes. However, not much is known about its function in *Drosophila*. In cells, its depletion does not yield a drastic phenotype with insignificant effects on cell growth. It appears similar to dMi-2 protein with about 77% sequence similarity. However, dCHD3 possesses only one PHD finger and lacks the C-terminal region downstream of the ATPase domain that is present in dMi-2 (CHD4) (Murawska et al., 2008). Nonetheless, dCHD3, can bind to DNA and

nucleosomes. Its ATPase activity is stimulated by DNA as well as nucleosomes and it remodels mononucleosomes bidirectionally, that is, both towards as well as away from the DNA ends (Murawska et al., 2008).

CHD4/Mi-2 was first identified in the serum of a dermatomyositis patient named Mitchell. It is a nucleosome stimulated ATPase and mainly takes part in the repression of gene expression through its association with the NuRD complex. It has been shown to be vital for proper lineage specification (Tyagi et al., 2016). *In vitro* CHD4 slides mononucleosomes away from the DNA ends. To investigate the mechanism through which CHD4 remodels chromatin, a crystal structure of human CHD4 (hCHD4) bound to nucleosomes was solved. It revealed that the hCHD4 binds to the nucleosomal DNA at super helical location (SHL)  $\pm 2$  and this causes a twist in the DNA. Subsequently, ATP binding then causes the closure of the ATPase motor and the propagation of the twist towards the dyad (see Figure 2.1A for nucleosome structure). Finally, the hydrolysis of ATP resets the ATPase motor and the remodeler can bind to the next DNA position (Farnung et al., 2020).

Like hCHD4, the *Drosophila* homolog, dMi-2 protein has two chromodomains (CD), also in the N-terminus. The chromodomains of other proteins such as dHP1 bind to methylated histone H3 tails and help in its recruitment to the chromatin. However, the dMi-2 chromodomains lack such a property but rather bind to DNA and this interaction has been proposed to be one of the 'early steps' in the remodeling process (Bouazoune et al., 2002; Jiang et al., 2009). Unlike dCHD3, dMi-2 has two PHD fingers which have been shown to read and bind to several PTM marks including acetylation and methylation. They can also bind unmodified histone tails. This binding is fundamental for gene expression regulation and nucleosome remodeling (Musselman and Kutateladze, 2011). Like hCHD4, dMi-2 also exists in NuRD complex (section 2.2.3), however, in addition to this it also exists in a dMEP-1 containing dMec complex. dMec represses pro-neural genes in non-neuronal cells and mediates the SUMOylation-dependent gene repression. In fact, dMec complex is the major dMi-2 containing complex in *Drosophila* embryos and in Kc cells (*Drosophila* embryo-derived cells) (Kunert et al., 2009; Stielow et al., 2008).

dMi-2 is important for somatic cell survival and germ cell development. Its interaction with Hunchback, a repressor of the HOX genes, has been implicated in the Polycomb-mediated gene repression of the homeotic genes (Kehle et al., 1998). Recent reports implicate dMi-2 in repressing hormone regulated genes in *Drosophila* (Kreher et al., 2017). Apart from its role

in gene repression, dMi-2 also facilitates gene expression. For example, dMi-2 plays an important role in the efficient transcription of the heat shock (HS) response genes. dMi-2 is recruited to HS genes through its interaction with poly-ADP ribose (PAR) polymers and nascent HS transcripts that are produced from the heat shock gene loci (Mathieu et al., 2012; Murawska et al., 2011).

Like hCHD4, recombinant dMi-2 is stimulated by nucleosomes and slides an end-positioned mono-nucleosome to the central position *in vitro* (Brehm et al., 2000). In fact, this property of mobilizing the nucleosomes allows dMi-2 to regulate the chromatin structure *in vivo*. For example, dMi-2 was shown to condense the polytene chromosomes and its over-expression or depletion disrupts their characteristic banding pattern (Fasulo et al., 2012). Mutations in the human homolog of dMi-2 have been associated with cancer. Recent findings suggest that analogous mutations disrupt the nucleosome remodeling function of dMi-2 by altering its ATPase activity, nucleosome -binding and -positioning. This indicates the importance of dMi-2 mediated chromatin remodeling for normal functioning of a cell (Kovač et al., 2018).

The third subfamily of CHD proteins (subfamily 3) include CHD6 to CHD9 proteins. These proteins contain SANT or SANT-like or SLIDE domains or Brahma/Kismet (BRK) domains. Therefore, this subfamily is highly diverse. The SANT and SLIDE-domains bind to DNA and mutations in these domains of CHD7 have been associated with CHARGE syndrome (Asad et al., 2016; Ryan et al., 2011; Tyagi et al., 2016). Chromatin remodelers in this family contribute to a wide range of processes including DNA damage response (Kargapolova et al., 2020), neural crest migration and differentiation (Asad et al., 2016) and gene expression regulation (Alendar et al., 2020; Wilkinson et al., 2015).

In *Drosophila*, the CHD protein containing a BRK domain is called Kismet. Kismet was first identified as a dominant suppressor of polycomb which represses homeotic genes in *Drosophila* (Daubresse et al., 1999). There are two isoforms of Kismet, one with a molecular weight of 574 kDa (kilo-Daltons) called KIS-L and another with a molecular weight of 225 kDa called KIS-S. The larger isoform contains an ATPase domain that catalyses ATP-dependent chromatin remodeling. Just like brahma, it possesses a 41 amino acid BRK-domain (BRM and KIS) in its C-terminal region which remains to be of unknown function. On polytene chromosomes, KIS-L co-localizes with RNA polymerase II (Pol II) on nearly all transcriptionally active sites. In fact, the loss of kismet leads to a dramatic reduction of elongating RNA Pol II and the elongation factor SPT6 and dCHD1 on polytene chromosomes. However, the

recruitment of RNA Pol II to the chromatin is not affected suggesting that KIS-L plays a role in the early steps of transcription elongation (Srinivasan et al., 2005). Recent studies have shown that Kismet plays an important role in activating the transcription of the steroid hormone receptor gene – ecdysone b1 (*ecr-b1*), by promoting H3K36 di- and tri-methylation and H4K16 acetylation (Latcheva et al., 2019).

### **1.2.3 Nucleosome remodeling and deacetylase (NuRD) complex**

As mentioned earlier, the chromatin is maintained in a highly dynamic fashion through compaction and decompaction by chromatin remodeling complexes allowing regulation of gene expression. One such complex combines the remodeling and deacetylation processes to give rise to nucleosome remodeling and deacetylase (NuRD) complex. NuRD is a macromolecular, highly conserved, multi-protein complex (Torchy et al., 2015; Zhang et al., 1999). In mammals, the NuRD complex consists of at least six components, with each component having at least two paralogs. These include:

- histone deacetylases – HDAC1 and HDAC2,
- the ATP-dependent remodeling proteins – CHD4 and its paralogs CHD3 and CHD5,
- the methyl CpG- DNA-binding domain proteins - MBD2 and MBD3, that can bind methylated and/or hydroxymethylated DNA,
- the metastasis associated proteins - MTA1, MTA2 and MTA3, that are involved in nucleosome recognition,
- the retinoblastoma binding proteins – RBBP4 and RBBP7, that can bind to histone tails,
- GATAD2A and GATAD2B, which are known to interact with MBD2 and MBD3 and with other CHD proteins, however, their precise functions remain elusive (Low et al., 2020).

NuRD complex was first described in 1998 by several groups reporting the coupled ATP-dependent remodeling and deacetylase complex (Tong et al., 1998; Wade et al., 1998; Xue et al., 1998; Zhang et al., 1998). The only other complex that combines two different chromatin regulating functions is Tip60/p400 complex. In Tip60/p400 complex, Tip60 is an

acetyltransferase which acetylates N-terminal tails of H2A and H4. The p400 exchanges H2A-H2B dimers for H2A.Z-H2B dimers within histone octamers (also see section 2.2.2.3) (Chen et al., 2013). One explanation proposed for such coupling of remodeling and deacetylase function is that the HDAC depends on ATP-dependent remodeling function for access to its targets. That is why, in the absence of ATP, only histone octamers can be deacetylated but not the nucleosomes (Torchy et al., 2015).

Depending upon the cellular, tissue, physiological and pathological contexts, the components of the NuRD complex could vary giving rise to several coexisting complexes. In mammals, the variety of NuRD complexes is increased by the isoforms of its components some of which can be mutually exclusive. For example, MTA1, MTA2 and MTA3 are mutually exclusive to each other as are the isoforms of MBD. To illustrate this, a NuRD complex is specifically found in mouse embryonic stem cells (mESCs) that contains MBD3C isoform of MBD, which uniquely has 50 amino acids in its N-terminal region allowing its interaction with WDR5, a histone H3 binding protein (Ee et al., 2017). This suggests that there is a great degree of heterogeneity of NuRD complexes with respect to all the alternative subunits and their cell-type-specific expression.

The stoichiometry of subunit components and the assembly of NuRD complex has been a long running debate. Reports have suggested that CHD4, the ATPase motor component, is not a stable subunit of the complex and associates with it only peripherally (Zhang et al., 2016). A recent study showed that, in mammals, a deacetylase module and a remodeling module form, which then combine to form the NuRD complex. The deacetylase module includes MTA, HDAC and RBBP components which have a stoichiometry of 2:2:4, respectively. The remodeling module includes MBD, GATAD2 and CHD components with a stoichiometry of 1:1:1, respectively (Low et al., 2020).

In *Drosophila*, the corresponding NuRD components are Mi-2 (CHD), RPD3 (HDAC), p55 (RBBP), MBD-like, MTA-like and Simjang/p66 (GATAD2). To measure the stoichiometry of the NuRD complex in *Drosophila* Schneider S2 cells, Ernest Laue and his colleagues performed label-free quantitative mass spectrometry based proteomic experiments. These revealed that MTA-like, Rpd3, MBD-like and p55 were present from single to several copies in different NuRD complexes whereas p66 (Simjang), Mi-2  $\alpha/\beta$  and Doc1 were sub-stoichiometric components. MTA1-like, MBD-like and Rpd3 were found in the ratio of 2:2:1, respectively (Zhang et al., 2016). This suggests that both in mammals as well as in *Drosophila*,



the deacetylation unit is at the core of the NuRD complex and the remodeling motor containing module at the periphery.

Functionally, NuRD complex was initially thought to be a *bona fide* transcriptional repressor because of its deacetylation activity. However, advances in the understanding of its role revealed that it plays a much more complex role in transcriptional regulation (Basta and Rauchman, 2015; Zhang et al., 1999). Genome-wide occupancy studies in mammalian cells showed that NuRD occupied most of the promoters and enhancers. At most of these loci, NuRD dampened the gene expression (Bracken et al., 2019). Soon it became clear that deacetylation was not the sole mechanism for NuRD to repress gene expression: NuRD increased the density of nucleosomes, thereby obstructing the transcription factors from binding their target sites, to represses gene expression at certain loci (Bracken et al., 2019). For example, a NuRD interacting transcription factor called Ikaros recruits NuRD to its target genes. At the target loci, NuRD induces a nucleosomal invasion resulting in eviction of RNA Pol II and a subsequent transcriptional repression. No deacetylation is required for this transcriptional shutdown, however, in the sequence of events, HDAC activity is later employed to maintain the target genes in a repressed state (Liang et al., 2017). Further, the genome-wide studies revealed association of NuRD with actively transcribed genes (Bornelöv et al., 2018). At the promoters of certain actively transcribed genes and regulatory regions, histone H3.3 is deposited. NuRD interacts with unmodified lysine residue 4 in histone H3.3 and this interaction was shown to be important for the expression of these genes (Kraushaar et al., 2018). Analysis revealed that NuRD played a role in fine-tuning of the expression of these genes. Thus, delineating from the earlier held assumption that suggested NuRD to be a canonical repressor of gene expression, recent models suggest NuRD to rather be a modulator of gene expression (Bracken et al., 2019).

One of the best examples of NuRD's fine-tuning role can be exemplified in the embryonic stem cells (ESCs) (Bornelöv et al., 2018). During the self-renewal state of ESCs, two naive pluripotency markers, Klf4 and Zfp42, are expressed while the primed pluripotency marker, Otx2, is repressed. For the differentiation to take place upon the removal of self-renewal factors, the Klf4 and Zfp42 genes are repressed while Otx2 is activated. NuRD complex is recruited to the enhancers of all the three genes in the self-renewal state. Upon removal of self-renewal factors, the cells that do not have NuRD complex recruited to the enhancers of these three genes, differentiate slower than wild type cells. The analysis revealed slower repression of Klf4 and Zfp42, and slower activation of Otx2 in the absence of NuRD on their

enhancers. There were distinct changes in the nucleosomal pattern of these enhancers upon loss of NuRD binding. These changes in the nucleosome patterning impaired the eviction of RNA Pol II from Klf4 and Zfp42 and recruitment of RNA Pol II at Otx2 genes. Hence, continued expression of pluripotency genes Klf4 and Zfp42 and impaired expression of primed pluripotency gene Otx2. This suggests that NuRD fine-tunes the gene expression at these loci by controlling the responsiveness of their enhancers (Bornelöv et al., 2018).

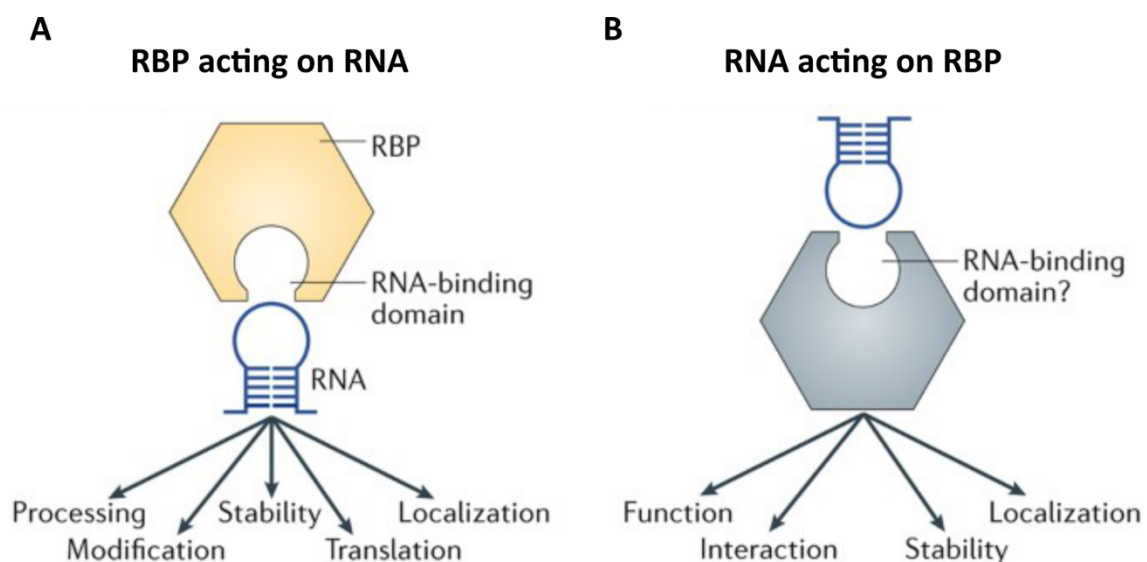
Besides ESCs, Mi-2 subunit of the NuRD complex has been shown to regulate the gene expression of the hormone regulated genes in *Drosophila*. In the presence of the hormone ecdysone, hormone responsive genes are upregulated and there is increased occupancy of dMi-2 at these genes. In presence of hormone, dMi-2 constrains the expression of these genes. Therefore, if dMi-2 is depleted, hormone leads to super-activation of ecdysone responsive genes. This supports the model that NuRD complex acts as a fine-tuner or a modulator of the gene expression (Kreher et al., 2017).

In addition to direct regulation of the gene expression, NuRD also facilitates the epigenetic regulation by other chromatin regulating complexes. For example, at certain pluripotency genes in ESCs, recruitment and trimethylation at H3K27 by polycomb repressive complex – 2 (PRC2) is dependent on the prior deacetylation of H3K27 by NuRD, allowing for a proper exit from pluripotency into lineage specificity during differentiation (Kim et al., 2015; Reynolds et al., 2012).

Owing to its critical role in gene expression regulation, several reports have implicated NuRD in cancer. The MTA1 component of NuRD has been found overexpressed in several cancers including breast, colorectal, oesophageal and prostate cancers. In fact, it has been noted that the higher the expression of MTA1 is, the more advanced is the grade of the tumor and the poorer is the prognosis (Lai and Wade, 2011). NuRD complex allows opening up of the chromatin at super enhancers in rhabdomyosarcomas, allowing binding of PAX3-FOXO1 fusion proteins which are drivers of tumorigenesis, leading to the downstream cascade of tumor signaling. In fact, loss of CHD4, the ATPase motor of the NuRD complex, has been shown to re-compact the chromatin architecture around super enhancers and depletion of PAX3-FOXO1 proteins (Marques et al., 2020). Furthermore, as mentioned in section 2.2.2.4, the mutations in CHD4 subunit of the NuRD complex have been implicated in several endometrial cancers implicating aberrant NuRD activities or mutations in its components in cancer.

### 1.3 RNA binding proteins

A large set of proteins have the ability to bind RNA and form a ribonucleoprotein (RNP) complex that have been shown to modulate the gene expression. Over 2000 such RNA binding proteins (RBPs) have been identified and the list is ever growing. RBPs identify a sequence or a structure in the RNA, often but not necessarily, through a well-defined structural motif called RNA binding domain (RBD). Several RBDs that have been extensively studied include RNA recognition motif (RRM), hnRNP K homology domain (KH domain) and DEAD-box helicase domain (Corley et al., 2020). However, several RBP complexes such as ribosome and spliceosome complexes lack such canonical RBDs indicating different mechanisms of protein-RNA interaction. RBPs have been termed as the ‘clothes’ of the mRNAs that help to cover or expose different regions, such as 5'- and 3'- untranslated regions (UTRs) and coding regions, within mRNAs, regulating their processing, stability, localization, translation and modification (**Figure 1.5A**) (Singh et al., 2015). Apart from RBPs exerting an influence on the RNAs, recent advances have shown that RNAs might as well exert their influence upon the RBPs in terms of their function, localization, stability and interactions (**Figure 1.5B**). Of late, several chromatin regulating proteins including ATP-dependent chromatin remodelers have been identified to bind RNA which impacts their role in gene expression regulation leading to a growing interest in dissecting this aspect of their biological function (Corley et al., 2020; Hentze et al., 2018).



**Figure 1.5: Protein-RNA interactions.**

**(A)** RNA binding protein (RBP) can bind to RNA and regulate its different aspects of life **(B)** Conversely, RNA binds to RBP and regulates its function. From (Hentze et al., 2018).

### **1.3.1 Chromatin regulators as RNA binding proteins**

A growing list of chromatin modifying proteins have been shown to be regulated by their interaction with RNA. Often, such proteins differ from the traditional RBPs, in that, they do not possess an established RBD or lack the specificity for RNA sequences or structures. Thereby, often such non-canonical RBPs are termed promiscuous in their RNA interactions (Hendrickson et al., 2016). Here, I detail some of the extensively studied non-canonical RBPs that are traditionally known for chromatin regulation.

#### **1.3.1.1 MOF**

MOF (male absent on the first), a component of the MSL (male specific lethal) complex, is a histone acetyltransferase (HAT) that acetylates the lysine residue 16 in histone H4 (H4K16). It is of pivotal importance for the dosage compensation in *Drosophila* males and its loss is lethal specifically to male flies (Akhtar et al., 2000). To compensate for the lack of an X-chromosome, the expression of the genes on the single X-chromosome in male (XY) flies is doubled to match the dosage of the X-chromosome related genes in the female (XX) counterparts. To achieve this, MSL complex is recruited specifically to the male X-chromosome in flies where it causes the global acetylation of H4K16, inducing the transcriptional upregulation of X-linked genes in males (Kind et al., 2008). Notably, the recruitment of the MSL complex to the male X-chromosome is mediated by its interaction with male-specific non-coding RNAs called rox1 and/or roX2, including the interaction of MOF with roX RNAs (Franke and Baker, 1999). In fact, simultaneous deletion of the roX RNAs from *Drosophila* male embryos has been shown to disrupt the localization of MSL complex from the X-chromosome. Additionally, a depletion of RNAs in *Drosophila* Schneider 2 (S2) cells by treatment with RNase A has been shown to cause loss of MOF from the nuclear compartment. Upon *in vitro* analysis of the RNA binding properties of MOF, it was established that two amino acids - tyrosine 416 and tryptophan 426, in its chromo-barrel domain, mediate its RNA binding and mutations of these two amino acids lead to complete absence of male progeny. MOF provides an instructive example for the importance of RNA binding in directing the activity of a histone modifying enzyme (Conrad et al., 2012; Keller and Akhtar, 2015).

### **1.3.1.2 PRC2 complex**

The polycomb repressive complex 2 (PRC2) has been extensively studied for RNA binding. Canonically known for its methyltransferase activity, where in, PRC2 complex marks lysine 27 of histone H3 (H3K27me3) which plays a crucial role during development and malignancy. It is a multiprotein complex and three of its core components – EZH2 (enhancer of zeste 2), SUZ12 (suppressor of zeste 12) and EED (embryonic ectoderm development) have been shown to bind RNA directly (Yan et al., 2019a). The first reports of PRC2 interaction with RNA came from the studies on X-chromosome inactivation. In mammalian females (XX), one X-chromosome is inactivated to account for the double copies of X-linked genes. In 1999, Xist (X inactive specific transcript) non-protein coding RNA was reported to be transcribed from the inactive X-chromosome and instead of its export from the nucleus, it was shown to act in cis and form ‘banded localization’ on the same chromosome it was transcribed from (Duthie et al., 1999). Soon after, in 2002, it was shown that PRC2 core component EED, stably associated with inactive X-chromosome in trophoblast stem cells (Mak et al., 2002). Several years later, in 2008, Jeanine Lee and her colleagues showed that short repeat RNA of 1.6 kb within Xist - called RepA, directly bound the EZH2 subunit of PRC2 complex and recruited it to the X-chromosome to induce the trimethylation of H3K27 and subsequent X-chromosome inactivation (XCI) (Wang et al., 2008).

Soon the role of RNA interaction in the recruitment and modulation of PRC2 function at different loci across the genome began to be investigated. PRC2 was reported to bind a non-coding RNA (ncRNA) that was transcribed antisense from the HOXC locus, hence called HOTAIR (HOX antisense intergenic RNA). This interaction of PRC2 with HOTAIR ncRNA was shown to have effects in trans and recruit PRC2 to the HOXD locus where PRC2 subsequently trimethylates the H3K27 to repress the genes in the HOXD cluster. In fact, depleting HOTAIR ncRNA lead to a depletion of PRC2 complex on the HOXD locus and a subsequent depletion of H3K27me3 levels. This indicates the essential role played by this RNA-PRC2 interaction to localize PRC2 (Rinn et al., 2007).

Reports of PRC2 interacting with several other ncRNAs such as Braveheart, MALAT1, PINC, Air, ANRIL among others suggested that although lacking a *bona fide* RBD, PRC2 might be a general RNA binding protein complex (Klattenhoff et al., 2013; Yan et al., 2019a). A large-scale study confirmed this view by finding that about 20% of long intergenic ncRNA (lincRNAs) expressed in different mammalian cells were bound by PRC2 complex (Khalil et

al., 2009). More interest grew in the dissection of this RNA binding property of PRC2 and the roles it carries. It was shown that PRC2 lacks specificity and bound to a wide range of RNAs, however, it preferred binding to G-quadruplex (G-4) forming and G-rich RNAs. Therefore, PRC2 is neither highly specific nor highly unspecific in its RNA binding, thereby, promiscuous in nature (Davidovich et al., 2013, 2015; Kretz and Meister, 2014; Wang et al., 2017). Contradicting the earlier hypothesis that RNA interaction recruited PRC2 to chromatin, several reports suggested that RNA binding prevented PRC2 from interaction with chromatin. In fact, it was proposed that binding of PRC2 to RNA or chromatin is 'mutually antagonistic' and a depletion of RNA was shown to cause an increment in PRC2 occupancy on the chromatin. Not only did RNA evict PRC2 from chromatin but structural analysis showed that RNA interacted with the catalytic active center of EZH2 component of PRC2, occluding it and thereby, inhibiting its histone methyltransferase activity (Beltran et al., 2016, 2019; Riising et al., 2014; Zhang et al., 2019). Solving the puzzle, Thomas Cech and his colleagues showed in induced pluripotent stem cells (iPSCs), that PRC2 interaction with RNA is essential for its recruitment at PRC2 target genes, however, this interaction leads to its eviction from actively transcribing PRC2 non-target genes (Long et al., 2020).

### **1.3.1.3 dMi-2/CHD4**

In *Drosophila* polytene chromosomes, dMi-2 redistributes to the heat shock gene loci 87A and 87C, upon a heat shock, where it is necessary for efficient transcription of the heat shock response genes. Further analysis of this recruitment lead to the hypothesis that dMi-2 was recruited to the body of the heat shock genes in a two-step manner, wherein, dMi-2 first interacted with the poly-ADP ribose (PAR) polymers that are produced immediately after the heat shock and then the nascent RNA transcripts produced from the heat shock genes. During *In vitro* binding assays, recombinant dMi-2 also bound to an *in vitro* transcribed RNA derived from one of the heat shock genes (Mathieu et al., 2012; Murawska et al., 2011). Later, Ingrid Grummt and her colleagues showed that a hypo-phosphorylated form of CHD4 bound to a long ncRNA (lncRNA) called PAPAS which recruits CHD4 to rDNA (ribosomal DNA) promoters in a heat shock dependent manner. This recruitment of CHD4 is essential to repress the ribosomal synthesis in an event of heat shock (Zhao et al., 2018). Although, established for its role as an ATPase motor protein (see section 2.2.2.4), these results gave an indication that dMi-2 might have some ability to bind to RNA.

Several of the *bona fide* chromatin regulators lack a known RBD, yet there is a growing list of chromatin associated proteins that bind RNA and are regulated in both their localization as well as function by this interaction. The processes and the surfaces through which chromatin regulators that lack a RBD interact with RNA are poorly understood and remain a matter of investigation.

## **1.4 How do proteins bind RNA?**

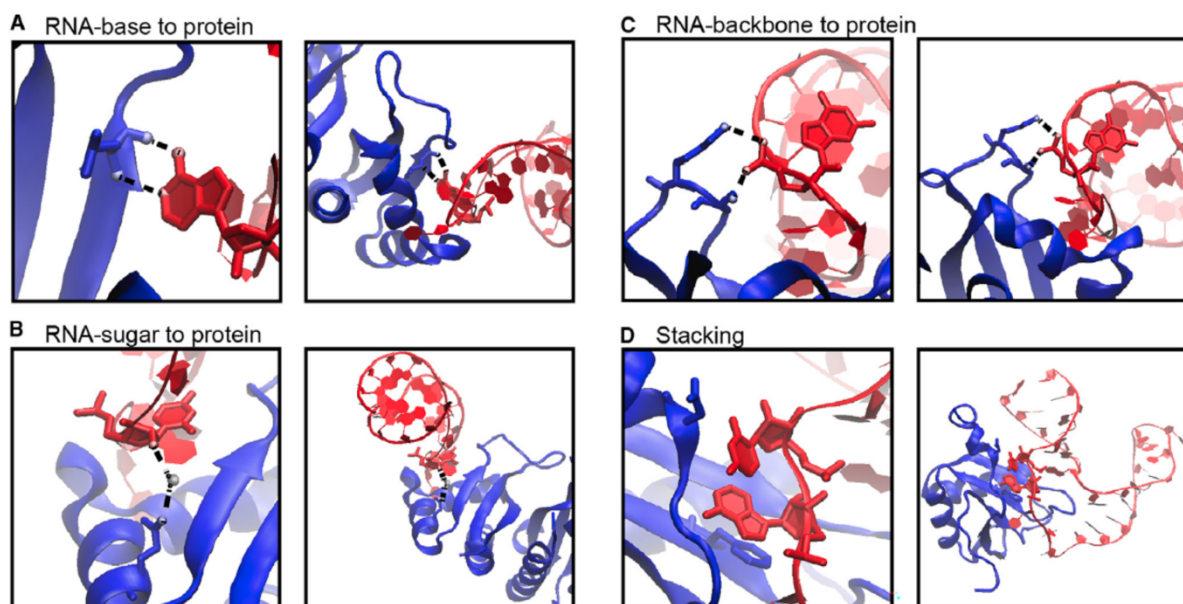
RNA binding proteins bind to RNA through highly diverse mechanisms involving different domains or surfaces. Various experimental techniques such as X-ray crystallography, nuclear magnetic resonance (NMR) and hydrogen-deuterium exchange mass spectroscopy (HDX-MS) shed light on the interacting surfaces and the amino acids and nucleotides involved in RNA-protein interactions (Corley et al., 2020; Long et al., 2017). Generally, there are domains specialized in binding to RNA and these domains, forming stable secondary and tertiary structures, are called RNA binding domains (RBDs). Often an RBP could possess several such RBDs that mediate and enhance its RNA interaction (Cléry and Allain, 2012). However, lately a plethora of RNA binding proteins have been discovered which lack such canonical RBDs. Several of these RBPs contain intrinsically disordered regions (IDRs) that serve as RNA binding sites. However, due to their disordered nature, the structural details of their interaction with RNA remain poorly understood (Järvelin et al., 2016). Nonetheless, the structural details of several other protein-RNA interactions are available and selected molecular characteristics of these interactions and domain structures will be briefly discussed below.

### **1.4.1 Molecular Interactions**

When a protein interacts with RNA, there is dynamic rearrangement at the local site of interaction in both RNA as well as protein. NMR and crystal structure studies of native and RNA bound proteins have shown that these rearrangements can be observed in the form of shifts in the RNA backbone and ‘flipping out’ of the RNA bases and amino acid residues. The sites of contact become rigid, locking the two molecules together while the adjacent regions loosen to absorb the changes in entropy (Corley et al., 2020; Leulliot and Varani, 2001; Matthews et al., 2016). At the local sites of interaction, there are different types of forces that bind the two molecules together, some of which are briefly discussed below.

### 1.4.1.1 Hydrogen bonds and Van der Waals interactions

Hydrogen bonds (H-bonds) are often the source of interaction between an RNA and a protein. The distances between H-bonds are 2.4-3Å and each bond carries 0.5-4.5 kcal/mol of energy (Auweter et al., 2006). The Van der Waals (VdW) interactions on the other hand, are the weakest of electrostatic interactions that form at a distance of above 3Å and carry forces of 0.5-1 kcal/mol. In the context of protein-RNA interactions, H-bonds and VdW interactions can occur between the 2'-OH group of the sugar, any RNA base or phosphodiester backbone (phosphate) of the RNA, and protein (**Figure 1.6**) (Teplova et al., 2011). Upon large scale analysis of several protein-RNA data sets, it was found that the H-bonds with the RNA base, 2'-OH group of sugar and phosphodiester backbone occur at an average frequency of 35.5%, 23.5% and 41%, respectively. This suggests that on average, most of the RBPs that bind to RNA through H-bonds, do so by interacting with the phosphate backbone of the RNA. However, the VdW interactions were found to be highly variable. The analysis further revealed that in about 71.5% of H-bonds and 76% of VdW interactions, RNA was contacted by the side chains of amino acids indicating H-bonds and VdW interactions are largely driven by amino acid side chains of the RBPs (Corley et al., 2020; Han and Nepal, 2007).



**Figure 1.6: Protein-RNA interactions**

Human NOVA1 and U1A KH domains visualized with VMD zoomed-in (left) and zoomed-out (right). Protein in blue and RNA in red. **(A)** Hydrogen bonds (H-bonds) between adenine base and main-chain atoms of Leucine. **(B)** H-bonds bridged by a water molecule between 2' OH in the sugar of RNA and glutamine. **(C)** H-bonds (two) between the phosphate backbone of



RNA (guanine) and serine and lysine residues. **(D)** Unpaired loop stacking interaction between adenine and cytosine nucleotides of RNA and aspartic acid and phenylalanine residues of protein. Adapted from (Corley et al., 2020).

#### **1.4.1.2 Hydrophobic and $\pi$ ( $\pi$ ) stacking interactions**

Hydrophobic interactions that account for about 50% of the protein-RNA interactions, are the attractive forces that are created between nonpolar molecules in order to minimize the contact with the water molecules. Generally, these interactions occur at a distance of 3.8-5.0Å and each interaction carries 1-2 kcal/mol of energy (Dill and MacCallum, 2012; Morozova et al., 2006; Onofrio et al., 2014). In the context of protein-RNA binding, hydrophobic interactions taking place between the RNA and the hydrophobic sidechain of a protein stabilize the protein-RNA interactions by forming a 'hydrophobic core' composed of hydrophobic amino acids and bases that are sequestered away from the solvent. For instance, the methionine-rich region of SRP54 protein, called M-binding domain, provides the hydrophobic surface for binding to SRP RNA (Akopian et al., 2013; Corley et al., 2020).

$\pi$  interactions are non-covalent interactions that occur between the ring of any nitrogenous base in the RNA and the  $\pi$  containing amino acids in the protein. Aromatic amino acid residues – histidine, tryptophan, tyrosine and phenylalanine, and charged residues including arginine, glutamine and aspartate possess  $\pi$  systems.  $\pi$  interactions tend to stack, hence called  $\pi$  stacking interactions (**Figure 1.6D**). These are generally strong, occurring over a distance of 2.7-4.3Å with 2-6 kcal/mol of energy per interaction. Analysis of the crystal structure data of several protein-RNA interactions has revealed that there can be several  $\pi$  interactions per structure, therefore, conferring considerable stability to the protein-RNA complexes. For example, in humans, U1A spliceosome protein is stabilized by several stacking  $\pi$  interactions with the polyadenylation inhibition element of an RNA (Corley et al., 2020; Wilson et al., 2016).

#### **1.4.2 RNA binding domains**

As mentioned earlier, canonical RBPs possess discrete RNA binding domains (RBDs). Often these RBDs are less than 100 amino acids in size and of those only few residues actually contact the RNA directly. However, a protein could possess several such RBDs that employ hydrogen bonds,  $\pi$  stacking formation and/or other forces of interaction as mentioned above,

to stabilize the protein-RNA complexes (Cléry and Allain, 2012). While there is a large number of known RBDs, I have selected a few of the best-established RBDs and will briefly discuss them below.

#### **1.4.2.1 RNA recognition motif**

The RNA recognition motif (RRM) is also called ribonucleoprotein domain (RNP). It is the most common RBD and is found in almost 0.5-1% of all human proteins. It is also the most extensively studied RBD and besides its RNA binding properties, it has been shown to also mediate protein-protein and protein-DNA interactions (Cléry and Allain, 2012). The total length of the RRM is 90 residues and it folds into two  $\alpha$ -helices and four  $\beta$ -sheets with a topology of  $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ . Within  $\beta$ -sheets 1 and 3, there are two stretches of conserved amino acids, one eight residues and another six residues long, called RNP1 and RNP2, respectively. In the three-dimensional space, there are a total of three conserved aromatic residues between RNP1 and 2 that remain exposed and face the solvent. It is these residues that are involved in directly contacting the RNA (Cléry and Allain, 2012; Corley et al., 2020). In total, one RRM can contact 2-8 nucleotides of an RNA and even though each RRM can have different specificity, most prefer GU-rich RNA sequences. The RNP 1 and 2 within RRM, form hydrogen bonds or stacking interactions to establish the contact often with nanomolar affinities. In some RBPs, there can be several RRM which bind to RNA at the same time, increasing the affinity as well as specificity. For example, in heterogenous nuclear RNPA1 (hnRNPA1), there are two RRM and both bind to RNA together (Beusch et al., 2017; Corley et al., 2020).

#### **1.4.2.2 K homology domain**

K homology (KH) domain was first discovered in 1993 in heterogenous nuclear ribonucleoprotein K (hnRNPK) as a pre-mRNA binding motif. It was reported to 'tenaciously' bind C-rich RNA sequences (Siomi et al., 1993). Later research revealed that KH domain can interact with 4 nucleotides at a time and is about 70 amino acids in size. It folds in three  $\alpha$ -helices and three  $\beta$ -sheets. Based on the differences in folding, there are two types of KH domains, type I (eukaryotic) - that assumes a  $\beta_1\alpha_1\alpha_2\beta_2\beta'\alpha'$  topology and type II (prokaryotic) - that assume a  $\alpha'\beta'\beta_1\alpha_1\alpha_2\beta_2$  topology. Between the three  $\alpha$ -helices of the KH domain, there is a conserved 'GXXG' motif that contacts the RNA through hydrogen bonds (Cléry and Allain, 2012; Valverde et al., 2008). KH domains rarely employ stacking interactions, therefore, the

proteins that bind RNA only through KH domains do so with lesser affinity as compared with RRM mediated binding, often with micromolar affinities. However, proteins that possess multiple KH domains have higher binding affinity than those with only one KH domain. For example, in the human E3 ligase protein MEX-3C, there are two KH domains – KH1 and 2 that mediate its binding to MEP10 RNA. The mutational analysis revealed that simultaneously perturbing the RNA binding ability of both the KH domains significantly decreased the binding affinity several folds compared to the effects of mutating only one KH domain suggesting a synergistic role in RNA binding (Yang et al., 2017).

### **1.4.2.3 Zinc Finger domain**

Zinc finger (ZnF) domains were initially reported to bind DNA, however, recent evidence suggests that several RBPs contact RNA through their ZnF domains. Typically, ZnF domains are about 30 amino acids long and fold in two  $\beta$ -sheets and one  $\alpha$ -helix forming a simple  $\beta\beta\alpha$  topology which is held together by a  $Zn^{2+}$  ion (Corley et al., 2020). Depending on the arrangement of cysteine and histidine residues that interact with the  $Zn^{2+}$  ion, there are three subtypes of ZnF domains that bind to RNA – zinc-knuckle (CCHC), CCCH, CCCC-type and CCHH-type. The zinc knuckles interact with the stem-loops in the RNA establishing contacts with the RNA bases in the loop and the phosphate backbone in the stem. The CCCH and CCCC-type ZnFs form hydrogen bonds with the RNA bases (Cléry and Allain, 2012). In fact, an RBP can have several such ZnFs and provide the specificity in RNA binding. For example, the structural analysis of an RBP called Muscle-blind-like 1 (MBNL-1) has revealed that it contains several CCCH-type ZnFs which can specifically recognize three nucleotide 5'-GCU-3' RNA sequences. The CCHH-type ZnFs can bind both single stranded as well double stranded RNA and DNA. It has been found in transcription factor TFIIIA (ZnF4-6) and it interacts with the phosphate backbone of the 5S rRNA (Cléry and Allain, 2012; Corley et al., 2020).

### **1.4.2.4 Cold shock domain**

A large set of proteins that are involved in cold adaptation across different forms of life, called cold shock proteins (CSPs), possess cold shock domains (CSDs) that are involved in chaperoning of RNAs under a downward shift in temperatures (Corley et al., 2020). Typically, CSDs are about 70 amino acids in length, however, could be longer in eukaryotic proteins and are highly conserved. These domains are folded into five anti-parallel  $\beta$ -strands (also

called  $\beta$ -barrels) that have been established as oligosaccharide and oligonucleotide-binding folds (OB-folds) (Amir et al., 2018). Within the CSDs, there are conserved RNP1 and RNP2, that are also found in RRM as discussed in section 2.3.2.1. The binding to RNA takes place by sequential stacking interactions and formation of hydrogen bonds with RNA bases, often leading to nanomolar affinities. The human YB-1 protein contains the CSD that is well characterized. Although, it possesses ability to also bind single stranded DNA, however, it binds to RNA with at least 10-fold higher affinity and has been shown to be involved in the regulation of translation initiation (Kljashtorny et al., 2015).

There are several other RBDs that have been well characterized and whose structural studies are available. However, it would be beyond the scope of this thesis to discuss each RBD.

### **1.4.3 Intrinsically disordered regions**

As mentioned above, while several RBPs bind RNA through dedicated domain structures, a growing number of RBPs are identified that lack such structured domains. Most of these RBPs possess intrinsically disordered regions (IDRs) that mediate their RNA binding. These regions are unstructured and are often composed of amino acid repeats such as RS repeats (arginine and serine repeats), RGG repeats (arginine and glycine repeats), R/K basic patches (arginine- or lysine-rich patches) and other motifs (Hentze et al., 2018). There are a number of RBPs that contain disordered RG repeats and such proteins are called SR proteins or SR-like proteins. The SR proteins are well characterized for their role in splicing but also mediate other RNA related processes such as export and translation. RS repeat elements are disordered, however, phosphorylation within the RS repeat residues reduces their structural flexibility. For example, the RS repeat elements in splicing factor SRSF1 (serine/arginine repeat splicing factor 1), adopt an arch-like rigid structure upon phosphorylation that provides a direct, non-specific, binding surface for RNA molecules. However, it is important to mention that IDRs do not necessarily require transition towards structural stability to bind RNA, as numerous IDRs bind to RNA despite the lack of any structural rigidity (Shepard and Hertel, 2009; Xiang et al., 2013). This is true for human nuclear ribonucleoprotein U (hnRNP-U) which does not contain any canonical RBD, but instead possesses RGG repeat containing IDRs which bind to Xist and PANDA lncRNAs and mediate its role in epigenetic regulation (Hasegawa et al., 2010; Puvvula et al., 2014). Several RBPs contain IDRs in addition to their canonical RBDs. In several of such cases, IDRs have been shown to increase the efficiency of RNA interaction. For example, Fragile-X mental retardation protein (FMRP) has two

canonical KH domains and RGG repeat IDRs. It preferentially binds certain mRNAs mainly through its KH2 domain, however, loss of RGG IDRs leads to significant loss in RNA binding suggesting that the IDRs work in cooperation with the KH domains (Ascano et al., 2012; Järvelin et al., 2016).

The IDRs in some proteins fold in a more structured manner once they come in contact with RNA. This has been observed with some of the RBPs that contain stretches of basic amino acids like arginine, called arginine-rich motifs (ARMs). Such IDRs provide a positively charged surface that faces the solvent and remains exposed for the negatively charged RNA to interact with. For example, the bovine immunodeficiency virus Tat proteins (Trans-activator of transcription) have ARMs which are disordered and bind to the 5'-end of the viral nascent RNA, called trans-activation response element (TAR). This binding induces the disordered Tat ARM to adopt a beta-turn structure (Puglisi et al., 1995). Similarly, IDRs can be composed of other basic amino acids like lysines. Such IDRs are called Poly-K or K-rich regions. Just like the ARM containing IDRs, poly-K IDRs provide a positively charged surface for RNA interaction. Of course, some IDRs contain a combination of lysine and arginine-rich residues (K/R-rich). Often IDRs have less sequence specificity but high affinity for RNA. In fact, the RNA binding mediated by IDRs can be as efficient as that mediated by a canonical RBD like RRM. For example, the homolog of protein SDA1 called SDAD1 has no RBD but in its C-terminal is poly-K IDR and its binding to RNA is as efficient as an RRM containing RBP *in vivo* (Corley et al., 2020; Järvelin et al., 2016; Strein et al., 2014). Therefore, the RNA binding mediated by IDRs can be as efficient as those mediated by *bona fide* RBDs. Hence, the strong RNA interactions reported for several canonical chromatin regulators that lack RBDs can be explained by their IDRs. Recently, IDRs have hence elicited a strong interest in protein-RNA biology.

### **1.5 Objectives**

During heat shock in *Drosophila*, dMi-2 is recruited to the heat shock genes where it supports their efficient transcription. Previous evidence has suggested that this recruitment takes place in two steps – dMi-2 first interacts with the PAR polymers which bring dMi-2 into close vicinity with the HS genes. In the second step, dMi-2 binds to nascent RNA transcripts from the HS genes and then gets recruited to the heat shock gene loci. In fact, recombinant dMi-2 was shown to bind to a part of Hsp70 derived RNA *in vitro* (Mathieu et al., 2012; Murawska et al., 2011). In addition, mammalian CHD4 was shown to interact with an unrelated long non-

coding RNA during heat shock. These isolated observations indicated the possibility of a conserved, general RNA binding function of dMi-2/CHD4.

In this study, I asked if dMi-2 binds RNA also in the absence of heat shock and aimed to systematically characterize its RNA binding properties. I sought to use an unbiased approach (iCLIP) to identify novel dMi-2 associated RNAs *in vivo* and to characterise dMi-2 binding to selected RNAs *in vitro*. Finally, I aimed to identify the RNA binding domain of dMi-2 and to investigate the functional consequences of the RNA binding.

## 2 Materials and Methods

### 2.1 Materials

#### 2.1.1 Cell lines

Cell line	Features	Description
Sf9	Immortalized ovarian cells	Spodoptera frugiperda (Sf)
S2	Drosophila melanogaster (late stage embryos, likely macrophage-like lineage)	Schneider 2 (S2)
S2 Cas9	Stably expressing Cas9, no protein tagged	See S2
S2 dMi-2-GFP	S2 cells expressing GFP tagged dMi-2 (Monoclonal)	See S2
S2 dMof-mCherry/dMi-2- GFP	S2 cells expressing mCherry tagged dMof and GFP tagged dMi-2 (Polyclonal)	See S2
S2 dLint1-GFP	S2 cells expressing GFP tagged dLint1	See S2

#### 2.1.2 Cell Culture

Materials	Supplier
Sf-900™ SFM	Thermo Fischer Scientific
Fetal bovine serum (FBS)	Sigma Aldrich
Penicillin-Streptomycin (Pen: Strep, 10,000 U/mL: 10,000 µg/mL)	Thermo Fischer Scientific
Schneider's Drosophila Medium	Thermo Fischer Scientific

### 2.1.3 Baculovirus

Baculovirus expressed protein; C-terminal tag	Source
dMi-2 WT FLAG (pFastBac Dual)	AG Brehm
dMi-2 $\Delta$ N FLAG (pVL1392)	AG Brehm
dMi-2 $\Delta$ C FLAG (pVL1392)	AG Brehm
dMi-2 691-1271 FLAG (pVL1392)	AG Brehm
dMi-2 377- 483 FLAG (pFastBac1)	This study
dMi-2 484 – 690 FLAG (pFastBac1)	This study
dMi-2 1-376 FLAG (pFastBac1)	This study
dMi-2 $\Delta$ 376 FLAG (pFastBac1)	This study
dMi-2 1-117 FLAG (pFastBac1)	This study
dMi-2 118-238 FLAG (pFastBac1)	This study
dMi-2 239-376 FLAG (pFastBac1)	This study

### 2.1.4 Plasmids

Plasmid	Description	Source
pVL1392 dMi-2 WT (1-1982)	Encodes full length dMi-2 C-terminally tagged with FLAG, for generation of recombinant Baculovirus	Bouazoune et al., 2002
pFastBacDual-dMi-2-FLAG	Encodes full length dMi-2 C-terminally FLAG tagged, for generation of recombinant Baculovirus	Ulla Kopiniak
pVL1392-dMi-2 $\Delta$ N FLAG (691-1982)	Encodes N-terminally truncated version of dMi-2, FLAG tagged on C-terminus, for generation of recombinant Baculovirus	Bouazoune et al., 2002
pVL1392-dMi-2 $\Delta$ C FLAG (1-1271)	Encodes C-terminally truncated version of dMi-2, FLAG tagged on C-terminus, for generation of recombinant Baculovirus	Bouazoune et al., 2002



pVL1392-dMi-2 691-1271	Encodes C-terminally FLAG tagged ATPase region of dMi-2, for generation of recombinant Baculovirus	Bouazoune et al., 2002
pFastBac1-dMi-2 1-376 FLAG	Encodes C-terminally FLAG tagged 1-376 amino acids of dMi-2, for generation of recombinant Baculovirus. The pFastBac1 vector was linearized with NotI restriction endonuclease and plasmid Gibson assembled.	This study
pFastBac1-dMi-2 377-1982 FLAG	Encodes C-terminally FLAG tagged 377-1982 amino acids of dMi-2, for generation of recombinant Baculovirus. The pFastBac1 vector was linearized with NotI restriction endonuclease and plasmid Gibson assembled.	This study
pFastBac1-dMi-2 1-117 FLAG	Encodes C-terminally FLAG tagged 1-117 amino acids of dMi-2, for generation of recombinant Baculovirus. The pFastBac1 vector was linearized with NotI restriction endonuclease and plasmid Gibson assembled.	This study
pFastBac1-dMi-2 118-238 FLAG	Encodes C-terminally FLAG tagged 118-238 amino acids of dMi-2, for generation of recombinant Baculovirus. The pFastBac1 vector was linearized with NotI restriction endonuclease and plasmid Gibson assembled.	This study
pFastBac1-dMi-2 239-376 FLAG	Encodes C-terminally FLAG tagged 239-376 amino acids of dMi-2, for generation of recombinant Baculovirus. The pFastBac1 vector was linearized with NotI restriction endonuclease and plasmid Gibson assembled.	This study

pFastBac1-dMi-2 377-690 FLAG	Encodes C-terminally FLAG tagged 377-690 amino acids of dMi-2, for generation of recombinant Baculovirus. The pFastBac1 vector was linearized with NotI restriction endonuclease and plasmid Gibson assembled.	This study
pRmHa-3 dMi-2 WT	Encodes N-terminally FLAG tagged full length dMi-2 for Copper induced expression in S2 cells	This study
pRmHa-3 dMi-2 Δ376	Encodes N-terminally FLAG tagged 377-1982 amino acids of dMi-2 for Copper induced expression in S2 cells	This study
pRmHa-3 dMi-2 ΔRRM	Encodes N-terminally FLAG tagged 11-238 fused with 377-1982 amino acids of dMi-2 for Copper induced expression in S2 cells	This study

### 2.1.5 Inhibitors

Inhibitors	Supplier
Aprotinin (10 µg/µL) Used: 10 µg/mL	AppliChem GmbH
Leupeptin (10 µg/µL) Used: 10 µg/mL	AppliChem GmbH
Pepstatin A (10 µg/µL) Used: 10 µg/mL	AppliChem GmbH
Phenyl-methane-sulfonyl-fluoride (PMSF, 0.2 M) Used: 0.2 mM	Roth
RiboLock RNase Inhibitor (RNAsin) (40 U/µL)	Thermo Scientific

## 2.1.6 Enzymes

Enzymes	Supplier
DNase I	VWR
Proteinase K	Roth
RNase A	AppliChem GmbH
RNase I	Thermo Fischer Scientific
Superscript III	Thermo Fischer Scientific
Turbo™ DNase	Thermo Fischer Scientific
Phusion™ High-Fidelity DNA Polymerase	Thermo Fischer Scientific

## 2.1.7 Oligonucleotides

Custom unmodified DNA oligonucleotides were purchased from IDT. The stocks were dissolved in Tris-EDTA (TE) buffer pH 7.5 at 100 pmol/μL and stored at -20°C.

### 2.1.7.1 Oligonucleotides for generating templates for *in vitro* transcription

The Oligonucleotides used to generate the templates for *in vitro* transcription contain a T7 promoter sequence at the 5'-end of the forward primer. The oligonucleotides for generating templates for double stranded RNA additionally contain T7 promoter sequences at 3'-end of the reverse primer. The T7 promoter sequences are depicted in lowercase while the gene specific DNA sequence is shown in uppercase.

Oligo name	Sequence
Hsp70T7+245F	taatacgactcactatagggCTCACACACAATGCCTGCTATTG
Hsp70T7+245R	CTGGTCGTTGGCGATAATCTCC
Hsp70T7+498F	taatacgactcactatagggGAGGACATGAAGCACTGGC
Hsp70T7+498R	CCTCGGGGGCAAATCTCTTG
Hsp70T7+1484F	taatacgactcactatagggGCTGATCGAGCGCAACTGC
Hsp70T7+1484R	GCCCTCATACACCTGAATGG
Hsp70T7+1781F	taatacgactcactatagggGCAGGCCGAGATTGATCGC
Hsp70T7+1781R	GCTCTCCAGGGCATTCTAG
Hsp70 230-merFwd	taatacgactcactatagggCCTACGGACTGGACAAGAAC

Hsp70 100-mer Rev	CCCTCGTCGATGGTCAG
PgantT7+245F	taatacgactcactatagggTCGGCAGGATGGAGGATC
PgantT7+245R	TAATGCTGGCATTGGAGTGCATC
PgantT7+498F	taatacgactcactatagggTCTTCCGGGCGATCCCAG
PgantT7+498R	TAGAAGCACATGACTATGGAGGC
PgantT7+1484F	taatacgactcactatagggCTGCATGTGCCGGGCG
PgantT7+1484R	GCAGCTGGAATGTGTGCGAC
PgantT7+1781F	taatacgactcactatagggAAGTGCCACGAGATGCTCGG
PgantT7+1781R	CAGCTGCCCGGAGGCAG
Hsp70ds100merT7R	taatacgactcactatagggCCCTCGTCGATGGTCAG
Pgantds100merT7R	taatacgactcactatagggCTTGATGAGGCGCAGCAG
Pgant230-mer Fwd	taatacgactcactatagggGTCATTGGTGCACGCGAGGC
Pgant100-mer Rev	CTTGATGAGGCGCAGCAG

### 2.1.7.2 Oligonucleotides for cloning in pFastBac1 vector to generate Baculovirus

All clonings were performed by Gibson assembly. The sequences overlapping the vector are depicted in lower case while the gene specific DNA nucleotides are shown in upper case. The FLAG-tag coding sequences are underlined.

Oligo name	Sequence
dMi2Fwd1_117	tcgacgagctcactagtcgcaccATGGCATCGGAGGAAGAG
dMi2Rev1_117_FLAG	Caggctctagattcgaaagc <u>ctacttgtcatcgtcgtcc</u> <u>ttgtagtc</u> TGCCGACTCCTTCTCCTTGG
dMi2Fwd118_238	tcgacgagctcactagtcgcaccATGTCATCCGGAATGCCATCTGT GG
dMi2Rev118_238_FLAG	caggctctagattcgaaagc <u>ctacttgtcatcgtcgtccttgtagtc</u> GACGGCC TCCTCGTAAATGTC
dMi2Fwd239_376	tcgacgagctcactagtcgcaccatgGAGGAGGAGGAAGAAGAGG
dMi2Rev239_376_FLAG	caggctctagattcgaaagc <u>ctacttgtcatcgtcgtccttgtagtc</u> GTGCTCAT GCTCGCCATC
dMi2Fwd377_483	tcgacgagctcactagtcgcaccATGCAGGACTACTGCGAGGTG

dMi2Rev377_483_FLAG	caggctctagattcgaaagcctacttgtcatcgtcgtccttgtagtcGCAGCTG CAACGAGGACAG
dMi2Fwd484_690	tcgacgagctcactagtcgcaccATGCCTCCGCTCACTGGCAAGG
dMi2Rev484_690_FLAG	caggctctagattcgaaagcctacttgtcatcgtcgtccttgtagtcGACCTTGA GCTTGGACTTGC
dMi-2_377_1982FLAG	caggctctagattcgaaagcctacttgtcatcgtcgtccttgtagtcGACGCCG GAATTATTCG

### 2.1.7.3 Oligonucleotides for sequencing

Oligo name	Sequence
dMi-2_WT_seq1F	GCATCGGAGGAAGAG
dMi-2_WT_seq2F	GAAGATGGCGAGCATG
dMi2_377_1982_seq1F	CAGCACCATGTATCTGG
dMi2_377_1982_seq2F	GAGCAACCAGAGCAAG
dMi2_377_1982_seq3F	GAGCCGGAGGATTGG
dMi2_377_1982_seq4F	GAGAAAGGCCAAGCG
dMi2_377_1982_seq5F	CCAAGGACCCGGAGG
dMi2_377_1982_seq1R	CGTTTATGATCGCGAAACGG
dMi2_377_1982_seq2R	GCTCCGGCATGCTGTAG
dMi2_377_1982_seq3R	GGTTCTGTCGAGCAGCTCG
dMi2_377_1982_seq4R	GTTGATCAGTGAGCAAGAGCC
dMi2_377_1982_seq5R	GCGTCGAGAGTGGCACGG
dMi-2 WT Seq1R	CAGAACTCACGCCACTTG
pRmHa-3 Fwd	GCAAAATCAAGTGAATCATCTC

### 2.1.7.4 Oligonucleotides for CHIP and qPCR

Oligo name	Sequence	Source
Hsp70 -154F	TGCCAGAAAGAAAACCTCGAGAAA	Boehm et al. 2003
Hsp70 -154R	GACAGAGTGAGAGAGCAATAGTACAGAGA	Boehm et al. 2003
Hsp83 Total_F	GGGTTTCTACTCCGCCTACC	Murawska et al. 2011
Hsp83 Total_R	CACGTACTGCTCGTCATCGT	Murawska et al. 2011
Vrille_ChIP6_Fwd	GTTTCTTCTGCCCAATGC	Kreher et al., 2017

Vrille_ChIP6_Rev	CCTCTTTGGCCGAAAAATCT	Kreher et al., 2017
USP_Fwd (52)	GCTCCTTTGAGCGACGAT	Kreher et al., 2017
USP_Rev (53)	CGAGAAGCTCTGGTTGAGGA	Kreher et al., 2017
EcR_All_Fwd	CAGGAGGACCAGATCACGTT	Kreher et al., 2017
EcR_All_Rev	TTCGCGAAGAATATTGAGTCC	Kreher et al., 2017
Broad_ChIP1_fwd	GCCGGCAATATTAGAAGTTCG	Kreher et al., 2017
Broad_ChIP1_rev	ATTGGATTGGATGGTGCAG	Kreher et al., 2017
GAPDH1_F	GAGCAAGGACTAAACTAGCCAAA	Meier et al. 2012
GAPDH1_R	CAACAGTGATTCCCGACCA	Meier et al. 2012
Intergenic2R_F	TGCTGACTGCCATCAAATTC	Meier et al. 2012
Inregenic2R_R	TACTGCTGTGACGGCTTTG	Meier et al. 2012

### 2.1.8 Ladders and dyes

Ladders and dyes	Supplier
GeneRuler 1kb DNA ladder	Thermo Fischer Scientific
GeneRuler 100 bp DNA ladder	Thermo Fischer Scientific
PageRuler™ Prestained Protein ladder	Thermo Fischer Scientific
6x Orange DNA loading dye	Thermo Fischer Scientific
2x RNA loading dye	Thermo Fischer Scientific
5x SDS loading buffer	Thermo Fischer Scientific
Spectra™ Multicolor High Range Protein Ladder	Thermo Fischer Scientific

### 2.1.9 Chemicals

Chemicals	Supplier
Acetic acid	Roth
Acrylamide mix (30%, 40%)	Roth
Agarose	Roth
Ammonium persulfate (APS)	Roth
Bovine serum albumin (BSA)	Thermo Fischer Scientific
Bradford dye	Bio-Rad Laboratories, Incorporated (Inc.)
Coomassie Brilliant Blue R-250	AppliChem GmbH

## Materials and Methods

Dimethyl sulfoxide (DMSO)	Serva
Dithiothreitol (DTT, 1M) Used: 1 mM	Sigma Aldrich
Ethylenediaminetetraacetic acid (EDTA)	Roth
Ethanol	Roth
Ethidium bromide	Roth
FLAG peptide	Sigma Aldrich
Formaldehyde	Polyscience, Inc.
Glycerol	Roth
Glycine	Roth
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	CalBiochem
Immobilon® Western Chemiluminescent HRP Substrate	Merck Milipore
Isopropanol	Roth
Methanol	Merck
Pierce™ 10x WB Transfer Buffer	Thermo Fischer Scientific
Potassium chloride (KCl)	Roth
Powdered milk	Roth
RNase away	Thermo Fischer Scientific Molecular Bioproducts
Sodium acetate	Roth
Sodium dodecyl sulfate (SDS)	AppliChem GmbH
SYBR® Gold	Thermo Fischer Scientific
Tetramethylethylenediamine (TEMED)	Roth
Triptolide	Biomol
Tris-aminomethane (TRIS)	Roth
TRIS-HCl	Roth
Triton® X-100	Roth
Tween® 20	Roth
Urea	Roth

### 2.1.10 Kits

Kits	Application	Suppliers
Bac-to-Bac Baculovirus Expression System	Sf9 Cell transfection for Baculovirus generation	Thermo Fischer Scientific
Dynabeads™ MyONE™ Silane cDNA Purification kit	iCLIP2	Thermo Fischer Scientific
Expand High Fidelity <sup>PLUS</sup> PCR system	PCRs for CRISPR tagging	Roche
MEGAscript™ T7 transcription Kit	<i>In vitro</i> transcription	Thermo Fischer Scientific
peqGOLD Total RNA kit	RNA isolation	VWR
ProNex® Size-selective purification system	iCLIP2 library preparation	Promega
QIAquick PCR purification kit	DNA isolation after ChIP	Qiagen
QIAGEN Plasmid Maxi Kit	DNA isolation	Qiagen
QIAquick Gel extraction kit	DNA isolation from agarose gel	Qiagen
SensiFAST™ SYBR® Lo-ROX kit	qPCR	Bioline
RNeasy Mini Kit	RNA isolation	Qiagen
SensiFAST™ cDNA Synthesis Kit	Reverse transcription	Bioline

### 2.1.11 Consumables

Consumables	Supplier
Amicon® Ultra Centrifugal Filters	Merck
25 mL Bio-Pure Reservoir	Diversified Biotech, Inc.
Blotting paper sheets	Sartorius Stedim Biotech
10 mL and 20 mL Gravity Chromatography Columns	Bio-Rad Laboratories, Inc.
Cuvettes	Sarstedt
15 mL and 50 mL Falcons	Greiner Bio-One



Filter pipet tips	Biozym Scientific GmbH
Gel Cassettes (1.0 and 1.5 mm)	Novex® (Thermo Fischer Scientific)
Gloves	Sempermed Unigloves®
LoBind Eppis	Eppendorf Biozym
96-well PCR plates	Thermo Fischer Scientific
Pasteur pipettes	BRAND GmbH & Co KG
Pipette tips	Sarstedt Gilson
10 cm and 15 cm plates	Sarstedt
T75 and T175 tissue culture flasks	Sarstedt
Roti®-PVDF (WB membrane)	Roth
Round bottom falcons	Greiner Bio-One
Super RX-N Fuji Medical X-ray film (WB)	Fujifilm

## 2.1.12 Devices

<b>Devices</b>	<b>Supplier</b>
Cell culture laminar air filter hood	Thermo Fischer Scientific
Centrifuge	Sigma Aldrich (Rotanta 460R) Thermo Fischer Scientific (Heraeus Fresco 21) Thermo Fischer Scientific (Heraeus™ Megafuge™ 1.0) Thermo Fischer Scientific (Sorvall RC-5B) Ultracentrifuge (Thermo Scientific)
ChemiDoc™ Touch Imaging System	Bio-Rad Laboratories, Inc.
Freezer (-20°C)	Liebherr
Freezer (-80°C)	Thermo Fischer Scientific
Fridge	Liebherr
Gel documentation	Intas Scientific Imaging
Incubator	Binder HmbH Infors HT (Multitron Pro)

Light microscope (CK2)	Olympus
Confocal microscope	Leica
Medical Film Developer	Konica Minolta
Microwave	LG
Thermocycler (T3000)	Biometra GmbH
Pipettes	Gilson
Pipette device	Heathrow Scientific
qPCR machine (Mx3000P)	Agilent Technologies
Rotating Wheel (Multi-Rotator PTR-35)	Grant Instruments Ltd.
Plate rocker (ST5D Cat)	Kobe
Sonicator (Bioruptor)	Diagenode
Spectrophotometer (Genesys 10vis)	Thermo Fischer Scientific
Spectrophotometer (NanoDrop)	Thermo Fischer Scientific
Stratagene Stratalinker UV 1800 Crosslinker	Marshall Scientific
Votrex	Scientific Industries, Inc.
Water bath	Grant Instruments Ltd.

### 2.1.13 Software

Software	Supplier
GraphPad Prism	GraphPad Software, Inc.
Image Lab™ 6.0.1	Bio-Rad Laboratories, Inc.
MxPro qPCR Software	Agilent Technologies

### 2.1.14 RNAs

RNA (length)	Source	Sequence (5'→3')
Hsp70 Aa #1 (100 nt)	This study ( <i>In vitro</i> )	CUCACACACAAUGCCUGCUAUUGGAAUCGAUCUGGGCACC ACCUACUCCUGCGUGGGUGUCUACCAACAUGGCAAGGUGGAG AUUAUCGCCAACGACCAG

	transcription)	
Hsp70 Aa #2 (100 nt)	This study ( <i>In vitro</i> transcription)	GAGGACAUGAAGCACUGGCCUUUCAAGGUUGUAAGCGAC GGCGGAAAGCCCAAGAUCGGGGUGGAGUAUAAGGGUGAG UCCAAGAGAUUUGCCCCCGAGG
Hsp70 Aa #3 (100 nt)	This study ( <i>In vitro</i> transcription)	CCUACGGACUGGACAAGAACCUCAAGGGUGAGCGCAAUGU GCUUAUCUUCGACUUGGGCGGCGGCACCUUCGAUGUCUCC AUCCUGACCAUCGACGAGGG
Hsp70 Aa #4 (100 nt)	This study ( <i>In vitro</i> transcription)	GCUGAUCGAGCGCAACUGCCGCAUUCGGUGCAAGCAGACU AAGACGUUCUCCACAUACGCGGACAACCAGCCCGGAGUCUC CAUUCAGGUGUAUGAGGGC
Hsp70 Aa #5 (100 nt)	This study ( <i>In vitro</i> transcription)	GCAGGCCGAGAUUGAUCGCAUGGUGAACGAGGCUGAAAAG UACGCCGACGAGGACGAGAAGCAUCGCCAGCGAAUAACCUC UAGAAAUGCCCUGGAGAGC
Hsp70 Aa #7 (230 nt)	Murawska et al., 2011	CCUACGGACUGGACAAGAACCUCAAGGGUGAGCGCAAUG UGCUUAUCUUCGACUUGGGCGGCGGCACCUUCGAUGUCU CCAUCCUGACCAUCGACGAGGGAUCACUGUUCGAGGUGCG CUCCACCGCCGGAGACACACACUUGGGCGGCGAGGACUUU GACAACCGGCUAGUCACUCAUCUGGCGGACGAGUUCAAGC GCAAGUACAAGAAGGAUCUGCGCUCCAACCCU
Hsp70 Aa dsRN A (100 bp)	This study ( <i>In vitro</i> transcription)	CCUACGGACUGGACAAGAACCUCAAGGGUGAGCGCAAUGU GCUUAUCUUCGACUUGGGCGGCGGCACCUUCGAUGUCUCC AUCCUGACCAUCGACGAGGG

Hsp70 Aa #6 (40 nt)	This study (IDT)	CCUACGGACUGGACAAGAACCUCAAGGGUGAGCGCAAUGU
Hsp70 Aa #8 (20 nt)	This study (IDT)	CCUACGGACUGGACAAGAAC
Pgant 35a #1 (100 nt)	This study ( <i>In vitro transcrip tion</i> )	UCGGCAGGAUGGAGGAUCGACGAAGGGAACGCAACGCCA CGGGCCGAGCUGAGCUACCAGGCCAGGGUCACCGUCGGA UGCACUCCAAAUGCCAGCAUUA
Pgant 35a #2 (100 nt)	This study ( <i>In vitro transcrip tion</i> )	UCUUCCGGGCGAUCCCAGACACUCGGCACAAGGUGUGC GAUCGCCAGGAGACCACGGAAGCGGAAAAUCUGCCCCA GGCCUCCAUAUGUCAUGUGCUUCUA
Pgant 35a #3 (100 nt)	This study ( <i>In vitro transcrip tion</i> )	GUCAUUGGUGCACGCGAGGCCGUGGGCGAUGUGCUCG UCUUCCUCGACUCGCAUAUCGAGGUCAAUCAGCAGUG GCUGGAGCCGCGUCGUCGCGCCUCAUCA
Pgant 35a #4 (100 nt)	This study ( <i>In vitro transcrip tion</i> )	CUGCAUGUGCCGGGCGAGGAGUCCAAGAAGUCGGCGGC AGCGCCCAUCUUUCAGCCUUGGCACUCGCGAAAACGUAA CUAUGUCGACACAUUCCAGCUGC
Pgant 35a #5 (100 nt)	This study ( <i>In vitro transcrip tion</i> )	AAGUGCCACGAGAUGCUCGGCGAUCAGCAGUGGCGGCAC ACCCGCAACGCCAACAGUCCCGUCUACAACAUGGCCAAAG GCACCUGCCUCCGGGCAGCUG
(GA) <sub>20</sub> (40 nt)	This study (IDT)	GA

(CA) <sub>20</sub> (40 nt)	This study (IDT)	CACACACACACACACACACACACACACACACACACACA
(G <sub>4</sub> A <sub>4</sub> ) <sub>4</sub> (40 nt)	This study (IDT)	AAAAGGGGAAAAGGGGAAAAGGGGAAAAGGGGAAAAAAAA
Poly (A) (40 nt)	This study (Dharma con, Inc.)	AA
Poly (C) (40 nt)	This study (Dharma con, Inc.)	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
Poly (U) (40 nt)	This study (Dharma con, Inc.)	UU
Poly (G) (40 nt)	This study (Dharma con, Inc.)	GG

## 2.1.15 Antibodies and antisera

### 2.1.15.1 Primary antibodies

Antibody	Host origin	Experiment	Dilution	Source
αFLAG	Rabbit, monoclonal	WB	1:2,000	Sigma Aldrich
αMi-2	Rabbit, polyclonal	WB	1:10,000	(Kehle et al., 1998)
αE(z)	Rabbit, polyclonal	WB	1:3,000	-

αRPD3	Rabbit, polyclonal	WB	1:10,000	-
αH3	Rabbit	WB	1:10,000	-
αLSD1	Rabbit, polyclonal	WB	1:10,000	-
dMOF	Rabbit, Polyclonal	IF	1:1,000	Asifa Akhtar

### **2.1.15.2 Secondary antibodies**

Antibody	Host origin	Dilution	Source
αRabbit	Polyclonal	1:30,000	GE Healthcare
αRabbit Alexa-633	Mouse	1:1,000	Thermo Scientific

## **2.2 Methods**

### **2.2.1 Cell Culture**

#### **2.2.1.1 Culturing Sf9 cells**

*Spodoptera frugiperda*-9 (Sf9) cells were cultured in Sf-900™ II SFM (Thermo Scientific). The Sf9 culture medium was supplemented with 1% Penicillin-Streptomycin (Gibco) and cells were cultured at 26°C with 80 revolutions per minute (rpm) shaking. The cells were cultured from 0.5 million per millilitre (mil/ml) to 4 mil/ml. Cells were split every 3 days.

#### **2.2.1.2 Culturing S2 cells**

Schneider-2 (S2) cells were cultured in Schneider's Drosophila Medium Thermo Scientific) supplemented with 1% Penicillin-Streptomycin (Pen-Strep) and 10% Fetal Bovine Serum (FBS). 0.5 mil/ml cells were seeded in a flask and cultured at 26°C for 4-5 days and split in a ratio between 1:10 to 1:20.

#### **2.2.1.3 Cell counting**

Cells were diluted 1:10 in respective medium. 10 µL cell suspension was transferred to the Neubauer cell counting chamber. Three of the four big squares, which are divided in 16 smaller squares, were counted and the mean was calculated and multiplied by the dilution factor (10). To calculate the cell-count per milliliter, the final cell number was multiplied by 10<sup>4</sup>.

#### **2.2.1.4 Freezing and thawing cells**

S2 or Sf9 cells were counted to 25 mil/ml in a 15 mL falcon, centrifuged to 1,000 relative centrifugal force (rcf) for 5 minutes at room temperature. The supernatant was discarded and the cell pellet resuspended in 500 µL of Freezing Buffer A. 500 µL of Freezing Buffer B was added dropwise while shaking the falcon. The cell suspension was transferred to a cryovial. The cryovial was placed in a 15 mL round bottom tube. This was then transferred to a 50 mL falcon filled with about 25 mL isopropanol. The cells were stored at -80°C.

Buffer	Composition
Freezing Medium A	60% Sf9 or S2 medium supplemented with 1% Pen-Strep + 40% FBS. Store at 4°C
Freezing Medium B	60% Sf9 or S2 medium supplemented with 1% Pen-Strep + 40% DMSO. Store at 4°C

Frozen Sf9 or S2 cells were taken out of the  $-80^{\circ}\text{C}$  freezer and quickly thawed in hands. The thawed cell suspension was washed by adding dropwise to a 15 mL falcon containing 14 mL respective culture medium. The cells were centrifuged at 1000 rcf for 5 minutes at room temperature. The supernatant was discarded and cell pellet again washed with 5 mL of respective medium. The centrifugation was repeated and final cell pellet was resuspended in 5 mL of respective medium and transferred to 25 mL of Sf9 culture medium in a 250 mL flask (Sf9 cells) or 15 mL of S2 culture medium in a T75 flask.

### 2.2.1.5 Infection of Sf9 cells

Proteins to be expressed in Sf9 cells were cloned in pFastBac1 vectors that provide a FLAG tag at the N-terminal of the protein of interest. The plasmid created as such were transformed into MAX Efficiency<sup>®</sup> DH10Bac<sup>™</sup> Competent E. coli. The Bacmids from these E. coli were extracted following the standard plasmid preparation protocol. The Bacmids were then transfected into Sf9 cells cultured in a 6-well plate using Cellfectin II reagent (Thermo Scientific). The cells were allowed to grow for 5-6 days. After visually inspecting the cells for signs of infection (cell lysis and increase in cell size), the supernatant containing the viral particles was saved as passage-1 (P1) virus. To amplify the P1 virus, Sf9 cells grown in a 10-cm dish were infected with 100-200  $\mu\text{L}$  of P1 virus and incubated for 5-6 days. Upon visual inspection of infection signs by using a light microscope, the supernatant was collected and saved as P2 virus. To amplify the P2 virus such that it has high viral titers, Sf9 cells were cultured in 15-cm dishes and infected with 100  $\mu\text{L}$  of P2 virus. Upon visual confirmation of infection, cells were incubated for 5-6 days or till 50-60% cell lysis was observed. The supernatant (P3 virus) was collected, briefly centrifuged and stored at 4°C.



The P3 Baculovirus was used for protein expression. 1 litre Sf9 cell culture at a density of 1 mil/ml were infected with 40-50 mL of P3 virus in a 5-litre flask and incubated for 48 hours at 26°C with shaking. Cells were harvested and proteins extracted.

## **2.2.2 Biochemical Methods**

### **2.2.2.1 Baculovirus protein extraction**

The Sf9 cells expressing the protein of interest were harvested by centrifugation at 3,200 rpm for 25 mins at 4°C using a swinging bucket rotor (Thermo Scientific, Heraeus). The supernatant was discarded and the cell pellet washed with 30 mL of cold 1XPBS (Phosphate buffer saline), pH 7.4. The centrifugation was repeated as earlier and supernatant discarded. The cellular pellet was resuspended and lysed in BC250 buffer supplemented with protease inhibitors and dithiothreitol (DTT), transferred to 50 mL falcons and incubated on ice for 15 mins. The lysate was snap-frozen in liquid nitrogen and stored at -80°C.

<b>Buffer</b>	<b>Composition</b>
1X PBS	140 mM NaCl 2.7 mM KCl 8.1 mM Na <sub>2</sub> HPO <sub>4</sub> 1.5 mM KH <sub>2</sub> PO <sub>4</sub> pH adjusted to 7.4 with HCl (Thermo Fisher Scientific). Stored at 4°C.
BC-250	0.4 mM EDTA pH 8.0 20 mM HEPES mP 7.9 10% Glycerol 250 mM NaCl Dissolved in autoclaved Milli-Q Water, filtered and stored at 4°C.

### **2.2.2.2 FLAG-immunoprecipitation (FLAG-IP)**

The cell lysates of FLAG-tagged proteins were thawed from -80°C and the freeze-thaw cycle was repeated. The lysate was cleared by centrifugation at 10,400 rpm for 75 mins at 4°C in a sorval centrifuge (Thermo Scientific, Sorvall). The supernatant was transferred to pre-chilled

50 mL falcons and the cell debris was discarded. The protein concentration of the whole cell extract was determined by Bradford assay.

An aliquot from the whole cell extract was taken as input, mixed with 5X SDS loading dye and stored at -20°C. To the rest of the extract, pre-equilibrated anti-FLAG M2 agarose (Sigma Aldrich) beads were added and incubated at 4°C for 4 hours on a rotator. After immunoprecipitation, the beads-protein solution was transferred to Bio-Rad Econo-pac gravity-flow columns, the flow-through was collected. The beads were washed with 40X bead volume of BC250, BC500, BC1000, BC500, BC250 and BC100 buffer, in that order. Each buffer was supplemented with protease inhibitor cocktail and DTT.

The purified protein of interest was eluted with twice the volume of beads with elution buffer containing FLAG-peptide diluted in BC100 (1:20), supplemented with protease inhibitors and DTT. As soon as the elution buffer was added, the beads were properly resuspended and 10 µL bead suspension was aliquoted out as Beads-IN. The elution was repeated once more. After collecting the elute twice, the beads were again resuspended in elution buffer and 10 µL of beads were aliquoted out as Beads-OUT. To the beads-IN and OUT samples, 10 µL of 2X SDS loading dye was added and samples stored at -20°C. The elutes were pooled and concentrated in pre-equilibrated centricons (Millipore). From the concentrated and purified protein, 21 µL were aliquoted out for analysis. The rest of the protein was aliquoted, snap frozen and stored at -80°C.

<b>Buffer</b>	<b>Composition</b>
Buffering capacity (BC) buffers	0.4 mM EDTA pH 8.0 20 mM HEPES mP 7.9 10% Glycerol 100 mM / 250 mM / 500 mM / 1000 mM NaCl. Dissolved in autoclaved Milli-Q Water, filtered and stored at 4°C.
FLAG elution buffer (EB)	FLAG peptide (5 mg/ml) diluted 1:20 in BC-100 buffer.

### **2.2.2.3 Protein concentration estimation**

Protein concentration was estimated using Bradford Dye (Bradford, 1976). The Bradford dye was diluted 1:5 and a standard curve was generated using BSA standard. The protein to be estimated was diluted 1:10 in water, 1 mL of 1:5 diluted Bradford Dye was added and

incubated at room temperature for 5 mins. The absorption was measured at 595 nm against a blank solution containing the 1:10 diluted buffer of the protein of interest. The standard curve was used to estimate the final concentration of the protein.

<b>Reagent</b>	<b>Composition</b>
Bradford dye	Bradford reagent concentrate diluted 1:5 in autoclaved Milli-Q water. Filtered and stored at 4°C.

In-gel estimation- Purified protein was loaded on a SDS polyacrylamide gel along with known BSA standards (Thermo Scientific) and electrophoresed. The gel was stained with Coomassie staining solution, destained, rehydrated and imaged with Bio-Rad ChemiDoc touch system. The protein bands were quantified along with the BSA standards using Image Lab software and concentration calculated against the BSA standards.

#### **2.2.2.4 Sodium dodecyl sulfate Polyacrylamide gel electrophoresis (SDS-PAGE)**

Polyacrylamide gels were poured with SDS and optimum concentration of acrylamide solution, depending on the size of the protein. The resolving gel was layered on top with 5% stacking gel. The gels were allowed to polymerise. Samples mixed with 5X SDS loading dye were denatured at 95°C for 5 mins and loaded on the SDS polyacrylamide gels and electrophoresed at 100 volts (V) till the bromophenol blue stain reached the resolving gel. The electrophoresis was then increased to 150V. The gels were run till the dye front reached the bottom of the gel.

<b>Solution</b>	<b>Composition</b>
Resolving gel	6% / 8% % / 10% / 12% or 15% Acrylamide Tris-HCl pH 8.8 Water SDS Ammonium persulfate TEMED Mixed and poured

Stacking gel	5% Acrylamide Tris-HCl pH 6.8 Water SDS Ammonium persulfate TEMED Mixed and poured
--------------	--

### 2.2.2.5 Coomassie staining and imaging

At the end of the electrophoresis, the gel was transferred to a Coomassie staining solution and incubated for 30 mins at RT while shaking. The gel was washed with de-ionized water thrice and destained using Coomassie destaining solution until the gel was ready for imaging. The gel was rehydrated in Coomassie storing solution for 30 mins at RT and imaged with Bio-Rad ChemiDoc touch system.

Solution	Composition
Resolving gel	6% / 8% % / 10% / 12% or 15% Acrylamide Tris-HCl pH 8.8 Water SDS Ammonium persulfate TEMED Mixed and poured
Stacking gel	5% Acrylamide Tris-HCl pH 6.8 Water SDS Ammonium persulfate TEMED Mixed and poured

### **2.2.2.6 RNase A treatment and Cell Fractionation**

The protocol was provided by Manuel Beltran and adapted for S2 cells (based on Zoabi et al., NAR 2014, Mendez et al., Mol Cell Biol 2000 and Beltran et. al., Genome Res 2016).

$2-4 \times 10^7$  cells per condition were aliquoted in 15 mL falcons. The cells were washed with 1X PBS at RT. Cells were permeabilized with 10 mL of 1X PBS supplemented with 0.05% Tween-20 and incubated at 4°C for 10 mins with intermittent agitation. The cell suspension was centrifuged at 1,200 rpm for 2 mins at 4°C and the supernatant was discarded. The cell pellet was resuspended in 950  $\mu$ L of 1X PBS (RT). The sample was split into two vials, 450  $\mu$ L each. One was labelled as sample and another Mock. The sample vial was taken to a different room while the mock sample was processed on the normal RNase-free bench. To the sample vial, 50  $\mu$ L of 10 mg/ml RNase A was added while to the Mock vial, 50  $\mu$ L of 10 mg/ml BSA was added. Each vial was incubated on separate rotating wheels at RT for 30 mins. Both the vials were centrifuged at 1,200 rpm for 2 mins at 4°C. The supernatant was discarded. The cell pellet was washed by pipetting up and down with 10 mL cold 1X PBS.

The cell suspension was centrifuged as earlier. The pellet was resuspended in 1 mL ice-cold buffer A supplemented with protease inhibitor and DTT (Care was taken to NOT mix by pipetting up and down as that could break the nuclei. The buffer was added to the wall and flicked gently). At this step and onwards, 2.5  $\mu$ L/mL RNase inhibitor - RNaseOUT was added to the Mock sample.

Triton X-100 was added to a final concentration of 0.1%, the tubes were flicked to mix and incubated on ice for 5 mins. To isolate the Cytoplasmic fraction (S1), the samples were centrifuged at 1,300 g for 4 mins at 4°C. The supernatant was separated, cleared by high-speed centrifugation at 20,000 g for 15 mins at 4°C and stored at -20°C. The nuclei were washed twice with 1 mL Buffer A (without pipetting up and down, only by gently flicking). The nuclei were lysed with 250  $\mu$ L of Buffer B (without pipetting up and down and only by gently flicking the tube as the chromatin is sticky and could stick to the pipette tip) supplemented with protease inhibitors and DTT. For the Mock sample, the Buffer B was additionally supplemented with RNaseOUT. Both samples were incubated on ice for 30 mins. The nucleoplasm (S2) was separated by centrifuged at 1,700 g for 4 mins at 4°C. The nucleoplasm was stored at -20°C. The chromatin pellet was washed twice with 1 mL of Buffer B and centrifuged at 1,700 g for 4 mins at 4°C (without pipetting up and down). The final chromatin pellet (P3) was resuspended in 200  $\mu$ L of 1X Laemmli buffer supplemented with  $\beta$ -Mercaptoethanol ( $\beta$ -Me). The chromatin samples were then sonicated for 7 cycles (30 sec ON, 30 sec OFF, at High output) in a Bioruptor (Diagenode). The samples were heated to

95°C for 5 mins and then stored at -20°C. The cytoplasm, nucleoplasm and chromatin fractions were loaded on 8-12% SDS PAGE. For the Mock sample, care was taken to clean the bench, the pipettes and all equipment with RNaseZap.

<b>Buffer</b>	<b>Composition</b>
Permeabilization solution	0.05% Tween-20 in 1x PBS
Buffer A	10 mM HEPES pH 7.9 10 mM KCl 1.5 mM MgCl <sub>2</sub> 0.34 M Sucrose 10% Glycerol 1 mM DTT, supplemented with protease inhibitors
Buffer B	3 mM EDTA 0.2 mM EGTA 1 mM DTT, supplemented with protease inhibitors and RNaseOUT (mock sample only)

### **2.2.2.7 Western Blot**

Samples to be analysed by Western blot (WB) were mixed with SDS loading dye, denatured at 95°C for 5 mins and loaded on SDS polyacrylamide gel along with a size marker and SDS PAGE was performed as described in 2.2.2.4. At the end of SDS PAGE, the proteins were transferred to a polyvinylidene difluoride (PVDF) membrane which was activated by brief exposure to methanol, at 400 mA (milliamperes) for 2 hr in cold conditions. The membrane was blocked with 5% milk in PBS-T for 30 mins at RT while shaking and then incubated overnight with primary antibody in 5% milk in PBS-T at 4°C on a shaker. The membrane was washed 3x with PBS-T for 5 mins each at RT on a shaker and incubated with the respective secondary antibody in 5% milk in PBS-T at RT for 2 hr on a shaker. The membrane was again washed 3x with PBS-T for 5 mins each at RT on a shaker. The secondary antibody is coupled with horseradish peroxidase (HRP) which reduces the peroxide from the WB detection solution. The reduction of peroxide provides the energy for the chemiluminescence of luminol in the WB detection solution. The chemiluminescence was detected by X-ray films which were developed by X-ray film processor (Kodak).

<b>Buffer</b>	<b>Composition</b>
10X SDS running buffer	For 1 Litre (L): SDS 10 g (gram) TRIS base 30.3 g Glycine 144.1 g Dissolve in 800 mL Milli-Q water, adjust the vol. to 1L and store at RT
Transfer buffer	10X Pierce™ WB Transfer buffer diluted 1:10 in Milli-Q Water. Store at 4°C
PBS-T	1x PBS supplemented with 1% Tween-20 Store at RT
Blocking milk solution	5% milk powder in 1x PBS-T Store at 4°C
WB detection solution	750 µL Immobilon™ Western HRP Substrate Peroxide Solution 750 µL Immobilon™ Western HRP Substrate Luminol Reagent Freshly prepared, mixed before use.

### 2.2.2.8 Agarose gel electrophoresis

DNA samples were verified for size and quality using agarose gels. The optimum percentage gels were prepared by dissolving DNA-grade agarose powder in 1x Tris-Acetic acid-EDTA (TAE) buffer, boiled briefly, supplemented with Ethidium bromide (EtBr) and poured in a chamber. The gel was allowed to polymerise. The DNA samples were mixed with DNA loading dye (1x final), loaded on the agarose gel along with a size marker and electrophoresed at 130 V for 45 mins. The gel was then exposed to ultraviolet (UV) radiations to visualize the DNA bands.

<b>Buffer</b>	<b>Composition</b>
TRIS-acetate-EDTA (TAE) buffer	40 mM Tris acetate pH 8.0 1 mM EDTA pH 8.0 Store at RT

### 2.2.2.9 Electrophoretic mobility shift assay (EMSA)

EMSAs were used to detect the binding of proteins with nucleic acids, DNA or RNA. When a protein binds DNA or RNA, the resulting protein-nucleic acids complex migrates slower this can be detected in native gels. In a 10  $\mu$ L reaction, DNA/RNA was incubated with the protein of interest in EMSA buffer supplemented with protease inhibitors, DTT and RNase inhibitor (for RNA EMSA only) for 15 mins at RT. The samples were loaded on a pre-run 5% native polyacrylamide gel at 100 V for 1 hr. The gel was stained with 1x SYBR gold stain and processed as mentioned in 2.2.2.10. The intensity of the shifted and unshifted bands were quantified using Image Lab software. After subtracting the background intensity, the bound fractions were calculated as-

$$\text{Fraction bound} = \frac{\text{Intensity of shifted band}}{(\text{intensity of unshifted band} + \text{intensity of shifted band})}$$

Percent bound was calculated as-

$$\text{Percent bound} = (\text{Fraction bound}) \times 100$$

The dissociation constant ( $K_d$ ) was calculated using the hill slope equation as

$$Y = \frac{B_{\text{max}} * X^h}{(K_d^h + X^h)}$$

Where,

$B_{\text{max}}$  = the maximum specific binding in the same unit as Y,

$K_d$  = ligand concentration needed to achieve a half-maximal binding at equilibrium in the same unit as X, and,

$h$  = hill slope.

Buffer	Composition
EMSA buffer	40 mM KCl 20 mM Tris pH 7.6 1.5 mM MgCl <sub>2</sub> 0.5 mM EGTA 10% Glycerol BSA (200 ng/ $\mu$ l) 0.4 Units (U) RNase inhibitor) 1mM DTT, supplemented with protease inhibitors



### 2.2.2.10 Native Polyacrylamide gel electrophoresis

EMSA reactions were analysed by native polyacrylamide gel electrophoresis. A 100 mL solution of 5% native polyacrylamide gel was prepared by mixing 12.5 mL of Rotiphorese Gel 40 (19:1, Acrylamide:Bisacrylamide) (Roth) with 5 mL of 10x Tris-Boric acid-EDTA (TBE) buffer (Thermo Scientific), supplemented with ammonium persulfate (APS) and tetramethylethylenediamine (TEMED) and poured in a glass chamber. The gel was allowed to polymerise at RT for 30 mins, pre-run for 30 mins at 100 V. After the EMSAs samples were loaded, the gel was electrophoresed for 1 hr at 100 V at RT in 0.5x TBE buffer. The gel was incubated in a 1x SYBR gold staining solution (Thermo Scientific) for 5 mins at RT with shaking. The gel was washed 2x with deionized water and imaged with Bio-Rad ChemiDoc Touch system.

Buffer	Composition
10x Tris-borate-EDTA (TBE) buffer	1 M Tris 0.9 M Boric acid 0.01 mM EDTA

### 2.2.2.11 Denaturing polyacrylamide gel electrophoresis (denaturing PAGE)

RNA samples were loaded on denaturing PAGE to verify the quality of the RNAs. The gels were poured just like the native gels as described in section 2.2.2.10, however, additionally supplemented with 8% urea. The RNA samples were mixed with RNA loading dye, heated to 65°C for 5 mins, instantly cooled for 1 min on ice and loaded on pre-run denaturing PAGE and electrophoresed at 100 V for 30 mins in 0.5x TBE buffer. The gels were processed as described in 2.2.2.10.

### 2.2.2.12 Circular dichroism (CD) spectroscopy

CD spectroscopy was used to determine the secondary structures of RNAs. CD spectroscopy measures the difference in the absorption of the right and left – circularly polarised light. 10 micromolar ( $\mu\text{M}$ ) RNA in 170  $\mu\text{L}$  volume was allowed to fold as described in Wang et al., Mol Cell, 2017. CD spectra were measured in a 0.1 cm path-length quartz cell. The measurements were taken from 220 nanometer (nm) to 350 nm with a data pitch of 0.1 nm and scanning speed of 50 nm/min and a band width of 1 nm. The cell holder temperature was regulated and the cuvette chamber was flushed with dry Nitrogen ( $\text{N}_2$ ) gas. Spectra were measured at temperatures indicated in the figures.

The CD spectroscopy data were analysed by using CAPITO CD analysis and plotting software (Wiedemann et al., 2013). The data were smoothed with Savitzky-Golay filter and CD values were plotted as mean residue ellipticity ( $\theta$  in grad cm<sup>2</sup> dmol<sup>-1</sup>).

Buffer	Composition
CD buffer	50 mM Tris-HCl pH 7.5 100 mM KCl 2.5 mM MgCl <sub>2</sub> 0.1 mM ZnCl <sub>2</sub> 2 mM 2-mercaptoethanol 0.1 mg/mL BSA 0.1 mg/mL fragmented yeast tRNA 5% v/v Glycerol

### 2.2.2.13 *in vitro* transcription

The *in vitro* transcription was used to produce some of the RNAs in this study. For any single stranded RNA to be produced, PCR template DNA was generated with T7 promoter sequence on its 5'-end (using forward primer with T7 promoter sequence). For double stranded RNA, the template DNA was produced using primers with T7 promoter sequence on both forward as well as reverse primers. *In vitro* transcription was set up using MEGAscript T7 transcription kit (Ambion) as-

Component	Amount
ATP solution (75 mM)	2 $\mu$ L
CTP solution (75 mM)	2 $\mu$ L
GTP solution (75 mM)	2 $\mu$ L
UTP solution (75 mM)	2 $\mu$ L
10x Reaction buffer	2 $\mu$ L
DNA template (PCR products)	0.1-12 $\mu$ g
Enzyme Mix (T7 RNA polymerase, 1U/ $\mu$ L)	2 $\mu$ L

Samples were mixed thoroughly and incubated 37°C for 16 hrs. The DNA template was digested by adding 1  $\mu$ L of TURBO DNase (2U/ $\mu$ L) and samples incubated at 37°C for 15 mins. The reactions were stopped by adding 30  $\mu$ L of Nuclease-free water (Thermo

Scientific). To precipitate the RNA, 0.1 vols of 3M Sodium acetate and 2.5-3 vols of ice cold 100% ethanol were added and samples incubated at -80°C for 1hr. Samples were centrifuged at 14,800 rpm for 30 mins at 4°C. The supernatant was discarded and RNA pellet was washed 2x with cold 75% ethanol. Supernatant was drained and RNA pellet air dried for 5-7 mins at RT and re-suspended in Tris-EDTA (TE) buffer pH 8.0. RNA was quantified by Nanodrop, quality checked by electrophoresis in denaturing-PAGE and stored at -20°C.

### **2.2.2.14 Nucleosome assembly**

Mononucleosomes were assembled on purified *Drosophila* histone octamers using 601 nucleosomal DNA with 80 base pairs (bp) of flanking DNA. The nucleosomal DNA and histone proteins were mixed in a 0.8:1 ratio (determined empirically by test assembly) in 2M-TE buffer. The mix was pipetted in a low binding tube. A dialysis membrane was hydrated and equilibrated in 2M-TE buffer. The assembly mix was transferred to the dialysis membrane and incubated in 500 mL of 1x 2M-TE buffer for 40 mins. The assembly mix was dialysed to 1M-TE buffer for 1 hr, 0.5M-TE buffer for 1hr and finally 0.05M-TE buffer for 1 hr. The dialysed nucleosome assembly mix was collected by poking the membrane at the top and pipetted out into a fresh low binding tube. The assembly was concentrated in an equilibrated 50,000 Dalton centricon (Milipore) by centrifugation at 3,000 rpm for 10 mins (until the assembly volume reaches about 200 µL). A glycerol gradient was poured with glycerol gradient buffer using AKTA prime. The nucleosome assembly was loaded carefully on the top of the glycerol gradient without disturbing the gradient. The tube was weighed and balanced and ultra-centrifuged (Thermo Scientific) at 35,000 rpm for 18 hrs at 4°C.

The fractions were collected by poking a hole at the bottom of the gradient tube and about 200 µL fractions were collected. 10 µL from each fraction were loaded on a 5% Native-PAGE for quality analysis. The best fractions were pooled and the mono-nucleosomes were stored at 4°C.

<b>Buffer</b>	<b>Composition</b>
(N)M-TE nucleosome assembly buffer	10 mM Tris-HCl pH 8.0 1 mM EDTA pH 8.0 2M / 1M /0.5 M or 0.05 M NaCl
Glycerol gradient buffer A	50 mM Tris pH 8.0 50 mM NaCl 0.1% NP-40

	10% Glycerol 1 mM $\beta$ -Mercaptoethanol
Glycerol gradient buffer A	50 mM Tris pH 8.0 50 mM NaCl 0.1% NP-40 30% Glycerol 1 mM $\beta$ -Mercaptoethanol

### 2.2.2.15 Nucleosome remodelling assay

A typical nucleosome remodelling reaction was set in a volume of 10  $\mu$ L. 50 nM end-positioned mono-nucleosomes with 80bp flanking DNA were mixed with different concentrations of dMi-2 Wild type (WT) protein in BC-100 buffer. Reactions were incubated for 5 mins at 26°C, then, 2 mM ATP and 6 mM MgCl<sub>2</sub> (final concentration) were added and reactions were incubated at 26°C for 45 mins. Reactions were stopped by adding competitor plasmid DNA (5  $\mu$ g/reaction, nearly 6 Kbp plasmid) and samples were incubated on ice for 10 mins and loaded on a 5% Native-PAA gel and run in 0.5x TBE buffer. Gels were stained with 1x SYBR gold staining solution, imaged with Bio-Rad ChemiDoc touch system and bands were quantified with Bio-Rad Image Lab software. For effects of RNA on remodelling activity of dMi-2, RNA was titrated in increasing concentrations before the addition of ATP and MgCl<sub>2</sub>. The IC<sub>50</sub> was calculated in GraphPad Prism software using the following equation-

$$Y = \text{Bottom} + (\text{Top} - \text{Bottom}) / (1 + (X / \text{IC}_{50}))$$

Where,

IC<sub>50</sub> is the concentration of agonist that gives a response half way between Bottom and Top. Top and Bottom are plateaus in the units on the Y axis.

### 2.2.2.16 Reverse transcript- quantitative polymerase chain reaction (RT-qPCR)

The gene expression levels were quantified using RT-qPCR. RNA was extracted from the cells using peqGold Total RNA extraction Kit (peqLab) following the Kit protocol. The RNA was reverse transcribed to cDNA using SensiFAST™ complementary cDNA synthesis Kit (Bioline) following the kit protocol in a thermocycler with the following program:

Temperature	Time
25°C	10 mins
42°C	15 mins
85°C	5 mins
4°C	∞

The cDNA was diluted 1:5 before qPCR analysis. The qPCR allows the quantitative analysis using a double stranded (ds) DNA intercalating fluorophore. The cycle threshold (Ct) value is the number of cycles till the fluorescence reaches an indicated value. In the experiments in this thesis, since the Triptolide treatment downregulates the global gene expression including the expression of *house-keeping genes*. Therefore, expression was calculated relative to DMSO treated controls which were set to 1 using the equations-

$$\Delta Ct = (\text{Mean Ct of Triptolide treated sample} - \text{Mean Ct of DMSO control})$$

$$\text{Fold change} = 2^{-\Delta Ct}$$

The standard deviation (SD) was calculated as-

$$\Delta SD = \sqrt{(\text{Mean SD of Triptolide treated sample})^2 + (\text{Mean SD of DMSO control})^2}$$

$$\text{Standard deviation} = ((\text{Fold change of sample} \times \text{LN}(2)^2 \times \Delta SD^2)$$

Where, LN(2) = Natural logarithm of 2

### 2.2.2.17 Chromatin immunoprecipitation-quantitative polymerase chain reaction (ChIP-qPCR)

Chromatin preparation-

ChIP-qPCR was used to quantify the abundance of proteins on chromatin. 100 million cells were seeded on a 15-cm dish and the protein-DNA contacts were crosslinked using 1% (final concentration, V/V) formaldehyde (Polysciences) for 10 mins at RT with shaking. The formaldehyde was quenched with 240 mM Glycine (final concentration) for 10 mins at RT with shaking. The cells were scraped and transferred to falcons and washed 2x with cold PBS, the supernatant was discarded. The cell pellet was lysed with 1 mL lysis buffer (supplemented with protease inhibitors and DTT) and incubated 10 mins on the ice. The cell lysate was sheared with sonication with 3 cycles of 30 seconds ON/OFF for 10 mins each at high power (Diogenode). The lysate was centrifuged at 14,800 rpm for 15 mins at 4°C. The supernatant (chromatin) was transferred to fresh tubes. 50  $\mu$ L of chromatin was aliquoted out for quality check.

Immunoprecipitation-

The chromatin (140  $\mu$ L per ChIP) was diluted to 1400  $\mu$ L with ChIP-IP buffer and pre-cleared with pre-equilibrated and blocked (in blocking buffer) protein A beads (80  $\mu$ L slurry) for 1 hr on a rotating wheel at 4°C. Pre-cleared chromatin was centrifuged 14,800 rpm for 10 mins at 4°C. The chromatin was transferred to fresh tube, 14  $\mu$ L aliquoted out as input (stored at -4°C) and beads discarded. To the pre-cleared chromatin, 35  $\mu$ L (1:1) pre-equilibrated and blocked, GFP-Trap beads (Chromotek) were added and samples incubated o/n on a rotating wheel at 4°C.

Washing-

The samples were centrifuged at 2000 rpm for 5 mins at 4°C, supernatant discarded and beads were washed 3x with low salt buffer, 3x with high salt buffer, 1x with LiCl buffer and 1x with TE buffer. Each wash was incubated on a rotating wheel at 4°C for 10 mins. After the last wash, beads were resuspended in 700  $\mu$ L of TE buffer and transferred to fresh tubes, centrifuged at 2000 rpm for 5 mins at 4°C. The supernatant was discarded.

Elution-

250  $\mu$ L of elution buffer (EB) was added to each sample and incubated for 20 mins on a rotating wheel at RT, centrifuged 2000 rpm for 5 mins at RT. The supernatant was transferred to fresh tubes. To the beads, 250  $\mu$ L of EB was added and incubated 20 mins at RT on a rotating wheel and then incubated 95°C for 10 mins, centrifuged full speed at RT for 5 mins. The supernatant was pooled with the first elute. To the input samples, 486  $\mu$ L of EB was added.

Decrosslinking-

20  $\mu$ L of 5M NaCl was added to all the samples and incubated at 65°C o/n with 800 rpm shaking.

Protein digestion-

To each sample, 10  $\mu$ L of 0.5 M EDTA, 20  $\mu$ L of 1M Tris-HCl pH 6.8 and 2  $\mu$ L of Proteinase K (10 mg/mL) were added and samples incubated at 45°C for 1 hr with 800 rpm shaking.

DNA purification-

DNA was purified using Qiaquick PCR purification Kit (Qiagen) following the kit protocol. The DNA was eluted with elution buffer provided in the kit. The DNA was diluted 1:4 with the same buffer and 5 µL were used for each qPCR reaction.

qPCR-

5 µL of purified DNA was mixed with 15 µL of qPCR master mix as-

qPCR Master mix per reaction-

Component	Concentration
2x SensiFAST SYBR Lo-ROX mix	1x
Forward primer	20 nM
Reverse primer	20 nM
Water	upto 15 µL

The master mix (15 µL) was pipetted in each well of a 96-well plate and 5 µL of purified DNA was then pipetted and mixed. The qPCR was performed with following program-

95°C	2 min	
95°C	5 s	} 40 cycles
60°C	10 s	
72°C	20 s	
55°C	10 s	
....		
95°C	10 s	

Melting curve: temperature was increased in 0.5°C steps after 10 s

The Ct values derived from the qPCR were used for the quantification of the % input as-

$$\Delta Ct = \text{Mean Ct}(\text{input}) - \text{Mean Ct}(\text{elute})$$

The percentage of input was calculated as-

$$\% \text{ input} = 2^{\Delta Ct}$$

The standard deviation (SD) was calculated as-

$$SD = \ln(2) \times \% \text{ input} \times \sqrt{SD(\text{input})^2 + SD(\text{elute})^2}$$

<b>Buffer</b>	<b>Composition</b>
Lysis buffer	50 mM Tris-HCl pH 8.0 10 mM EDTA 1% SDS
ChIP-IP buffer	16.7 mM Tris-HCl pH 8.0 1.2 mM EDTA 167 mM NaCl 1.1% Triton X-100 0.01% SDS
Blocking buffer	2 mG/mL BSA 2% Fish skin gelatin 2x low salt buffer (1x final)
2x Low salt buffer	40 mM Tris-HCl pH 8.0 4 mM EDTA 300 mM NaCl 2% Triton X-100 0.2% SDS
Low salt buffer	20 mM Tris-HCl pH 8.0 2 mM EDTA 150 mM NaCl 1% Triton X-100 0.1% SDS
High salt buffer	20 mM Tris-HCl pH 8.0 2 mM EDTA 500 mM NaCl 1% Triton X-100 0.1% SDS
LiCl buffer	10 mM Tris-HCl pH 8.0 1 mM EDTA 250 mM LiCl 0.1% NP-40
Elution buffer (EB)	100 mM NaHCO <sub>3</sub> 1% SDS



### 2.2.2.18 RNase A and Immunofluorescence

300  $\mu$ L of 1 mil/mL S2 cells expressing endogenously tagged dMi-2-GFP on the C-terminal were seeded on Concanavalin-A coated coverslips placed inside a 24-well plate and allowed to spread for 1 hr at 26°C. The medium was removed and cells washed twice with 1xPBS. The cells were permeabilized with 0.05  $\mu$ g/ $\mu$ L Digitonin in 1x PBS for 15 mins at RT. The cells were washed twice with 1x PBS. The RNA was depleted by treating the cells with 10  $\mu$ g/ $\mu$ L RNase A in 1x PBS for 10 mins at RT. The control cells were treated with 1x PBS alone. Cells were washed twice with PBS and fixed with 4% paraformaldehyde (PFA) for 15 mins at RT. The cells were again washed twice with 1x PBS. Since, dMi-2 is GFP tagged, it was visualized directly under a confocal microscope. However, to visualize the effects on a known RNA binding protein, MOF, which loses its nuclear occupancy upon RNA depletion, cells were treated with anti-Mof primary antibody after blocking with 3% BSA in a humidified chamber for 2 hr. The cells were incubated with secondary antibody against Rabbit primary antibody, incubated for 1 hr in humidified conditions. The secondary antibody solution (3% BSA in PBS) was supplemented with DAPI and Phalloidin stains. The coverslips were washed twice in 1x PBS and mounted on glass slides with Mowiol mounting media. Slides were allowed to set at 4°C o/n and imaged with Confocal microscope (Zeiss).

Antibody solutions-

Solution	Components
Primary antibody solution	Anti-Mof (1:1000, Rabbit) in 3% BSA
Secondary antibody solution	Alexa-633 anti-rabbit antibody (1:1000) DAPI stain (1:1000) Phalloidin-647 stain (1:1000)

### 2.2.2.19 Gibson assembly

All the constructs in this study were generated by Gibson assembly as described in (Gibson et al., 2009). In brief, the insert DNA fragments were generated with primers having at least 15-20 nucleotides over-lapping nucleotides at the selected site of insertion in the vector on either side. The vector was cut at the selected site with restriction enzymes and purified using Qiagen gel extraction kit following the kit protocol. The insert DNA fragments were also purified by gel extraction. The PCR generated insert DNA fragments and the restriction digestion generated vector DNA were mixed with 1.33x isothermal master mix as-

The 1.33x isothermal master mix (15  $\mu$ L) was thawed on ice. To this 5 $\mu$ L of DNA (insert + vector) were added in equimolar amounts. Samples were incubated at RT for 2 mins and then immediately at 50°C for 60 mins in a thermocycler. 5  $\mu$ L from the reaction were immediately transformed in competent *E. coli*. The transformed bacteria were then plated on Luria Broth-Agar (LB-A) plates with specific antibodies and incubated o/n at 37°C. The plates were verified for colonies which were used for plasmid extraction.

5x isothermal reaction buffer-

<b>Component</b>	<b>Volume</b>
1 M Tris-HCl pH 7.5	3 mL
1 M MgCl <sub>2</sub>	300 $\mu$ L
100 mM dGTP	60 $\mu$ L
100 mM dATP	60 $\mu$ L
100 mM dCTP	60 $\mu$ L
100 mM dTTP	60 $\mu$ L
1 M DTT	300 $\mu$ L
PEG-8000 (Polyethylene glycol)	1.5 g
100 mM NAD (Nicotineamine adenine dinucleotide)	300 $\mu$ L
Water	6 $\mu$ L Aliquoted as 320 $\mu$ L in 1.5 mL tubes and stored at -20°C.

1.33x isothermal master mix-

<b>Components</b>	<b>Volume</b>
5x isothermal master mix	320 $\mu$ L
T5 exonuclease (NEB)	0.64 $\mu$ L
Phusion™ High Fidelity DNA Polymerase (Thermo Scientific)	20 $\mu$ L
Taq DNA Ligase (Thermo Scientific)	0.16 $\mu$ L
Water	Up to 1.2 mL Aliquot as 15 $\mu$ L in 0.2 mL PCR tubes and store at -20°C.

### 2.2.2.20 individual nucleotide resolution cross-linking and immunoprecipitation-2 (iCLIP2)

The iCLIP2 protocol as described in (Buchbender et al. 2019) was used. Cells expressing the protein of interest which were seeded in 15 cm dishes were washed 2x with cold PBS and all liquid was discarded. Protein-RNA crosslinking was done with exposure to 254 nm UV light (300mJ/cm<sup>2</sup>) using a Stratagene Stratalinker while the culture dish was placed in ice-cold water bath for cooling. The cells were harvested by scraping and collected in a vial and centrifuged for 5 mins at 2,000 rpm at 4°C. The liquid was discarded. The cell pellet was resuspended in 4 packed cell volumes of RIPA buffer supplemented with protease inhibitors and DTT, incubated for 10 mins on ice, mixed every 3 mins and centrifuged at 13,000 rpm for 20 mins at 4°C. Pellets were discarded and the whole cell extracts (supernatant) transferred to fresh tubes, quantified by Bradford assay and certain volumes selected for immunoprecipitation, three vials each condition and labelled 10<sup>-3</sup>, 10<sup>-4</sup> and 10<sup>-5</sup>.

Buffer	Composition
RIPA buffer	50 mM Tris-HCl pH 7.4 1% NP-40 0.1% SDS 150 mM NaCl 5 mM EDTA Store at 4°C (supplemented with protease inhibitor and RNase inhibitor on use).

#### DNase and partial RNase digestion-

To the whole cell extract, 2 volumes (vol) of RQ-1 buffer was added. Subsequently, TURBO™ DNase (1:500 vol.) and RNaseOUT (1:1,000 vol.) were added to the extract and mixed gently. RNase I was added to a final concentration of 10<sup>-3</sup>, 10<sup>-4</sup> and 10<sup>-5</sup> respectively. Samples were incubated at 37°C for 3 mins with 800 rpm shaking and immediately placed on ice for 3 mins. To each sample, 20 µL (1/50 vol.) of 5M sodium chloride (NaCl) was added and samples centrifuged at 16,400 rpm for 5 mins at 4°C. The extracts were transferred to fresh tubes.

Buffer	Composition
RQ1 buffer	40 mM Tris-HCl pH 8.0 10 mM MgSO <sub>4</sub>

	1 mM CaCl <sub>2</sub> Store at 4°C (supplemented with 1mM DTT and protease inhibitors on use)
TBS1000-T buffer	50 mM Tris pH 7.4 1000 mM NaCl 0.05% Tween-20 Store at 4°C (supplemented with 1mM DTT and protease inhibitors on use)
PNK Buffer	70 mM Tris-HCl pH 7.5 10 mM MgCl <sub>2</sub> 0.05% NP-40 Store at 4°C (supplemented with 1mM DTT and protease inhibitors on use)

#### Immunoprecipitation (IP)-

Since our protein of interest was fused with GFP, the GFP trap beads were washed 2x with 1 mL cold TBS-T. To each sample, the washed and equilibrated GFP-trap beads were added and samples incubated for 2 hrs at 4°C on rotation. After IP, the samples were centrifuged at 2,000 rpm for 5 mins at 4°C and the flow through was discarded. The beads were washed 4x with TBS1000-T (transferred to fresh tubes after third wash) and 2x with PNK buffer.

#### On-bead phosphatase treatment

<b>1x phosphatase mix</b>	<b>20x</b>
4 µL 10x phosphatase buffer (Fermentas)	80 µL
1.5 µL SAP (1U/µL, Fermentas)	30 µL
0.5 µL RNaseOUT (40U/µL, Invitrogen)	10µL
43 µL DMPC-water	680 µL
40 µL total vol.	40 µL each

To the beads 40 µL of the 20x phosphatase mix was added and incubated at 37°C for 20 mins. Beads were washed 2x with TBS400-T (eppis were changed in second wash) and 2x with PNK buffer.

## On-bead 3'-end RNA linker ligation

<b>1x ligase mix:</b>	<b>20x</b>
1.5 $\mu$ L 3'-linker (100 pmol/ $\mu$ L)	30 $\mu$ L
2.5 $\mu$ L 10x T4 RNA-ligase buffer (Thermo)	50 $\mu$ L
0.65 $\mu$ L RNA-ligase (10U/ $\mu$ L, Thermo)	13 $\mu$ L
0.25 $\mu$ L RNaseOUT (40U/ $\mu$ L, Invitrogen)	5 $\mu$ L
2.50 $\mu$ L BSA (1mg/mL, The)	50 $\mu$ L
17.60 $\mu$ L DMPC-water	352 $\mu$ L
25 $\mu$ L total vol.	25 $\mu$ L each

To the beads, 25  $\mu$ L of 1x ligase mix was added and incubated at 16°C in a thermomixer o/n. Samples were washed 2x with PNK buffer.

## On-bead PNK treatment:

<b>1x PNK mix</b>	<b>20x</b>
2 $\mu$ L 10x PNK buffer (NEB)	40 $\mu$ L
2 $\mu$ L ATP (SCP801; 12.33 pmol/ $\mu$ L)	40 $\mu$ L
0.5 $\mu$ L T4 PNK Enzyme (NEB)	10 $\mu$ L
0.5 $\mu$ L RNaseOUT (40U/ $\mu$ L, Invitrogen)	10 $\mu$ L
15 $\mu$ L DMPC-water	300 $\mu$ L
20 $\mu$ L total vol.	20 $\mu$ L each

To the beads, 20  $\mu$ L of 1x PNK mix was added and incubated at 37°C for 20 mins in a thermomixer. Beads were washed 1x with cold TBS-T and 1x with cold PNK buffer.

## SDS PAGE and Blotting-

The beads were mixed with 20  $\mu$ L of 2x LDS loading buffer supplemented with DTT. Samples were denatured for 10 mins at 70°C while shaking and loaded on 4-12% Bis-Tris gel along with a pre-stained size marker and electrophoresed at 200 V for 5 hrs. After this the proteins were transferred to a nitrocellulose membrane by blotting o/n at 15 V at 16°C. The protein size marker bands were drawn on a plastic bag using a radioactive pen after placing the plastic bag on top of the membrane. Expose the membrane whilst in the plastic bag to X-ray

film for 16 hrs. The X-ray films were developed and the autoradiograms were used to detect the protein-RNA complex of interest. The corresponding area was marked on the nitrocellulose membrane and further processed.

RNA isolation-

The selected area in the nitrocellulose membrane containing the protein-RNA complex of interest with intermediate RNase I ( $10^{-4}$ ) treatment was cut along with the corresponding area in the control lane. To the nitrocellulose membrane pieces, 400  $\mu$ L of PNK buffer was added and 20  $\mu$ L of Proteinase K (20 mg/mL). Samples were incubated at 37°C for 20 mins with 1,000 rpm shaking. 400  $\mu$ L of PK buffer and 7 molar (M) urea were added to each tube and incubated for a further 20 mins at 55°C with shaking at 1,000 rpm.

CLIP-PK Buffer

<b>Buffer</b>	<b>CLIP-PK Buffer (Final conc.)</b>	<b>Volume (mL)</b>
1M Tris-HCl pH 7.4	100 mM	1.5
5M NaCl	50 mM	0.150
0.5M EDTA	10 mM	0.300
10% SDS	1%	1.5

CLIP-PK-Urea Buffer

<b>Buffer</b>	<b>CLIP-PK-Urea Buffer (Final conc.)</b>	<b>Volume (mL)</b>
1M Tris-HCl pH 7.4	100 mM	1.5
5M NaCl	50 mM	0.150
0.5M EDTA	10 mM	0.300
Urea	7M	6.306 grams

To each tube, 800  $\mu$ L of RNA-grade phenol-chloroform-isoamylalcohol (25:24:1) was added and samples incubated at 30°C for 5 mins with shaking at 1,000 rpm. To separate the aqueous layer containing the RNA, samples were centrifuged at 13,000 rpm for 5 mins at RT using phase-lock gel (5PRIME). The aqueous layer was transferred to a new tube without disturbing the gel. RNA was precipitated by adding 1  $\mu$ L glycoblue (Ambion, 9510), 80  $\mu$ L of 3 M sodium acetate (NaOAc) pH 5.5 and 0.7 vol. ice-cold isopropanol, mixed and incubated

o/n at -80°C. Samples were centrifuged at 14,800 rpm for 20 mins at 4°C, supernatant was discarded, RNA pellet washed with 0.9 mL of 80% ethanol, centrifuged again full speed for 5 mins. The RNA pellet was air-dried for 3 mins and dissolved in 5 µL of RNase-Free water and transferred to a PCR tube.

Reverse Transcription (RT)-

To the RNA sample, 1 µL (0.5 pmol/µL) RT-oligo primer was added along with 1 µL (10 mM) dNTP mix.

RT Thermal program

Temperature	Incubation time
70°C	5 mins
25°C	Hold

To this solution of 7 µL, add the RT mix as-

RT mix

Components	Volume
Water	7 µL
5x RT Buffer	4 µL
0.1M DTT	1 µL
RNase inhibitor (40U/µL)	0.5 µL
Superscript III	0.5 µL

The reverse transcription reaction was set in a thermocycler with following temperature and time cycles-

Temperature	Incubation time
25°C	5 mins
42°C	20 mins
50°C	40 mins
80°C	5 mins
4°C	Hold

At the end of RT, 1.65  $\mu$ L of 1 M sodium hydroxide (NaOH) was added and samples incubated at 98°C for 20 mins. To eliminate the radioactivity, 20  $\mu$ L of 1 M HEPES-NaOH pH 7.3 was added. The samples were cleaned up using MyONE clean-up kit following the kit protocol which allows cDNA to bind the beads, beads are washed and then 5  $\mu$ L of water was added directly without removing the beads.

Second linker ligation to the 3'-end of the cDNA-

To the cDNA-bead solution, 2  $\mu$ L (10  $\mu$ M) of second adapter and 1  $\mu$ L DMSO (100%) were added. Samples were heated to 75°C for 2 mins and immediately cooled on ice for 1 min. We had a total of four tubes, therefore, we added four different second linkers as-

Sample	Linker sequence
dMi-2-GFP replicate -1	NNNNNGCCTAANNNN (#3)
untagged control replicate -1	NNNNNTCAAGTNNNN (#8)
dMi-2-GFP replicate -2	NNNNNAAGCTANNNN (#10)
untagged control replicate -2	NNNNNCGAAACNNNN (#21)

Ligation-mix

Sample	Volume
Water	0.3 $\mu$ L
10x RNA ligase buffer (with DTT) (NEB)	2 $\mu$ L
100 mM ATP	0.2 $\mu$ L
50% PEG 8000	9 $\mu$ L
High conc. RNA Ligase	0.5 $\mu$ L

To the 8 $\mu$ L cDNA-second linker mixture, 12  $\mu$ L ligation mix and an additional 1  $\mu$ L of RNA ligase was added (final vol. 21  $\mu$ L). Samples were mixed by stirring and agitation o/n at 22°C with 1,100 rpm shaking. The cDNA ligated to the second linker were cleaned up using MyONE clean up kit following the kit protocol. Like in the first MyONE clean up step, the cDNA were coupled to the beads, washed, air-dried and cDNA eluted in 23  $\mu$ L of water. This time the elute was separated from the beads into fresh vials.



First Polymerase chain reaction (PCR, cDNA amplification)

Sample	Volume
cDNA	22.5 $\mu$ L
P5Solexa_s and P3Solexa_s primer mix (10 $\mu$ M each)	2.5 $\mu$ L
2x Phusion HF MasterMix	25 $\mu$ L

The samples were incubated in a thermocycler with following program-

Temperature	Incubation time (seconds)
98°C	30
98°C	10
65°C	30
72°C	30
72°C	3 mins
16°C	Hold

} 6 cycles

The excess primers and primer-dimers were removed in a subsequent size selection step using ProNex size selection kit following the kit protocol. The sample to ProNex bead ratio was determined (1:2.95) and this ratio was used for the samples. At this step, the cDNA fragments longer than 20 nucleotides (nt) were retained while smaller fragments were discarded. This allows unambiguous mapping of the resultant reads. A 20 nt cDNA fragment is extended to a length of 75 nt due the addition of two linkers. All the products below 75 nt were excluded from ProNex selection.

Second PCR amplification (Library preparation)

The cDNA fragments selected after first size selection were mixed with PCR mix as-

Sample	Volume
Water	8 $\mu$ L
cDNA ( $\geq$ 75 nts)	10 $\mu$ L
P5Solexa and P3Solexa primer mix (10 $\mu$ M)	2 $\mu$ L
2x Phusion HF MasterMix	20 $\mu$ L

The samples were incubated in a thermocycler with following program-

Temperature	Incubation time (seconds)
98°C	30
98°C	10
65°C	30
72°C	30
72°C	3 mins
16°C	Hold

} 11 cycles

After the library preparative PCR, 4  $\mu$ L of each library was loaded on a 6% native PAGE and stained with EtBr to verify the quality of the library. The library was again subjected to size selection using the ProNex size selection kit following the kit protocol. At this stage of amplification, the size of a 20 nt cDNA fragment is extended to 155 nt (P5Solexa and P3Solexa primers are each 60 nt). Therefore, fragments longer than 155 nt were retained and the shorter fragments were discarded along with primer-dimers and excess primers. The final library was quantified using a Bioanalyzer and sequenced.

### 2.2.3 Bioinformatics Pipeline

The raw iCLIP sequencing reads were assigned to the different libraries by the specific barcode part of the unique molecule identifier (UMI) and deduplicated based on total sequence identity with an in-house script. UMIs were removed from the read sequences and added to the read identifier. Additionally, poly(A) tails with at least eight consecutive A's were removed from reads and trimmed reads longer than 19 nucleotides were aligned to the

*Drosophila melanogaster* genome (dm6) using bwa-mem (0.7.17-r1188, arXiv:1303.3997v2) with minimum score output (-T) set to 19. Alignments were deduplicated based on UMIs and mapping position using UMI-tools (1.0.1, doi:10.1101/gr.209601.116) dedup functionality with default parameters. Only reads mapping uniquely to the genome were considered for further analysis while the rest of the reads were discarded. The uniquely mapped reads were attributed to transcript annotations based on their genomic coordinates using featureCounts from the SubRead (v1.6.4, doi:10.1093/bioinformatics/btt656) software suite for non-overlapping features (-O) and only counting strand specific alignments (-s 1).

Sample Correlation- The read counts from the two biological replicates were checked for correlation by calculating the squared Pearson Correlation Coefficient for paired samples. For visualization a linear model was fitted using the same metric.

Metagene Analysis- A metagene representation was generated based on read starts from both libraries using bamCoverage, calculateMatrix and plotProfile from the deepTools suite (3.4.3, doi:10.1093/nar/gkw257) with options “-bs 1 --ignoreDuplicates --minMappingQuality 1 --exactScaling --Offset 1”. Additionally, a detail graphic for the TTS of all transcripts with iCLIP reads was generated by manually plotting cross link positions derived from alignment start position minus one within +/-500 nucleotides of the TTS. Transcripts were sorted by mapped number of reads and their proximity to TTS. Relative density of cross links was calculated for the respective region.

Intron-Exon coverage- The cumulative coverage across the last intron-exon junction of all transcripts with at least two exons was determined using a custom script and plotted for +/- 10 nucleotides around the junction.

DE-Seq analysis- To rule out abundant iCLIP targets due to overall expression in the cell, we used unbiased RNA-Seq libraries from the same cell line (lenz et. al., 2021) as a reference in a relative expression analysis. Reads were mapped identical to iCLIP reads but without prior trimming. Uniquely mapping reads were then counted using featureCounts as described above. In contrast to iCLIP, RNA-Seq read counts were normalized to FPKM (fragments per kilobase in transcript per million reads in library) to account for the difference in protocols (there is one read per cross-linked transcript molecule in iCLIP, while RNA-Seq can produce multiple reads per molecule especially for longer transcripts).

DESeq2 (doi:10.1186/s13059-014-0550-8) was used to identify mRNAs stably captured in iCLIP experiments than expected by proportional abundance in RNA-Seq. We chose a threshold for adjusted p-value of  $< 0.05$  and putative log fold change  $> 4$  to identify mRNAs that we regard as sufficiently enriched in iCLIP above expected abundance levels.

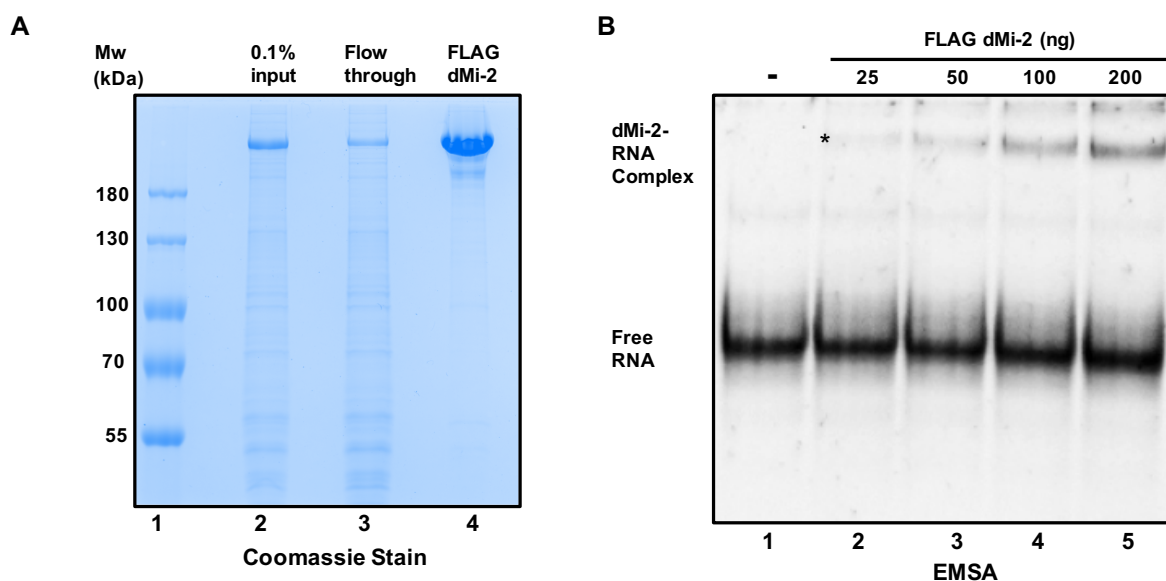
Position relative to cross-link site- Genomic sequences  $\pm 200$  nucleotides around all identified cross-linking sites were extracted and absolute as well as relative base content per relative position to the cross-linking site calculated to identify base specific biases at the cross-linking site. Relative base content per position was converted to percent of the median content for each base within the 400-nucleotide window. Additionally, a sequence logo was generated with genomic sequences of  $\pm 20$  nucleotides around the cross-linking site using kpLogo (v1.1, doi:10.1093/nar/gkx323) and parameters “-gapped -startPos 21”.

## 3 Results

### 3.1 Chapter 1

#### 3.1.1 dMi-2 is an RNA-binding protein

Murawska et al., 2011, suggested dMi-2 to be recruited to heat shock genes through a 2-step mechanism in which a possible interaction with RNA *in vivo* is postulated to play an important role. dMi-2 was shown to bind RNA by Murawska et al., 2011, however, the molecular aspects underlying the binding to RNA were not studied in detail. I sought to characterize the molecular basis of dMi-2 interaction with RNA and started by re-producing the dMi-2 RNA binding experiment in Murawska et al., 2011. I generated the 230 nucleotide RNA fragment, used in that study, by *in vitro* transcription from Hsp70Aa gene. I infected *Spodoptera frugiperda*-9 (Sf9) cells with Baculovirus expressing N-terminally FLAG-tagged dMi-2. The recombinant FLAG dMi-2 was purified by FLAG-immunopurification using anti-FLAG-M2 beads and the protein was eluted using FLAG-peptide. (**Figure 3.1.1A, Lane 4**). I performed an electrophoretic mobility shift assay (EMSA) using the recombinant dMi-2 and the RNA. If a protein binds the RNA, the protein-RNA complex that is formed is separated from the unbound RNA on a native gel wherein the bound complex appears as a shifted/retarded band. When I added dMi-2 to the RNA, I saw a shifted band (**Figure 3.1.1B, lane 2, asterisk mark**) and as I increased the amount of dMi-2, I observed an increased appearance of an RNA-dMi-2 complex confirming that, indeed dMi-2 is an RNA-binding protein *in vitro* (**Figure 3.1.1B**).



**Figure 3.1.1: dMi-2 binds RNA *in vitro***

**A)** FLAG Immunoprecipitation of N-terminally FLAG tagged dMi-2 protein. 15  $\mu$ L of eluted dMi-2 (lane 4) was loaded on an SDS gel along with 0.1% input (14.4  $\mu$ g) and 14.4  $\mu$ g of flow through (lanes 2 and 3, respectively) and electrophoresed. The gel was stained with Coomassie Brilliant Blue. The additional faint bands running below the main (dMi-2) band are likely the degradation products of dMi-2. Molecular weight marker was loaded in lane 1. **B)** Native polyacrylamide gel showing the EMSA of a 230-mer RNA fragment *in vitro* transcribed from Hsp70Aa gene. The asterisk in lane 2 marks the dMi-2-RNA complex. 160 ng (nanogram) of RNA in each reaction was incubated with 25, 50, 100 and 200 ng of FLAG-dMi-2 in lanes 2, 3, 4 and 5, respectively, and loaded on a native polyacrylamide gel. RNA only was loaded in Lane 1 as free RNA control and samples electrophoresed. The gel was stained with 1X SYBR Gold stain and imaged with Bio-Rad ChemiDoc touch system.

### 3.1.2 dMi-2 binds RNA promiscuously

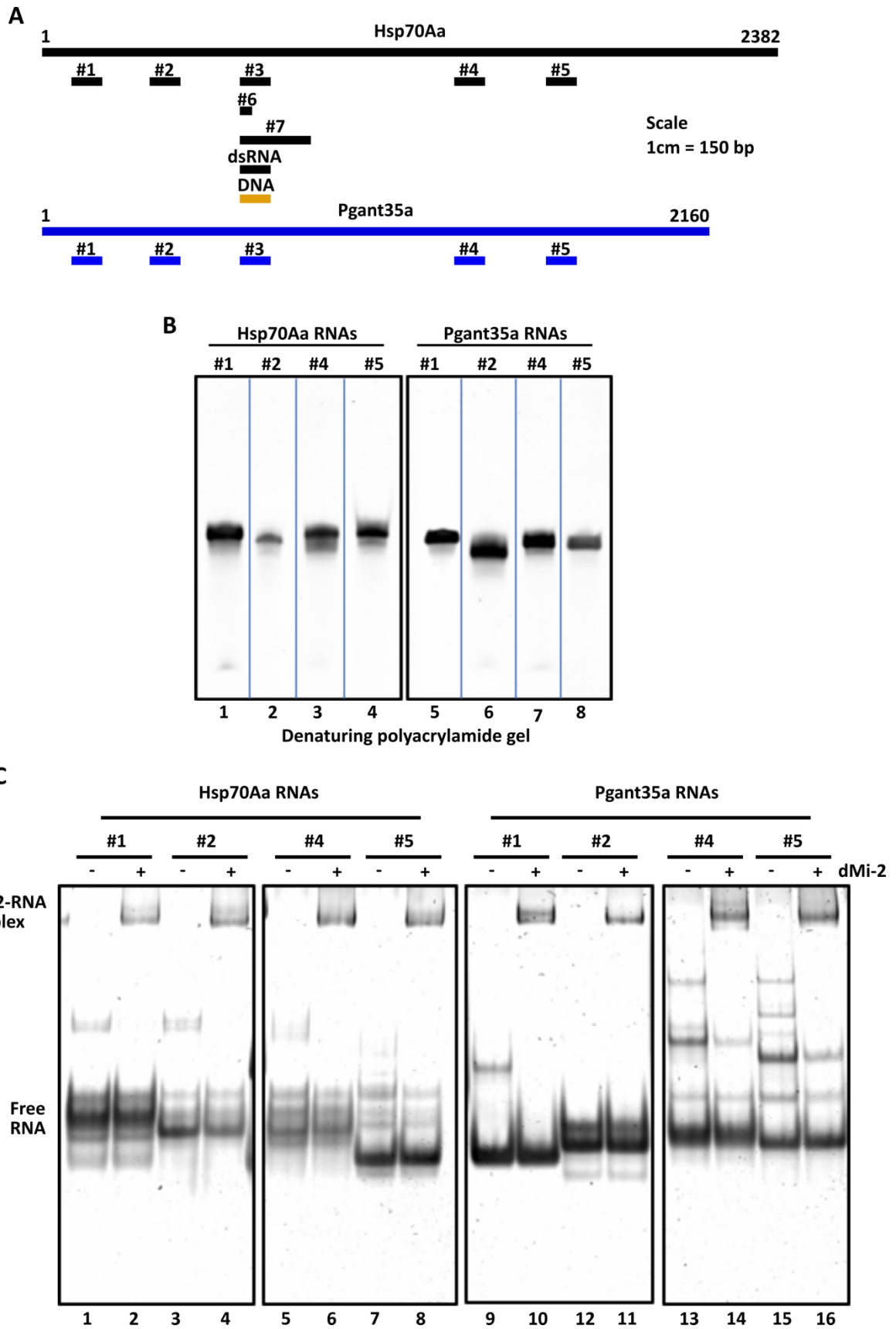
After confirming that *in vitro* dMi-2 binds to a 230-mer RNA derived from the Hsp70Aa gene, I asked if this binding was specific to only that region of the Hsp70Aa RNA or if dMi-2 could bind to different regions of the Hsp70Aa RNA. Also, I asked if reducing the length of the RNA impacted the binding. To answer this, I generated four RNA fragments from different regions of Hsp70Aa gene by *in vitro* transcription, named, Hsp70Aa #1, #2, #4 and #5), each 100 nucleotides in length (**Figure 3.1.2A, upper panel**). Each RNA is different, in terms of

sequence, from the others and from the 230-mer RNA used in Figure 3.1.1B (**Figure 3.1.2A, upper panel**). The RNAs were loaded on a urea denaturing polyacrylamide gel and electrophoresed to verify the quality of the RNAs (**Figure 3.1.2B, left panel, lanes 1-4**), confirming that RNAs were of good quality appearing as a single band. The slight differences in running patterns between the RNAs could likely arise from incomplete denaturation of the secondary structures. The RNAs on the native EMSA gel appeared as multiple bands (**Figure 3.1.2C, lanes 1,3,5 and 7**) rather than as single band seen in Figure 3.1.2B, lanes 1-4. This could be due to impurity, degradation or different secondary structure conformations of the RNAs. However, since, these RNAs form a single distinct band in the denaturing gel, we can conclude that the multiple bands on the native gel are very likely due to different conformations of the RNAs. Indeed, predicting the secondary structures of the RNAs using SPOT-RNA software suggests that they can form distinct folded structures (**Supplementary Figure S3.1.2 A-D**). Upon testing these RNAs in EMSA, dMi-2 bound to each of the RNAs (**Figure 3.1.2C, lanes 2, 4, 6 and 8**) irrespective of their sequence and predicted structural differences. These results suggest that dMi-2 can bind to different regions of Hsp70Aa RNA which are composed of different sequences and that reducing the length of RNAs from 230 nucleotides to 100 nucleotides did not impact the binding of dMi-2 to RNA.

Having established that dMi-2 binds to different regions within Hsp70Aa RNA, I asked if the binding of dMi-2 to RNA is specific to Hsp70Aa RNA only or if it binds to other RNAs as well. To answer this, I selected Pgant35a, an unrelated gene which produces a transcript of similar length as Hsp70Aa (**Figure 3.1.2A, lower panel**). I generated four RNAs from this gene, each 100 nucleotides in length, by *in vitro* transcription and electrophoresed them on urea denaturing polyacrylamide gel to check their quality (**Figure 3.1.2B, lanes 5-8**). All the RNAs ran as distinct single bands on the denaturing gel suggesting the purity and good quality of the RNAs. Each RNA has a distinct sequence that is different from the other and from the RNAs that were generated from Hsp70Aa gene. I also predicted the secondary structures of these RNAs using SPOT-RNA software (**Supplementary Figure S3.1.2 E-H**). Each RNA also has a distinct predicted secondary structure. Similar to the RNAs generated from Hsp70Aa gene, the RNAs generated from Pgant35a gene, also formed multiple bands on the native gel (**Figure 3.1.2C, lanes 9, 11, 13, and 15**), indicating the possibility of formation of several conformations in native conditions as opposed to denaturing conditions. I tested dMi-2 for binding to these RNAs in EMSA (**Figure 3.1.2C, lanes 9-16**). Interestingly, dMi-2 bound to all of the RNAs generated from Pgant35a gene (**Figure 3.1.2C, lanes 10, 12, 14 and 16**),

suggesting the promiscuous nature of dMi-2 interaction with RNA which indicates that dMi-2 could bind to a diverse set of RNAs. This hints toward the possibility of dMi-2 being a general RNA binding protein which could potentially bind to several RNAs *in vivo* as well.





### Figure 3.1.2: dMi-2 binds RNA promiscuously

**A)** Schematics representing the regions of Hsp70Aa (upper panel) and Pgant35a (lower panel) genes from which RNAs were *in vitro* transcribed. The orange represents the Hsp70Aa 100bp double stranded DNA **B)** 150 ng of RNAs derived from Hsp70Aa and Pgant35a genes were loaded on urea denaturing polyacrylamide gels (Lanes 1-4 and 5-8, respectively) and electrophoresed. **C)** EMSAs showing that dMi-2 binds all the RNAs from (B) promiscuously. For EMSAs 150 ng of each RNA and 500 ng of dMi-2 was used per reaction. RNA only as control was loaded in lanes 1, 3, 5, 7, 9, 11, 13 and 15, and electrophoresed. Both the gels were stained with 1X SYBR Gold and imaged with Bio-Rad ChemiDoc touch system.

### 3.1.3 dMi-2 binds some nucleic acids with higher affinity than others *in vitro*

Having established that dMi-2 can bind to several RNA molecules *in vitro*, I asked if there are quantitative differences in binding affinity. To this end, I generated 100-nucleotide RNAs from Hsp70Aa and Pgant35a genes, named Hsp70Aa #3 and Pgant35a #3. (**Figure 3.1.2A**). Both of these RNAs were tested for quality (**Supplementary Figure S3.1.3 A and B, respectively**). Each RNA appeared as a single discreet band suggesting the purity of the RNAs. Hsp70Aa #3 RNA runs slightly higher than the 100-nucleotide single stranded RNA size marker, that could perhaps be due to incomplete denaturation which might allow some form of the secondary structure to form leading to an unexpected running behaviour. I used 250 nanomolar (nM) of each RNA and titrated dMi-2 from 0.1 to 1.5 micromolar ( $\mu\text{M}$ ) in the EMSA. As expected, dMi-2 bound to both the RNAs (Figure 3.1.3A, lanes 2-10 and lanes 12-20). In order to determine the quantitative differences in binding affinity of dMi-2 between these two RNAs I quantified the bound fraction and the unbound fraction of RNAs in each well using the Bio-Rad Image Lab quantification program. The binding affinity of each RNA was measured in terms of the approximate dissociation constant (app.  $K_d$ ), which measures the concentration of the protein at which half of the RNA probe is bound by the protein. Hence, the lower the  $K_d$  value, the stronger the affinity of the protein to that RNA. The quantification of the EMSA results showed that dMi-2 bound the Hsp70Aa #3 RNA better than the Pgant35a #3 RNA (Figure 3.1.3B). The dissociation constant (app.  $K_d$ ) of Hsp70Aa #3 RNA was  $0.38 \pm 0.06$  ( $\mu\text{M}$ ) while it was  $0.63 \pm 0.28$  ( $\mu\text{M}$ ) for pgant35a #3 RNA, suggesting that dMi-2 has nearly 2-fold higher affinity for Hsp70Aa #3 RNA over the Pgant35a #3 RNA. From this data, I conclude that, indeed, dMi-2 could bind some RNAs with higher affinity over other RNAs *in*

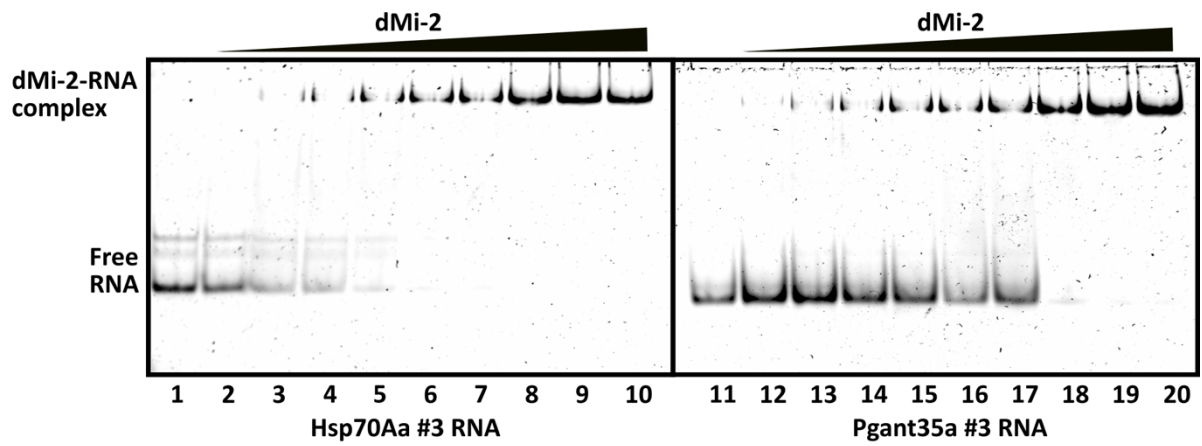
*vitro*. This difference in the affinity could arise either from sequence of the RNA that dMi-2 prefers binding to or their secondary structures.

dMi-2 has been shown to bind to double stranded DNA (dsDNA) by (Bouazoune et al., 2002). I asked if dMi-2 binds double stranded RNA (dsRNA) as well and whether the binding affinities are different between dsRNA and dsDNA. To this end, I used the 100bp DNA sequence of Hsp70Aa #3 fragment from above as template to generate a double stranded RNA by *in vitro* transcription using two T7 promoters, one on each strand. The template DNA of this fragment without T7 promoters was used as double stranded DNA (Supplementary Figure S3.1.2C), hence, these two nucleotide fragments are analogous in sequence to each other. To ascertain the formation of double stranded RNA, I treated the dsRNA with or without RNase III. RNase III is a ribonuclease that specifically cleaves dsRNA. I used Hsp70Aa #3 single stranded RNA as a control (Supplementary Figure S3.1.2D). The RNase III specifically cleaved the dsRNA (Supplementary Figure S3.1.2D, lane 4) and expectedly had no effect on the ssRNA (Supplementary Figure S3.1.2D, lane 3). Having substantiated the double stranded nature of the Hsp70Aa 100bp dsRNA, I compared the binding of FLAG dMi-2 to dsRNA and dsDNA fragments with analogous sequences in EMSA experiments. I observed that dMi-2 binds to dsRNA (Figure 3.1.2C, lanes 1-10) in addition to dsDNA (Figure 3.1.2C, lanes 11-20). dMi-2 bound to both the nucleic acids with similar dissociation constants (app.  $K_d$ ) of  $0.86 \pm 0.01$  ( $\mu\text{M}$ ) for dsRNA and  $0.82 \pm 0.01$  ( $\mu\text{M}$ ) for dsDNA (Figure 3.1.2D), suggesting that dMi-2 has similar binding affinity to both dsRNA and dsDNA of 100bp length *in vitro*.

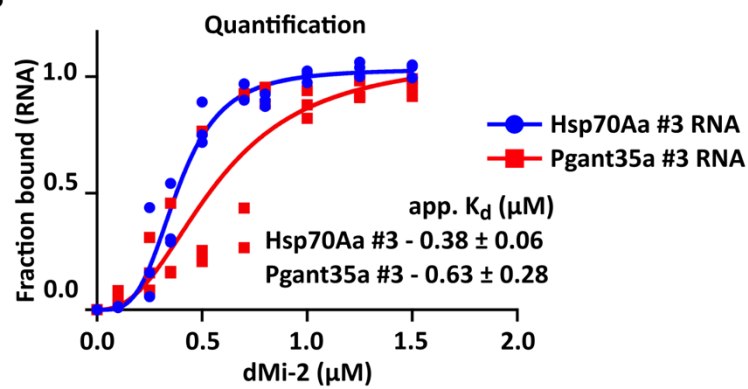
**Table 1: Approximate dissociation constants for binding of various nucleotides to dMi-2**

Nucleic acids	app. $K_d$ ( $\mu\text{M}$ )
Hsp70Aa 100-mer ssRNA	$0.38 \pm 0.06$
Pgant35a 100-mer ssRNA	$0.63 \pm 0.28$
Hsp70Aa 100bp dsRNA	$0.86 \pm 0.01$
Hsp70Aa 100bp dsDNA	$0.82 \pm 0.01$

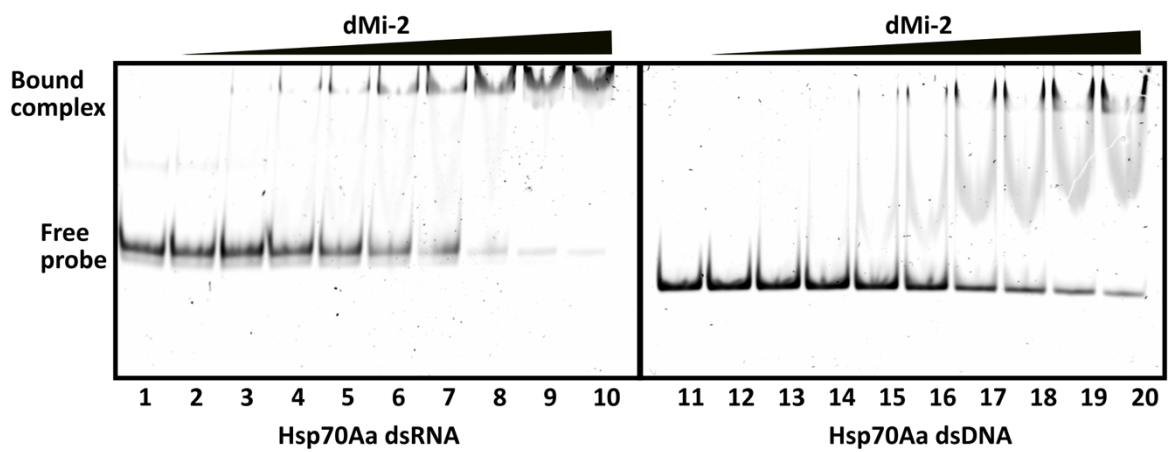
A



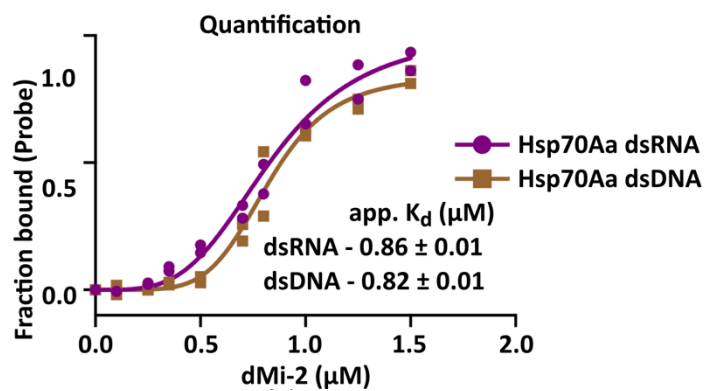
B



C



D



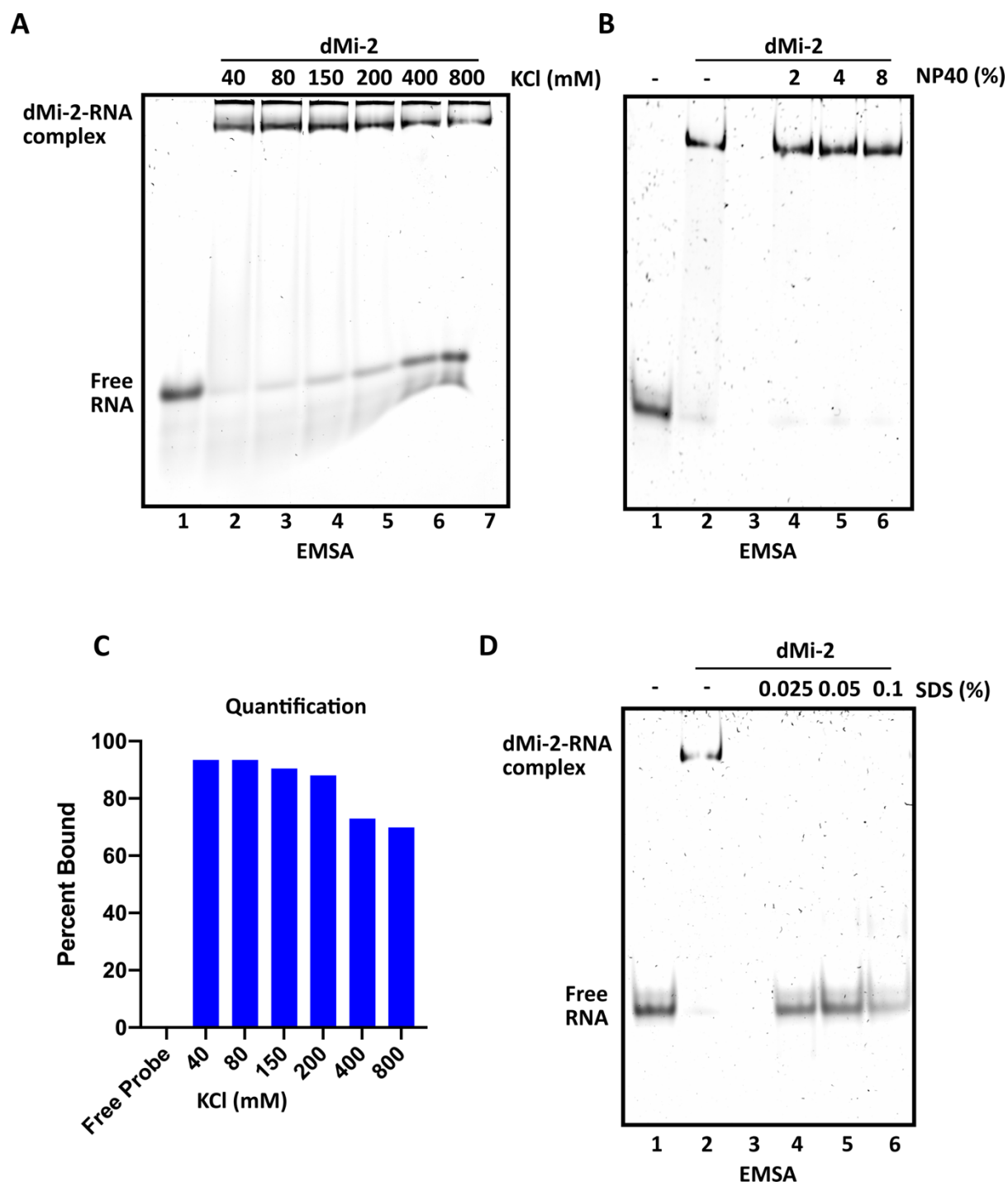
**Figure 3.1.3: dMi-2 binds some nucleic acids with more affinity than others**

**(A)** EMSAs of ssRNAs Hsp70Aa #3 (lanes 1-10) and Pgant35a #3 (lanes 11-20). 250 nM RNA was used in each reaction and dMi-2 was titrated from 0.1 to 1.5  $\mu$ M. Representative image of 3 independent replicates is shown. **(B)** Quantification of the EMSA in (A). Data points from three independent biological replicates are shown. **(C)** EMSA of dsRNA and dsDNA from Hsp70Aa (100bp) which are analogous in sequence. 250 nM dsRNA or dsDNA was used for each reaction and dMi-2 was titrated from 0.1 to 1.5  $\mu$ M. Representative image of two independent replicates. **(D)** Quantification of the EMSA in (C). Data points from two independent biological replicates are shown. Each sample was loaded on a native polyacrylamide gel and electrophoresed. The gels were stained with 1X SYBR Gold and imaged with Bio-Rad ChemiDoc touch system. The image quantifications were performed using Bio-Rad Image-Lab software. The  $K_d$  were determined by using Hill Slope equation.

**3.1.4 dMi-2 binds RNA in high salt and detergent conditions**

The results above substantiate that dMi-2 is a general RNA binding protein *in vitro*. Next, I sought to determine if dMi-2 would be able to bind to RNA smaller than 100 nucleotides. I tested dMi-2 binding to 40-mer RNA, and it bound well, data not shown. Therefore, I decided to use 40-mer RNAs as well, henceforth. Next, I sought to determine the stability of the dMi-2-RNA complex. The standard EMSA reaction buffer contains 40 millimolar (mM) potassium chloride (KCl); to determine the salt sensitivity of dMi-2 binding to RNA, I increased the salt concentration in the EMSA reaction buffers up to 800 mM KCl (**Figure 3.1.4A**). The formation of some RNA-protein complexes such as pentatricopeptide repeat protein-10 (PPR-10)-RNA complex was shown by (McDermott et al., 2018) to be highly sensitive to salt with dissociation constant increasing nearly 450-fold from  $1.8 \pm 0.2$  picomolar (pM) at 75 mM salt to  $800 \pm 400$  picomolar (pM) at 300 mM salt. Similarly, Fox1, another RNA binding protein, was shown by (Auweter et al., 2006), to be influenced by salt in its RNA interaction with the dissociation constant ( $K_d$ ) increasing over 300-fold from 0.06 nM at 75 mM salt to 18.7 nM at 500 mM salt. Likewise, PRC2, a chromatin modifying protein that has been extensively characterized for RNA binding, loses the crosslinking to HOTAIR 400 RNA at 400 mM salt, as reported by (Wang et al., 2017). Surprisingly, over 60% of dMi-2-RNA complex can withstand even the maximum amount of salt that I tested i.e. 800 mM KCl in the EMSAs (**Figure 3.1.4A, upper**

**panel lane 7 and Figure 3.1.4C).** The electrophoresis of the gels in such high salt conditions causes retardation of the nucleic acids which is the reason for the free RNAs running unevenly between low salt (**Figure 3.1.4A, lanes 1-4**) and high salt conditions (**Figure 3.1.4A, lanes 5-7**). Also, the dMi-2-RNA complex is most stable at low to intermediate salt conditions i.e. 40 mM to 150 mM (**Figure 3.1.4A, lanes 2-4**), after which the complex appears to gradually disintegrate, with increasing free RNA appearing as the salt is increased, however, still retaining stability to the extent that over 60% of the RNA remains bound to dMi-2 even at 800 mM salt (**Figure 3.1.4A and Figure 3.1.4C**).



**Figure 3.1.4: dMi-2 binds RNA in high salt and detergent conditions**

**A)** EMSA showing binding of dMi-2 to RNA in increasing salt conditions. 34.5 ng of Hsp70Aa #6 RNA (Figure 3.1.2A) and 1.12  $\mu$ g dMi-2 was used for each reaction. RNA only was loaded in lane 1 as free RNA control. **B)** EMSA showing binding of dMi-2 to RNA in increasing NP40 conditions. 34.5 ng of Hsp70Aa #3 RNA and 763 ng of dMi-2 was used for each reaction. No NP40 was added in lanes 1 and 2. Lane 3 was left empty. **C)** Quantification of EMSA in (A). **D)** EMSA showing binding of dMi-2 to RNA in increasing SDS conditions. 34.5 ng of Hsp70Aa

#3 RNA and 763 ng of dMi-2 was used in each reaction. No SDS was added in lanes 1 and 2. Lane 3 was left empty. Each sample was loaded on native polyacrylamide gel and electrophoresed. Gels were stained with 1X SYBR Gold and imaged with Bio-Rad ChemiDoc touch system.

Further, I tested the effects of two detergents on the stability of dMi-2-RNA complex. I used NP40, a mild ionic detergent (**Figure 3.1.4B**) and sodium dodecyl sulfate (SDS), a harsher anionic detergent (**Figure 3.1.4D**). dMi-2-RNA complex could withstand 8% of NP40 in the reaction buffer. Beyond this concentration, the reactions became too viscous for proper pipetting. On the other hand, there was no formation of the dMi-2-RNA complex in presence of even 0.025% of SDS (**Figure 3.1.4D, lanes 4-6**). The possible explanation could be that the SDS either de-stabilized the complex or denatured the protein such that it no longer bound the RNA. To conclude, the data suggests that dMi-2-RNA interaction is very stable and the complex can withstand 800 mM salt and mild detergent conditions compared to several other RNA binding proteins including PPR-10, Fox-1, and PRC2, further substantiating that dMi-2 is a strong RNA binding protein.

### **3.1.5 Mapping the RNA binding region of dMi-2**

The dMi-2 protein is 1982 amino acids long (**Figure 3.1.5A and Supplementary Figure S3.1.5A**). It has a N-terminal region that consists of several domains including plant homeodomain (PHD) and chromodomains (CD). The C-terminal part of the protein is separated from the N-terminal part by an ATPase or Helicase domain (**Figure 3.1.5A**). I sought to identify the RNA binding regions in dMi-2. To that end, I used three dMi-2 truncation mutants - dMi-2  $\Delta$ N, dMi-2  $\Delta$ C and dMi-2 (691-1271) from (Bouazoune et al., 2002) (**Figure 3.1.5A**). All three mutant are N-terminally FLAG tagged. I infected Sf9 cells with Baculovirus expressing these truncation mutants and purified them by FLAG immunopurification. The proteins were eluted from the anti-FLAG beads using FLAG peptide (**Supplementary Figure S3.1.5A**). I performed EMSAs with these truncation mutants to test their binding to RNA. Often, protein-RNA complexes are resolved in the native polyacrylamide gel, however, with large proteins, there could be some complexes that do not enter the gel. Such complexes could be a result when two or more protein molecules bind the same RNA molecule. The resultant protein-RNA complex could be too large for gel resolution and therefore, such complexes remain stuck in the well. Such unresolvable complexes usually form with large



proteins. Indeed, we observe such complexes with FLAG dMi-2 wild type (WT), dMi-2  $\Delta$ N and dMi-2  $\Delta$ C, as pointed out by arrow heads (**Figure 3.1.5B, lanes 2, 3 and 4**). Therefore, it is important to consider the signal in the well as bound complex. The EMSA experiment shows that dMi-2 WT bound to RNA, as expected. The truncation mutant dMi-2  $\Delta$ C, which has the N-terminal region but doesn't have the C-terminal, also bound to the RNA (**Figure 3.1.5B, lane 4, Figure 3.1.5C and Supplementary Figure S3.1.5B, lanes 5 and 6**). Importantly, the dMi-2  $\Delta$ N mutant, which has no N-terminal part, largely lost binding to RNA with very faint signal from a gel resolved complex as pointed out by an asterisk (**Figure 3.1.5B, lane 3, asterisk mark, Figure 3.1.5C and Supplementary Figure S3.1.5B, lanes 3 and 4**) suggesting that the major RNA binding regions of dMi-2 reside within the N-terminal part of dMi-2. However, as marked by an arrow head, an unresolvable complex appears in the well (**Figure 3.1.5B, lane 3**). Also, since there is a loss of signal from the free RNA in lane 4 as compared with the RNA only control in lane 1, this mutant might as well bind to RNA to some extent, suggesting that some RNA binding regions of dMi-2 might reside outside of the N-terminal region. In case of dMi-2 (691-1271), which has the ATPase domain, no binding to RNA could be detected (**Figure 3.1.5B, lane 5 and Supplementary Figure S3.1.5B, lanes 7 and 8**) suggesting that this domain on its own does not bind RNA *in vitro*. From this data, I conclude that the major RNA binding regions of dMi-2 are within the N-terminal part of the protein.

Having located the major RNA binding region of dMi-2 in its N-terminal part, I asked which region within the N-terminal part contributes to the RNA binding of dMi-2. The first 376 amino acids of dMi-2 contain several positively charged amino acids (22.34%). If the binding of dMi-2 to RNA is entirely driven by charge, that is, interaction between positively charged amino acids with negatively charged nucleic acids, then the RNA binding region should be within the first 376 amino acids. To answer this, I created two new truncation mutants. First, I generated a truncation which consisted of only the first 376 amino acids, dMi-2 1-376 (**Figure 3.1.5A**). This region includes the HMG-box domain whose function remains unknown and has several positively charged amino acids. Next, I generated another truncation mutant which lacked the first 376 amino acids, dMi-2  $\Delta$ 376. This mutant contains the tandem PHD fingers and the chromodomains, the ATPase domain and the C-terminal part (**Figure 3.1.5A**) and it has only 15.81% positively charged amino acids. Both of these mutants were N-terminally FLAG-tagged. I infected Sf9 cells with Baculovirus expressing these truncation mutants. The proteins were purified by FLAG immunopurification and eluted with FLAG

peptide (**Supplementary Figure S3.1.5C and E, lane 2**). Next, I tested these mutants in EMSAs for binding to RNA. dMi-2 (1-376) bound to RNA (**Figure 3.1.5B, lane 6 and Supplementary Figure S3.1.5F, lanes 1-4**). On the other hand, the dMi-2  $\Delta$ 376 mutant also bound to RNA to some extent with a faint gel resolvable complex and an unresolved complex in the well (**Figure 3.1.5B, lane 7, arrow head**). The signal was stronger as compared to dMi-2  $\Delta$ N suggesting that this mutant binds RNA better than the dMi-2  $\Delta$ N mutant (**Figure 3.1.5C**). This potentially implicates the additional PHD and chromodomains in dMi-2  $\Delta$ 376 to play some role in RNA binding. These data suggest that the major RNA binding regions of dMi-2 are within the first 376 amino acids and that some RNA binding regions may reside outside of this region. It also supports the hypothesis that the binding to RNA may be entirely charge driven.

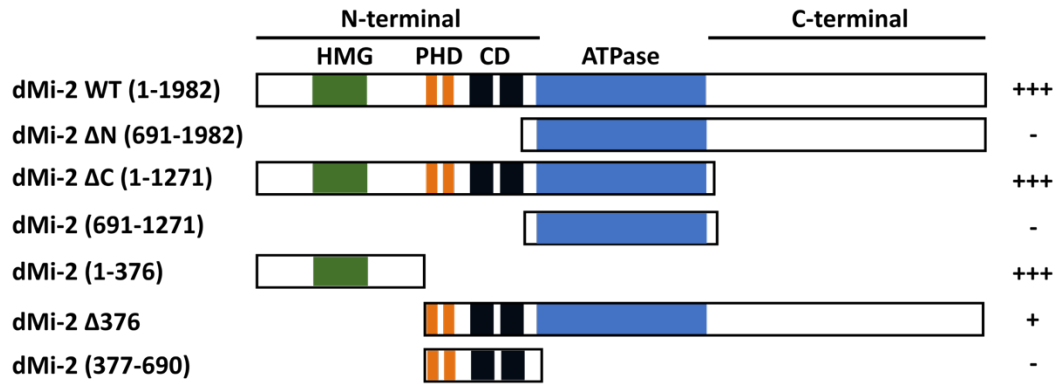
Since the above data indicate that the PHD fingers and the chromodomains (dMi-2 377-690) might as well bind to RNA *in vitro*. Therefore, to test that, I generated a N-terminally FLAG tagged dMi-2 (377-690) mutant and expressed this protein in Sf9 cells and FLAG immunopurified it (**Supplementary Figure S3.1.5E, lane 3**). I tested dMi-2 (377-690) for binding to RNA in EMSA. No RNA binding could be detected in the EMSA (**Figure 3.1.5B, lane 8, Figure 3.1.5C and Supplementary Figure S3.1.5F, lanes 5-7**), suggesting that dMi-2 (377-690) might not bind to RNA on its own, however, it may contribute to the RNA binding when fused to the rest of the protein. Together, I conclude that the first 376 amino acids of dMi-2 which has 22.34 % positively charged amino acids, largely appear to be sufficient for RNA binding with possibly additional contribution by PHD fingers and chromodomains (377-690).

Having tracked a major RNA binding region of dMi-2 to the first 376 amino acids, I went on to further characterize the RNA binding regions within the first 376 amino acids. To that end, I generated mutants containing only the first 117 amino acids, dMi-2 (1-117); dMi-2 (118-238) which contains the HMG-box domain of unknown function and dMi-2 (239-376) (**Figure 3.1.5D**). Each mutant was N-terminally FLAG tagged and expressed in Sf9 cells. The proteins were purified by FLAG immunopurification and eluted by FLAG peptide (**Supplementary Figure S3.1.5G**). Importantly, dMi-2 (1-117) and dMi-2 (118-238) contain similar numbers of positively charged amino acids, 17.94% and 17.35%, respectively, while dMi-2 (239-376) contains 26.08% positively charged amino acids. If the above hypothesis that the binding is entirely driven by charge is true, then dMi-2 (239-376) should bind to RNA the best while dMi-

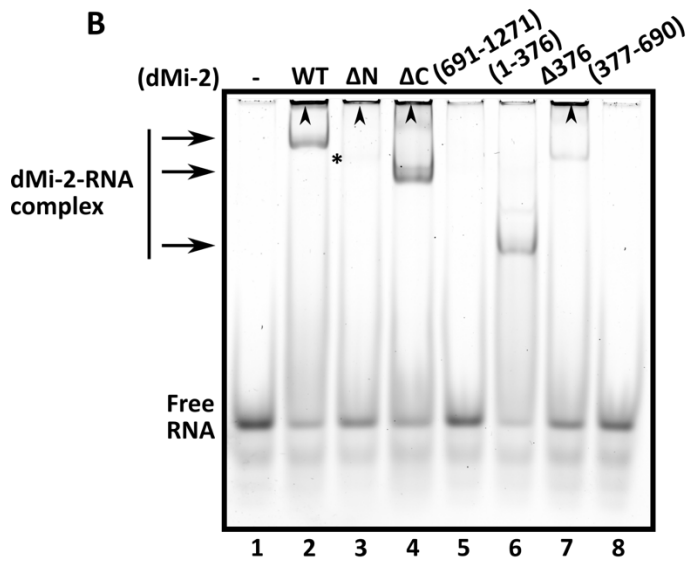
2 (1-117) and dMi-2 (118-238) should bind with equal affinity. Next, I tested these truncation mutant proteins for binding to RNA in EMSA. Interestingly, I observed that both fragments flanking the HMG-box domain bound to RNA while no RNA binding could be detected in dMi-2 (118-238) mutant in our assay (**Figure 3.1.5E**). This suggests that within the N-terminal region of dMi-2, the major contributors to RNA binding are likely 1-117 and 239-376 amino acids and that the binding to RNA is not entirely driven by charge. Several proteins, including chromatin regulators, bind to RNA without the presence of any canonical RNA binding domains. In such proteins, the presence of intrinsically disordered regions (IDRs) have been implicated in RNA binding. dMi-2 does not possess any canonical RBD binding domain and I observed that its RNA binding is not entirely driven by charge. Therefore, I asked if dMi-2 possesses IDRs which may contribute to its RNA interaction. To this end, I predicted the presence of IDRs in the region that my experimental data showed to be majorly involved in RNA binding, that is, residues 1-376. I used Predictor of Natural Disordered Regions (PONDR) software with default settings. The HMG-box (118-238) folds into a defined structure (**Figure 3.1.5F**). Indeed, the HMG-box was predicted as a well-defined structure with PONDR. The regions flanking HMG-box that appeared to bind RNA in the EMSA, that is, 1-117 and 239-376 amino acids, were predicted as IDRs suggesting that the RNA binding of dMi-2 may be contributed by the IDRs in the N-terminal regions.

Together, I conclude that major RNA binding sites of dMi-2 are in the first 376 amino acids in the N-terminal region and that 1-117 and 239-376 amino acids contribute to the RNA binding of dMi-2. The intrinsically disordered nature of these regions might play an important role in the RNA binding of dMi-2. This data does not rule out the possibility of some contributions to RNA binding outside of the N-terminal region.

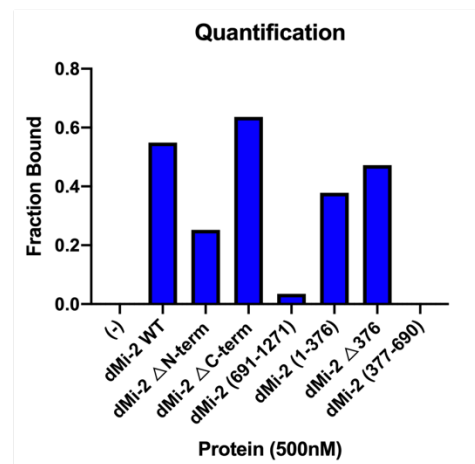
A



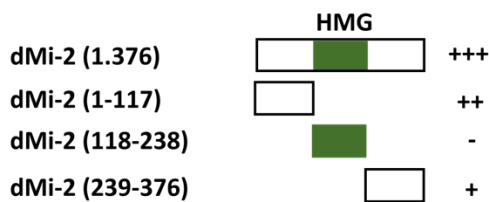
B



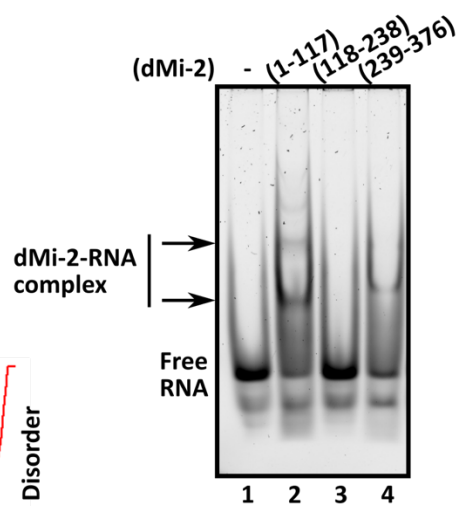
C



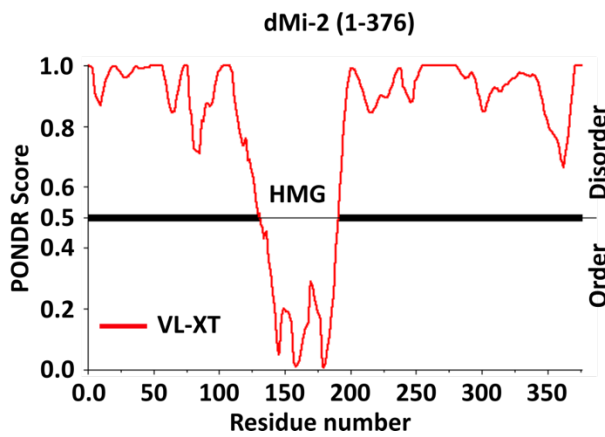
D



E



F



### Figure 3.1.5: Mapping the RNA binding region of dMi-2

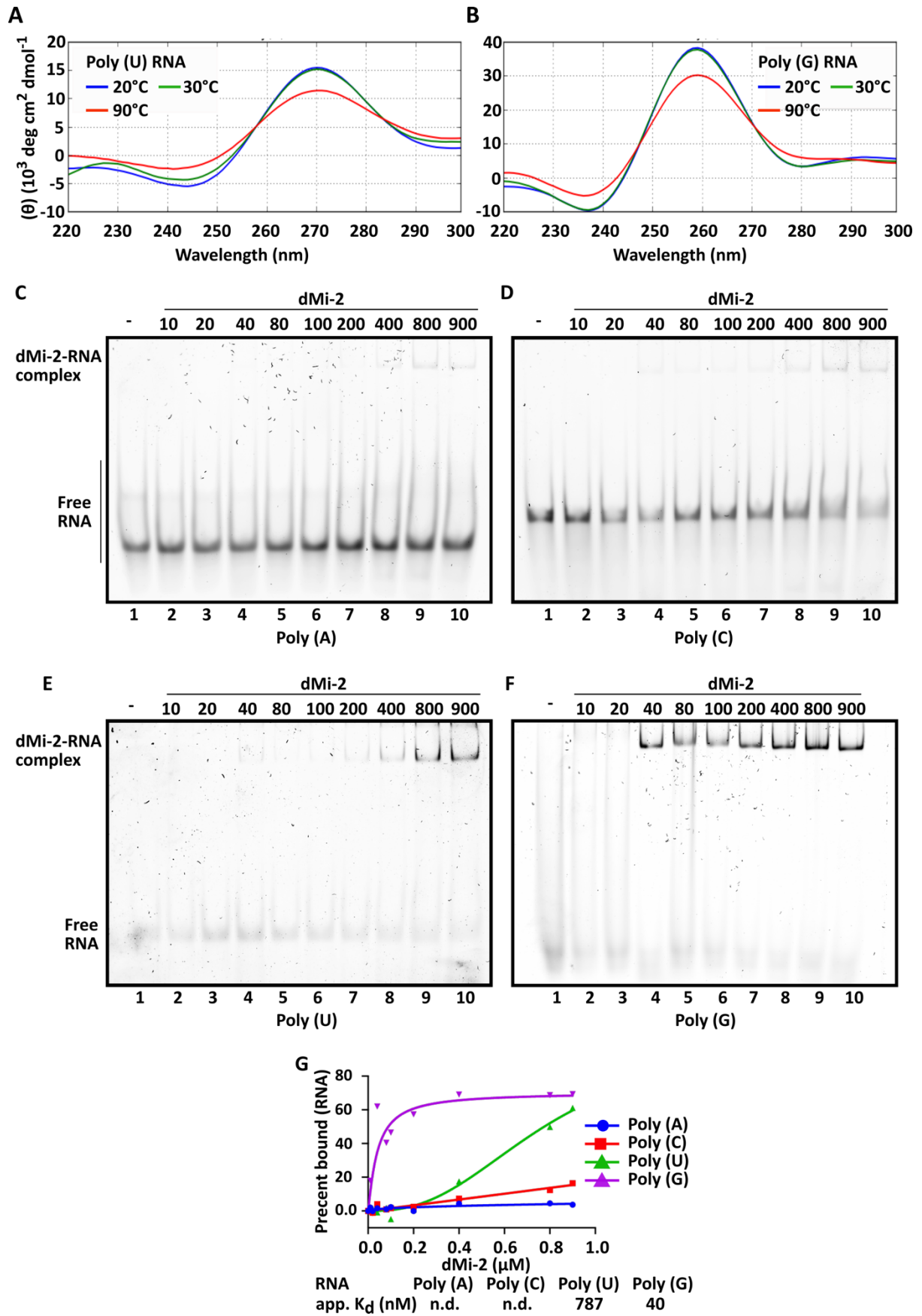
Schematics representing wild type and truncation mutants of dMi-2. High mobility group box (HMG), Plant homeodomains (PHD) and chromodomains (CD) and ATPase domains are marked (A). (B) EMSA of dMi-2 WT and mutants. 250 nM RNA was used for each reaction against 500 nM of protein. RNA only was loaded as control in lane 1. Reactions were loaded on a native polyacrylamide gel and electrophoresed. The bound complexes are shown by arrows on the left. The unresolved complexes in the well are pointed out by arrow heads in lanes 2, 3, 4 and 7. Asterisk mark in lane 3 points out a gel resolved complex. (C) Quantifications of the EMSA in (B). The bound fractions (gel resolved plus unresolved complex in the well) and unbound fractions were quantified using Bio-Rad Image Lab Software. (D) Schematics representing the truncation mutants of first 376 amino acids. HMG-box domain is denoted. (E) EMSA of dMi-2 (1-117), (118-238) and (239-376). 250 nM RNA was used for each reaction with 1.5  $\mu$ M proteins. RNA only was loaded as free RNA control in lane 1. Arrows on the left show the bound complexes. EMSA samples were loaded on native polyacrylamide gel and electrophoresed. Each gel was stained with 1X SYBR Gold and imaged with Bio-Rad ChemiDoc touch system. (F) Software prediction of IDRs in the N-terminal part of dMi-2. PONDR was used with optimal settings to predict IDRs. PONDR Score 1.0 (y-axis) indicates highest score for prediction of IDR while PONDR Score of 0.0 indicates a well folded and structure prediction. HMG-box region is predicted as ordered and highly structured region while regions comprising (1-117) and (239-376) are predicted as disordered or IDRs. VL-XT is the combination of three predictors. One is trained on Variously characterized Long disordered regions. Two predictors are trained on X-ray characterized Terminal disordered regions.

### 3.1.6 dMi-2 binding to homopolymer RNAs

Next, I asked if dMi-2 had a specific binding preference to a certain nucleotide over the others i.e. if dMi-2 preferred binding to Adenine (A), Uracil (U), Cytosine (C) or Guanine (G) nucleotides. To answer this, I used homopolymers of (A), (U), (C) and (G), each 40 nucleotides long. The repeats of Guanines of such a length have been shown by (Wang et al., 2017) to form a G-quadruplex structure. Therefore, I tested the Poly (G) RNA for the formation of G-quadruplex structures by circular dichroism (CD) spectroscopy. The characteristic

absorbance pattern for a G-quadruplex structure are a negative peak between 237-241 nm (nanometre) and a positive peak between 260-265 nm. I used the Poly (U) RNA as a control. It is unstructured and has been shown to produce a negative peak between 240-250 nm and a positive peak at 270 nm in CD spectroscopy. The absorbance pattern of my Poly (G) RNA in CD spectroscopy supports that it indeed formed a G-quadruplex structure with the absorbance peaks that are characteristic of a G-quadruplex structure (**Figure 3.1.6B**). The structure was stable even at temperature as high as 90°C. On the other hand, the Poly (U) RNA absorbance peaks suggested an unstructured RNA (**Figure 3.1.6A**).

Having established that my Poly (G) RNA formed a G-quadruplex structure, I tested the binding of dMi-2 to all homopolymer RNAs, including Poly (G) RNA, in EMSA. The structure of Poly (G) RNA doesn't allow it to form a discrete band in a native gel, it rather runs as a smear as compared to other RNAs which form discrete bands. dMi-2 bound to Poly (A) and Poly (C) RNAs with very low affinities (**Figure 3.1.6C – D, and 3.1.6G, quantification**). On the other hand, dMi-2 bound to Poly (U) and Poly (G) RNAs with comparatively higher affinities (**Figure 3.1.6E-F, and 3.1.6G, quantification**). It bound to Poly (U) RNA at higher concentrations and with an approximate dissociation constant (app.  $K_d$ ) of 787.8 nanomolar (nM). Importantly, dMi-2 bound to Poly (G) RNA with the highest affinity and had an approximate dissociation constant (app.  $K_d$ ) of 40.88 nanomolar (nM) (**Figure 3.1.6F-G**). These data suggest that dMi-2 has low affinities for binding to Adenine and Cytosine nucleotides while it has higher affinities for binding to Uracil and Guanine nucleotides, and that it, particularly prefers binding to Guanines, perhaps due to the G-quadruplex structure.



### Figure 3.1.6: dMi-2 binding to homopolymer RNAs

CD spectroscopy of poly (U) and Poly (G) RNAs, each 40-mer **(A)** and **(B)**. 10  $\mu$ M RNAs were allowed to fold in binding buffer and CD spectra measured between 220 to 300 nm. The data were analysed by CAPITO CD analysis and plotting software. The data were smoothed with Savitzky-Golay filter and CD values were converted to mean residue ellipticity ( $\theta$  in grad  $\text{cm}^2 \text{dmol}^{-1}$ ). Poly (G) RNA forms a characteristic G-quadruplex absorbance pattern with negative peaks between 237-241 nm and positive peaks between 260-265 nm. Poly (U) RNA was used as a control and it appears unstructured. The measurements were conducted at three different temperatures, 20°C (Blue), 30°C (Green) and 90°C (Red). At least 3 measurements were taken at each temperature and mean values plotted. **(C – F)** EMSAs of Poly (A), (U), (C) and (G). 52.8 nM RNA was used in each reaction and dMi-2 titrated from 10 nM to 900 nM. Samples were loaded on native polyacrylamide gels and electrophoresed. The gels were stained with 1X SYBR gold and imaged with Bio-Rad ChemiDoc touch system. **(G)** Quantification of EMSA gels in (C-F). The percentage of bound RNA fraction was quantified using Bio-Rad Image Lab Software.

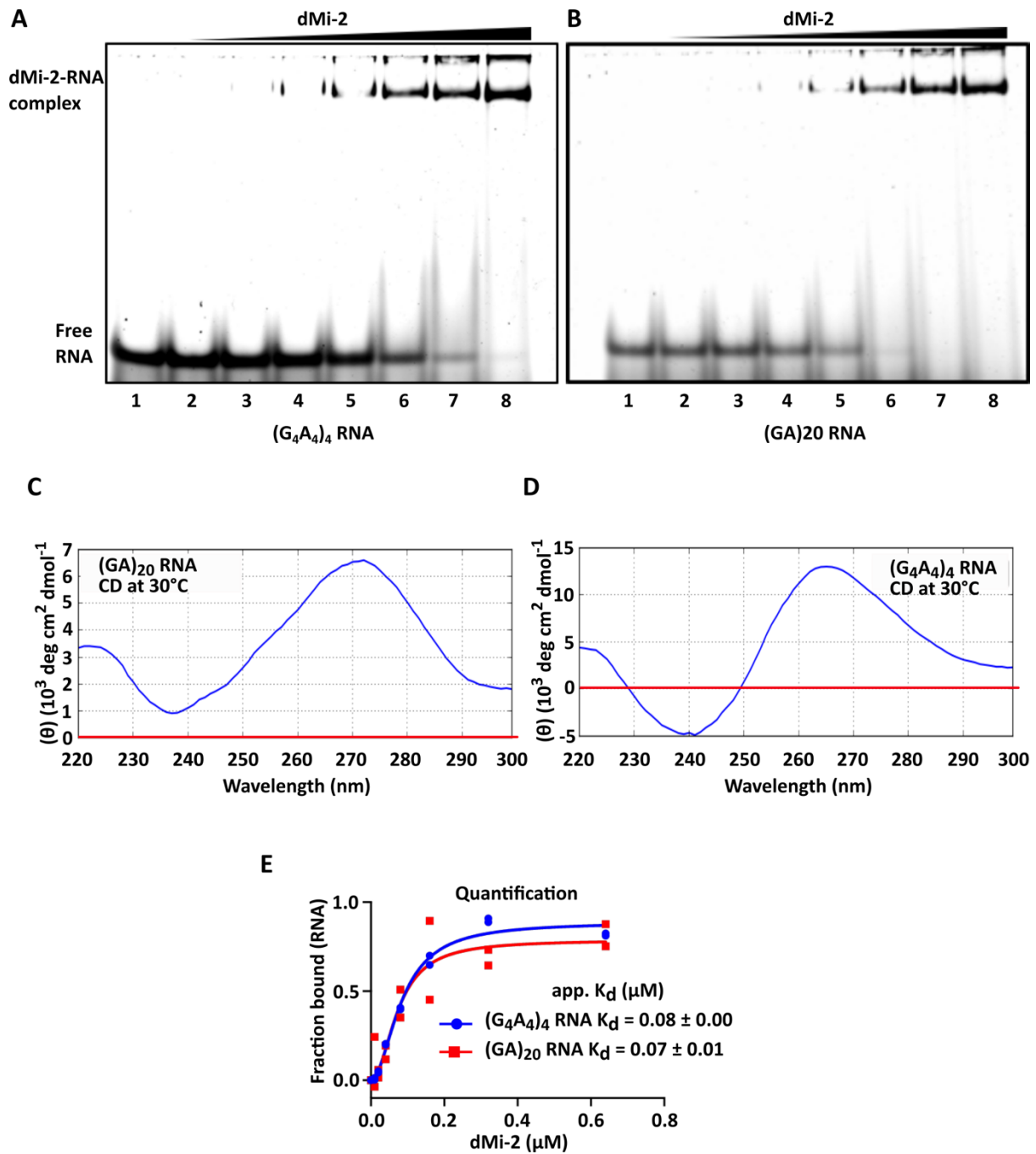
### 3.1.7 dMi-2 binds to G-rich RNA

Some RNA binding proteins, PRC2, for example, has been shown to have higher affinity for binding to G-quadruplex forming RNAs including Poly (G) of the same size as Poly (G) used in this study i.e. 40 nucleotides. Since dMi-2 bound better to Poly (G) RNA compared to Poly (U), (C) and (A) RNAs **(Figure 3.1.6)**, and since I showed that my Poly (G) RNA forms a G-quadruplex structure, I asked if the better binding to Poly (G) RNA was because of its structure. To address this, I used two RNAs -  $(G_4A_4)_4$  RNA and  $(GA)_{20}$  RNA. Each of these RNAs are 40 nucleotides long. When two or more tetrads or blocks of G-nucleotides are regularly interspersed by other nucleotides, for example, Adenines, the G-nucleotide blocks form a G-quadruplex structure. Since,  $(G_4A_4)_4$  RNA has four G-repeat blocks separated by non-G nucleotides, it forms a G-quadruplex (also called a G4) structure. On the other hand, although,  $(GA)_{20}$  RNA has the same number of G-nucleotides as  $(G_4A_4)_4$  RNA, however, since it lacks the G-repeat blocks, it does not form a G-quadruplex structure.



I hypothesized that if the better binding of dMi-2 to Poly (G) RNA was entirely due to the G-quadruplex structure of that RNA, then I should observe a similarly higher affinity for  $(G_4A_4)_4$  RNA compared to  $(GA)_{20}$  RNA. To test this hypothesis, I first verified whether the  $(G_4A_4)_4$  RNA indeed forms G-quadruplex structures or not. I used CD spectroscopy to test the formation of the structures. The  $(G_4A_4)_4$  RNA absorbance patterns in CD spectroscopy determined a negative peak between 237-241 nm and a positive peak between 260-265 nm, a pattern that is characteristic of G-quadruplex structure (**Figure 3.1.7D**). I measured the absorbance at 20°C, 30°C and 90°C to account for the stability of the structures. At 90°C, the structure was disrupted, suggesting that the G-quadruplex structure formed by  $(G_4A_4)_4$  RNA is less stable at 90°C compared to the G-quadruplex structure of Poly (G) RNA (data not shown). Since, I perform the binding experiments at room temperature i.e. 26°C, I therefore, can be sure of the formation of G-quadruplex structures as it remained stable between 20°C (data not shown) and at 30°C (**Figure 3.1.7D**). On the other hand, the control  $(GA)_{20}$  RNA had its own distinct absorbance pattern, confirming that it did not form a G-quadruplex structure (**Figure 3.1.7C**).

Next, I tested these RNAs in EMSAs for binding to dMi-2. dMi-2 bound to both the RNAs (**Figure 3.1.7A-B and Figure 3.1.7E, quantification**). The binding affinity of dMi-2 to G-quadruplex forming  $(G_4A_4)_4$  RNA was  $0.08 \pm 0.00 \mu\text{M}$  while for control  $(GA)_{20}$  RNA, which does not form a G-quadruplex structure was  $0.07 \pm 0.01 \mu\text{M}$  (**Figure 3.1.7E**). Since, the binding affinities of G-quadruplex forming  $(G_4A_4)_4$  RNA and the non-G-quadruplex forming  $(GA)_{20}$  RNA are very similar, it suggests that the G-quadruplex structure does not confer an added advantage for binding to dMi-2. Therefore, the higher affinity of binding that was observed for Poly (G) RNA is unlikely due to its G-quadruplex structure and might be due to the G-nucleotides itself. From this I conclude that dMi-2 prefers binding to G-nucleotide rich RNAs *in vitro*.



**Figure 3.1.7: dMi-2 binds to G-rich RNAs**

EMSAs showing the binding of dMi-2 to G-quadruplex forming  $(G_4A_4)_4$  RNA and to non-G-quadruplex forming  $(GA)_{20}$  RNA (**A - B**). 250 nM RNA was used for each reaction and dMi-2 was titrated from 10 nM (lanes 2) to 640 nM (lanes 8). Each RNA only controls was loaded in lane 1. Each EMSA is a representative figure of at least two independent replicates. The reactions were loaded on native polyacrylamide gels and electrophoresed. The gels were stained with 1X SYBR gold and imaged with Bio-Rad ChemiDoc touch system. Circular

dichroism (CD) spectroscopy of (GA)<sub>20</sub> and (G<sub>4</sub>A<sub>4</sub>)<sub>4</sub> RNAs (**C** and **D**, respectively). (G<sub>4</sub>A<sub>4</sub>)<sub>4</sub> RNA forms G-quadruplex structures with absorbance spectra for negative peak between 237-241 nM and positive peak between 260-265 nM. CD spectroscopy was performed at three different temperatures, 20°C (not shown), 30°C (Blue) and 90°C (not shown). The (GA)<sub>20</sub> RNA does not form a G-quadruplex structure. The data were analysed by CAPITO CD analysis and plotting software. The data were smoothed with Savitzky-Golay filter and CD values were converted to mean residue ellipticity ( $\theta$  in grad cm<sup>2</sup> dmol<sup>-1</sup>). The red-line denotes the zero (0) value on Y-axis. (**E**) Quantification of the bands in gels in (A), and (B). The bands were quantified using Bio-Rad Image Lab software. The data points of two independent biological replicates are plotted.

**Table 2: RNA sequences**

RNA	Sequence
Poly (A)	(A) <sub>40</sub>
Poly (U)	(U) <sub>40</sub>
Poly (C)	(C) <sub>40</sub>
Poly ((G)	(G) <sub>40</sub>
(G <sub>4</sub> A <sub>4</sub> ) <sub>4</sub> RNA	A <sub>6</sub> (G <sub>4</sub> A <sub>4</sub> ) <sub>4</sub> A <sub>2</sub>
GA repeat RNA	(GA) <sub>20</sub>

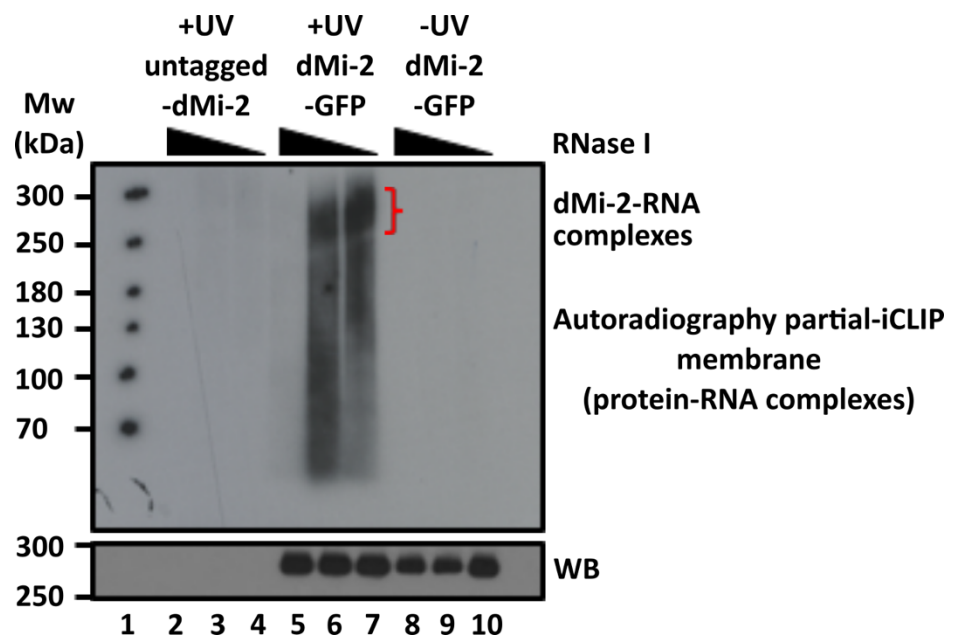
## 3.2 Chapter 2

### 3.2.1 dMi-2 Binds RNA *in vivo*

My results in chapter 1 have established that dMi-2 has an inherent property to bind RNA *in vitro*. I hypothesized that dMi-2 might bind RNA *in vivo* as well. To answer this, I performed a partial individual-nucleotide resolution UV cross-linking and immunoprecipitation (iCLIP) experiment. The technique makes use of 254 nm (nanometre) Ultraviolet (UV) light ( $300\text{mJ}/\text{cm}^2$ ) to cross-link RNAs to protein in native physiological conditions. The cross-linked protein-RNA complexes are immuno-precipitated using an antibody against the protein of interest. The RNAs are trimmed in a controlled fashion using RNase I and any DNA contaminations are removed by DNase treatment. The 5'- end of the RNA is then phosphorylated with a radio-labelled phosphate, for example,  $^{32}\text{P}$ , using a kinase. The radio-labelled RNA-protein complexes are resolved by electrophoresis in a gel and transferred to a nitrocellulose membrane. The membrane can be autoradiographed for the detection of RNA signals. If a protein binds RNA *in vivo* then there should be signal on the autoradiogram at the position at which the RNA-protein complex of interest runs.

I used two *Drosophila melanogaster* Schneider-2 (S2) cell lines, one of which has the dMi-2 protein endogenously tagged with green fluorescence protein (GFP) on its C-terminal end (dMi-2-GFP) and a control cell line that has no GFP tag, (untagged-dMi-2) (Lenz et al., 2020 (in press)). As an additional control, I also included dMi-2-GFP cells without any UV treatment, that is, no cross-linking for protein-RNA complexes (minus UV dMi-2-GFP). I performed GFP immunoprecipitation, using GFP-trap beads to immunoprecipitate the GFP fused dMi-2. The partial-iCLIP was performed as mentioned above and samples were electrophoresed on a 4-15% gradient polyacrylamide gel. The gel resolved protein-RNA complexes were transferred to a nitrocellulose membrane and the membrane was autoradiographed (**Figure 3.2.1, upper panel**). As expected, I observed no signal in the untagged-dMi-2 sample which lacks a GFP-fusion protein (**Figure 3.2.1, upper panel, lanes 2-4**). dMi-2 has a molecular weight of nearly 224 kilo Daltons (kDa) and the molecular weight of GFP is 27 kDa. Therefore, we expected dMi-2-GFP fusion protein to run above 250 kDa. We detected signals in UV treated dMi-2-GFP cells at the expected molecular weight suggesting cross-linking of dMi-2 to RNA *in vivo* (**Figure 3.2.1, upper panel, lanes 5 - 7**). No signal was detected in the lane with highest concentration of RNase I (lane 5) suggesting complete digestion of RNAs. However, as the

RNase I concentrations decreased from  $10^{-3}$  Units/ $\mu$ l in lane 5 to  $10^{-5}$  Units/ $\mu$ l in lane 7, there appeared increased signals at the expected size (red closing bracket). This suggests that the signals that were detected are indeed RNA and not any contaminating DNA that survived the DNase treatment. There also appeared signals throughout in lanes 6 and 7. Those signals are very likely either from the RNA-binding proteins that co-purified with dMi-2-GFP or are the degradation products of dMi-2-GFP bound to RNA. No signal was detected in the (minus) -UV dMi-2-GFP samples (**Figure 3.2.1, upper panel, lanes 8-10**). To establish that the signal was indeed coming from RNA cross-linked with dMi-2-GFP, I probed the nitrocellulose membrane with anti-dMi-2 antibody. As expected, no dMi-2 signals were detected in untagged-dMi-2 samples, lanes 2-4 (**Figure 3.2.1, lower panel**). On the other hand, dMi-2-GFP signals were detected in lanes 5-10 suggesting that the signals in the autoradiogram were indeed a result of cross-linked RNAs that were closely associated with dMi-2 *in vivo*.



**Figure 3.2.1: dMi-2 Binds RNA in vivo**

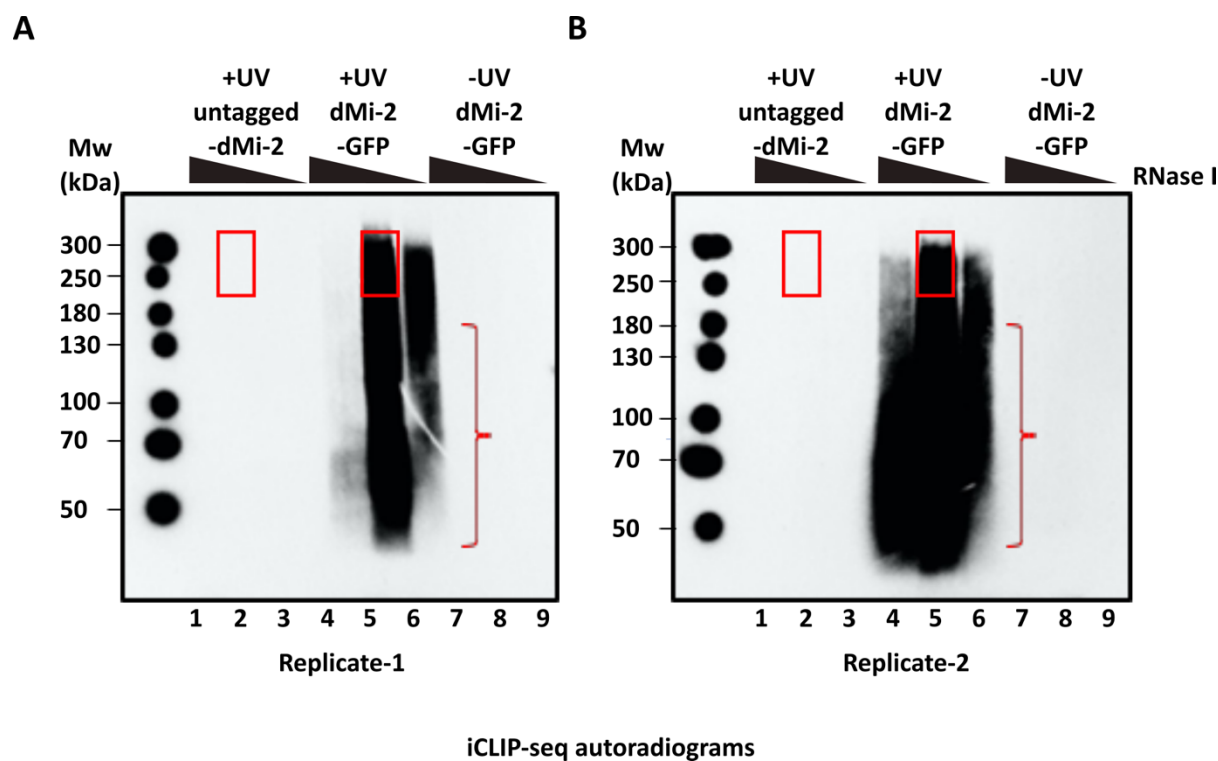
Partial-iCLIP showing that dMi-2-GFP binds RNA *in vivo*. Autoradiography of partial- iCLIP membrane (upper panel) and the Western blot (WB) against dMi-2 (lower panel). The dMi-2-GFP cells were treated with UV radiation to crosslink the RNA bound by dMi-2-GFP *in vivo*. As controls, we similarly treated untagged-dMi-2 cells with UV and dMi-2-GFP cells were not treated with UV. The protein extracts, 3 mg each, were used for GFP-immunoprecipitation and partial-iCLIP was performed to capture the RNA - dMi-2-GFP complexes. The samples were treated with DNase, and RNA fragments were trimmed under a controlled RNase I treatment. The samples were loaded on a gradient gel and electrophoresed. The protein-

RNA complexes were transferred to a membrane and autoradiographed. Red closing bracket indicates dMi-2-RNA complexes. Protein molecular weights are indicated on the left.

### 3.2.2 dMi-2 iCLIP

Having shown that dMi-2 binds RNA *in vivo* in an unbiased partial-iCLIP experiment, I next went on to identify the RNAs that dMi-2 binds to in the Drosophila S2 cells using iCLIP sequencing. The method is a continuation of the partial iCLIP whereby, the protein-RNA complexes resolved in a gel are retrieved from the membrane. The RNAs are purified and a cDNA library is prepared which is sequenced to identify the RNAs bound by the protein.

The same cell lines and experimental setup as described above (3.2.1) was used. I performed an additional biological replicate for the iCLIP-seq. The protein-RNA contacts were cross-linked by UV treatment and subsequently protein-RNA complexes were resolved on a gel and autoradiographed as in 3.2.1 (**Figure 3.2.2 A and B**). As expected, there were radio-labelled-RNA signals from dMi-2-GFP samples in both of our replicates (**Figures 3.2.2 A and B, lanes 4-6**). As the RNase I concentrations decreased from lanes 4 to lanes 6, increased radio-labelled-RNA signals were detected, confirming that the signals were indeed from RNA. No signals were detected in untagged dMi-2 samples which do not have any GFP tag (lanes 1-3) suggesting that the signals obtained in lanes 4-6 were specifically from dMi-2-GFP bound to RNA. Similarly, no signals were detected in dMi-2-GFP samples which were not treated with UV light (lanes 7-9), further substantiating that the signals are specific to dMi-2-GFP. As was observed in 3.2.1, there were signals throughout the lanes 4-6 (right closing brackets), which could either come from RNAs cross-linked to proteins that co-purified with dMi-2 or RNAs that were bound to degradation products of dMi-2. Next, I selected the moderately RNase I treated ( $10^{-4}$  Units/ $\mu$ l) samples from dMi-2-GFP and untagged dMi-2 samples in both replicates because the RNA fragments in these samples are of the optimal size, such that the RNA fragments are not too long, which could cause hindrances during immunoprecipitation, or too small, which do not allow proper alignment of the reads after sequencing (Huppertz et al., 2014). The region where dMi-2-GFP-RNA complexes were expected to run (**Figure 3.2.1A and B, red boxes**) and their corresponding regions in the untagged dMi-2 controls were selected and cut out from the membrane.

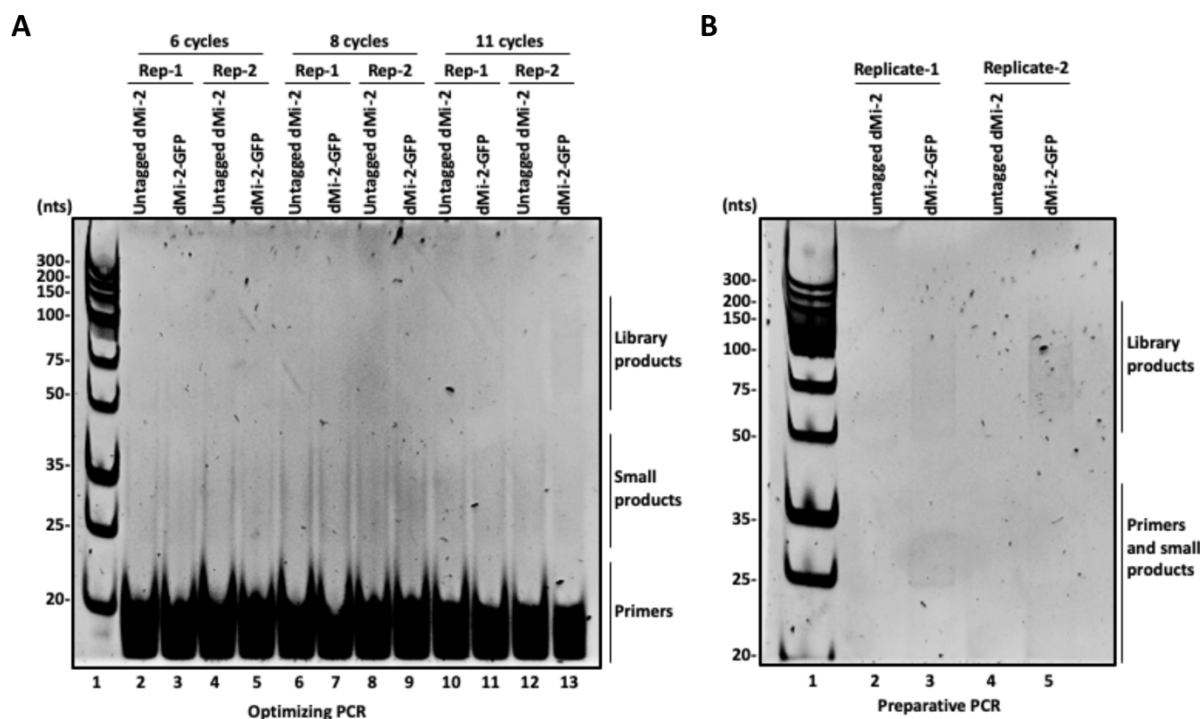


**Figure 3.2.2: dMi-2 iCLIP-seq membranes**

iCLIP experiments of dMi-2-GFP replicate-1 (**A**) and replicate-2 (**B**). 26.10 mg of whole cell extract from each cell type or treatment were used for replicate-1 and 28.10 mg for replicate-2. The dMi-2-GFP cells were treated with UV light to cross-link protein-RNA contacts. Untagged dMi-2 cells treated with UV and untreated dMi-2-GFP cells were used as controls. After GFP-immunoprecipitation and downstream steps, radio-labelled RNA-protein complexes were loaded on gradient polyacrylamide gels and electrophoresed. The complexes were transferred to membranes by blotting. The membranes were autoradiographed. Radio-labelled RNA signals were detected in dMi-2-GFP samples at the expected molecular weight of dMi-2-GFP (Figure 3.2.1A and B, lanes 4-6). No signals were detected at corresponding molecular weight in both the controls (lanes 1-3 and lanes 7-9). Signals below the dMi-2-RNA signals could be either degradation products of dMi-2-GFP or co-purifying protein that also bound to RNA (Right closing brackets). The red boxes indicate the region of the membranes that were cut out for library preparation. Molecular weight is indicated on the left by dots drawn on a plastic wrap using a radioactive pen while the membrane was placed inside the plastic wrap.

The regions cut out from the nitrocellulose membranes were treated with Proteinase-K to digest the proteins. The UV cross-linking causes the formation of a covalent bond between the RNA and the protein that were in contact with each other in physiological state. Therefore, the Proteinase-K, which exclusively hydrolyses peptide bonds, digests the protein up to the cross-link site, where it leaves a single amino acid attached to the RNA. Next the RNAs were eluted from the nitrocellulose membranes and precipitated using phenol-chloroform-isoamylalcohol. The RNAs hence retrieved were reverse transcribed to cDNA.

Next, we optimized library preparation with 6, 8 and 11 cycles of PCR amplification to generate a library that yields sufficient reads yet does not contain too many PCR duplications. The best results were obtained with 11 cycles (**Figure 3.2.3A, lanes 10-13**). Therefore, we performed 11 cycles in the library preparative PCR and loaded 4  $\mu$ L (microliter) of each library on a native polyacrylamide gel and electrophoresed to verify the quality of the iCLIP-seq library (**Figure 3.2.3B**). As expected, no products were detected in both the replicates of untagged dMi-2 samples (**Figure 3.2.3B, lanes 2 and 4**). PCR products suitable for library preparation (50-100 bp) were detected in dMi-2-GFP samples in both replicates (**Figure 3.2.3B, lanes 3 and 5**). Primers and smaller PCR products were also detected. These unwanted products were removed in the subsequent size selection step. The final libraries were analysed for quality with a bioanalyzer and sequenced.





**Figure 3.2.3: iCLIP library preparation**

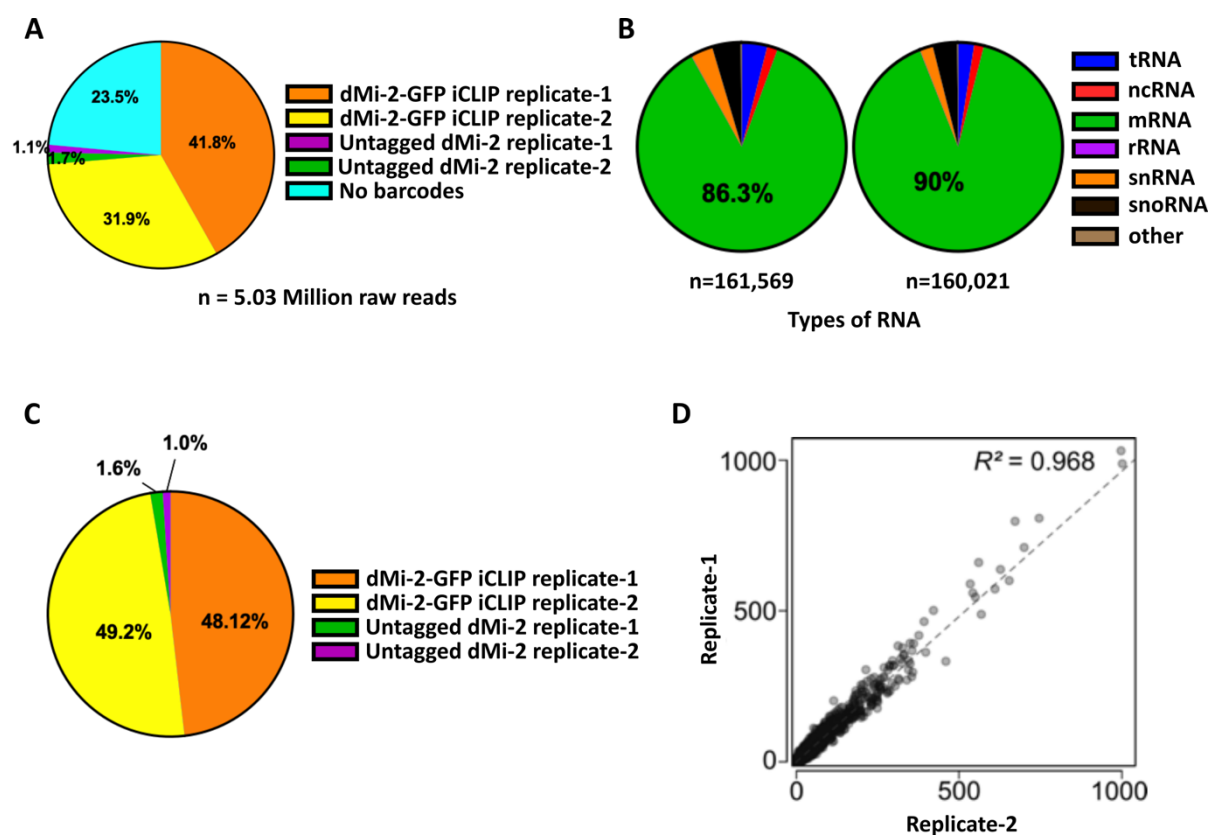
Optimization of library preparation PCR (**A**). 10  $\mu\text{L}$  of PCR reactions were set up with each sample and both the replicates; 6, 8 or 11 cycles of PCR were performed and the entire reactions loaded on a 7.5% Native polyacrylamide gel along with a size marker and electrophoresed. The gel was stained with Ethidium bromide solution. The free primers and small PCR products are denoted. Amplimers best suited for library preparation (50-100 bp), denoted as library products were most prominent in PCR reactions with 11 cycles, lanes 11 and 13. (**B**) Preparative PCR of replicate-1 and replicate-2 with 11 PCR cycles using P5 and P3-Solexa primers. 4  $\mu\text{L}$  out of each PCR reaction of a total of 40  $\mu\text{L}$  was loaded on a native polyacrylamide gel along with a size marker and electrophoresed. The gel was stained with ethidium bromide solution. The primers and small PCR products indicated on the right, were removed in the subsequent size selection step. The library products were observed in dMi-2-GFP samples, each replicate, lanes 3 and 5, as indicated on the right. No products were detected in untagged dMi-2 control samples in either replicate, lanes 2 and 4.

**3.2.3 dMi-2-GFP promiscuously binds to different types of RNAs *in vivo***

Next, the iCLIP data was analysed after sequencing. Combining the sequencing reads from the two replicates of dMi-2-GFP and untagged dMi-2 controls, we had a total of just over five million reads (**Figure 3.2.4A**). Of these, 41.8% came from replicate-1 of dMi-2-GFP and 31.9% came from replicate-2. Further, 23.5% of the reads had no barcodes and hence their origin could not be determined. These reads were discarded from the downstream analysis. Our control samples for iCLIP, that is, untagged dMi-2 replicates-1 and -2, expectedly yielded very few reads, 1.1% and 1.7%, respectively. These were also discarded from any downstream analysis (**Figure 3.2.4A**). Next, we characterized the iCLIP sequencing reads into the types of RNAs they were emanating from. dMi-2-GFP replicates-1 and -2 produced a total of 161,569 and 160,021 reads mapping uniquely to RNAs, respectively. Of these, a majority of the reads stem from mRNAs - 86.3% and 90% in replicate-1 & -2, respectively, suggesting that dMi-2 is predominantly associated with protein coding RNAs (**Figure 3.2.4B**). The remaining reads mapped to tRNAs, rRNAs, snRNAs, snoRNAs and other non-coding RNAs, indicating that dMi-2-GFP might also be associated with these RNAs to some extent *in vivo*.

Of the 279,665 total mRNA reads, 49.2% were produced by replicate-1 of dMi-2-GFP while 48.1% of the reads originated from replicate-2 (**Figure 3.2.4C**). The control samples, that is, untagged dMi-2 replicate-1 and -2, as expected produced only a comparatively small number of reads representing just 1.6% and 1.0% of the total mRNA reads (**Figure 3.2.4C**). Next, we went on to measure the strength of the relation between the two replicates of dMi-2-GFP by linear regression, comparing the mRNA read data sets of the two replicates. The closer the squared Pearson correlation coefficient ( $R^2$ ) is to 1, the greater is the correlation between two data sets. The mRNA data sets of dMi-2-GFP replicates-1 and -2 produced a squared Pearson correlation coefficient ( $R^2$ ) of 0.968, indicating that the two replicates are highly correlated (**Figure 3.2.4D**).

Taken together, these data indicate that dMi-2-GFP promiscuously binds to several types of RNAs *in vivo*, however, it predominantly binds to mRNAs. Further, the sequencing data sets from our two biological replicates show a very similar relative representation of different types of RNA. Furthermore, both data sets are highly correlated which allowed us to pool the two data sets for further analysis.



**Figure 3.2.4: dMi-2 binds to different types of RNAs *in vivo***

iCLIP sequencing data analysis **(A)**. A total of 5.03 million raw reads from dMi-2 GFP replicates-1 and -2, indicated in orange and yellow, accounting for 41.8% and 31.9% reads, respectively. The control untagged dMi-2 replicates-1 and -2, indicated in green and violet, accounted for 1.7% and 1.1% reads, respectively. 23.5% of the reads, indicated in cyan, had no barcode and were thus not included in our downstream analysis. **(B)** The distribution of different types of RNAs represented in replicates-1 and -2 of dMi-2-GFP were plotted as pie charts, the total number of reads in each replicate are indicated below the pie charts. Most of the reads in both the replicates, 86.3% and 90%, indicated in green, were mRNAs reads. **(C)** A total of 279,665 reads were mRNA reads. Of these, 48.1% originated from replicate-1 of dMi-2-GFP, indicated in orange and 49.2% from replicate-2, indicated in yellow. The control untagged dMi-2 replicates-1 and -2, indicated in green and violet, represent 1.6% and 1.0% mRNA reads, respectively. **(D)** Squared Pearson correlation coefficient ( $R^2$ ) of the two-mRNA data sets from replicate-1, on x-axis, and replicate-2, on y-axis. The  $R^2 = 0.968$  indicates high correlation between the two replicates.

**3.2.4 dMi-2 GFP binds 3'-ends of mature mRNAs**

iCLIP allows the precise determination of the cross-linked RNA nucleotide. Therefore, it is possible to analyse the distribution of dMi-2-GFP binding sites along the RNA molecules. To this end, we mapped the dMi-2-GFP cross-linking sites across a length normalized metagene. The distribution of the cross-linking sites showed a very specific pattern with few to no mapping towards the transcription start site (TSS) while most of the cross-linking sites accumulated towards the transcription end site (TES) (**Figure 3.2.5A**). This indicates that dMi-2 binds RNAs at the 3'-ends *in vivo*. To further confirm this, we aligned all annotated genes with the transcription end site (TES) at position zero (0) in their genomic context and visualized the cross-linking sites with varying intensities of blue (light blue: 1 supporting read; black: more than 4 supporting reads) for 200 nucleotides up- and down-stream of the TES. The majority of all cross-linking sites indeed mapped towards the 3'-end of the transcripts, confirming that dMi-2-GFP predominantly binds the 3'- end regions of RNAs (**Supplementary figure S3.2.5A, lower panel**). We calculated the cumulative density

of the cross-linking sites around the TES and as expected from the alignment plot, we found a sharp peak towards the 3'-end of the RNAs (**Supplementary figure S3.2.5A, density plot**), indicating binding towards the end of RNAs.

We found mRNAs from a total of 5,471 genes with at least 5 iCLIP cross-linking reads per mRNA across the two replicates of our iCLIP data. We asked whether dMi-2-GFP was bound to the genes of any of these RNAs that we found in our iCLIP data. To answer this, we used the dMi-2-GFP chromatin occupancy (ChIP-seq) data from Lenz et al., 2021 (in press). Our analysis showed that out of 5471 genes whose RNAs were bound by dMi-2-GFP, 2406 genes had significantly enriched dMi-2-GFP bound to their genes as well (**Figure 3.2.5B**). Next, we asked if RNA binding of dMi-2 plays a role in its recruitment to the chromatin. On chromatin, dMi-2 largely binds at the promoter regions, that is, towards the 5'-end of genes (Lenz et al., 2021 (in press)). If RNA binding plays a substantial role in recruiting dMi-2 to chromatin then we would expect the ChIP-seq profile of at least some of the above mentioned 2406 genes, to match their iCLIP profile. Using the above mentioned dMi-2-GFP ChIP-seq data set, we plotted the dMi-2-GFP chromatin binding sites of the selected 2406 genes which also had dMi-2-GFP bound to their mRNAs, across a metagene. Interestingly, we observed a sharp ChIP-seq peak at the 5'-end of the metagene (**Figure 3.2.5C**). This indicates that even on the genes whose RNA are bound by dMi-2-GFP, the dMi-2-GFP is located at the promoter region which is in contrast to its occupancy on RNA. This does not support the hypothesis that RNA binding is involved in recruiting dMi-2 to chromatin.

Some RNA binding proteins bind to nascent RNAs, for example, PRC2. Others bind to processed and mature mRNAs, for example, cap-binding eukaryotic initiation factor 4F (eIF4F) (Pelletier et al., 2015). Some RNA binding proteins can bind to both nascent as well as to processed RNAs. We sought to determine whether dMi-2 binds to nascent RNAs or to processed RNAs. Since, dMi-2 predominantly cross-linked near the 3'-end of the RNAs, we therefore determined the average read coverage along the last intron-exon junctions of RNAs that had a minimum of two exons. We considered 15 nucleotides from either side of the intron-exon junction and plotted the

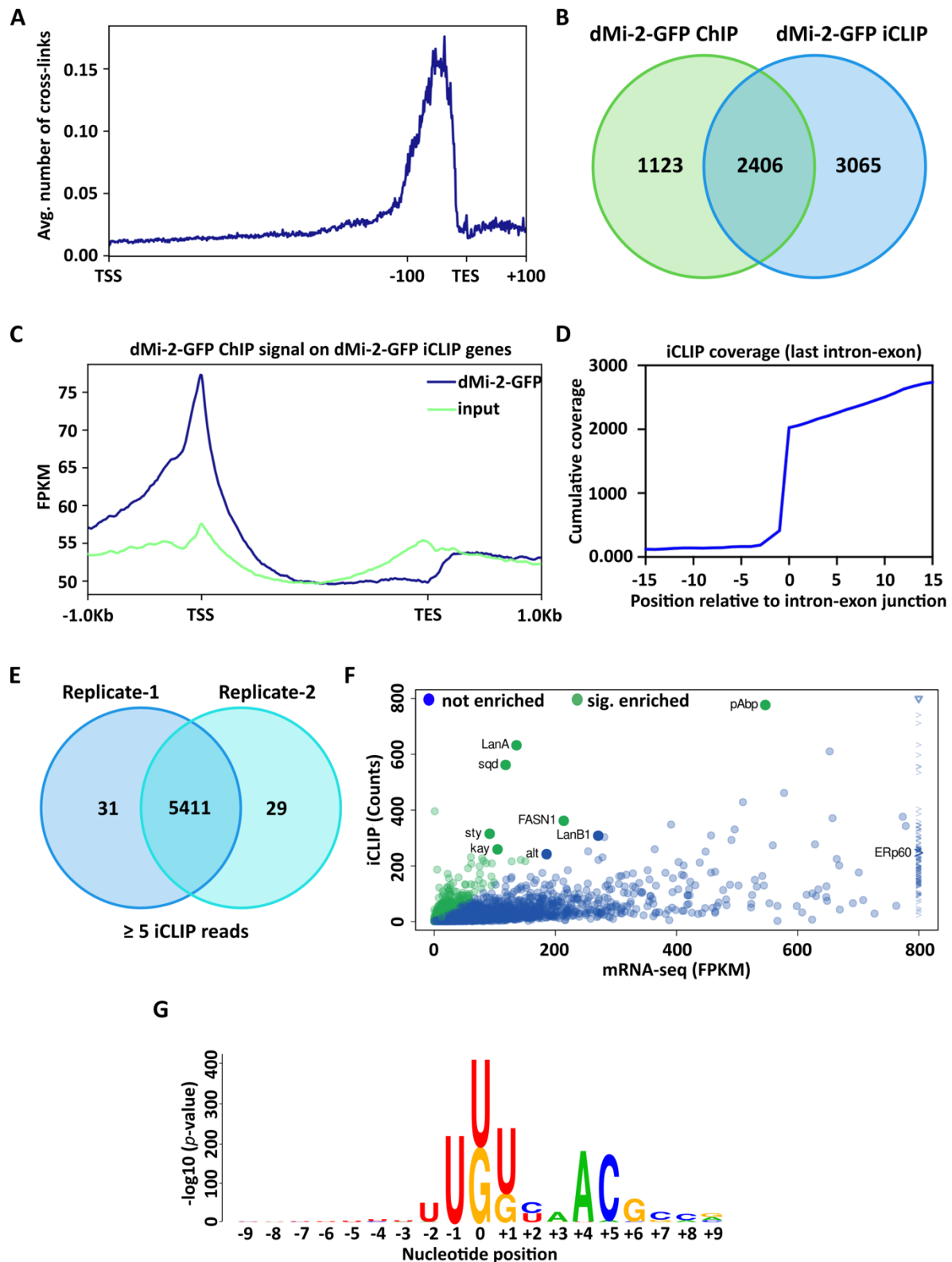
cumulative read coverage. We found low coverage across the intronic region, compared to the coverage in the exonic region (**Figure 3.2.5D**). While this finding does not rule out that dMi-2-GFP binds to introns and exons of nascent RNA, it is more compatible with the view that most binding events involve processed (spliced) RNA molecules.

From the 5,471 genes which had a minimum of 5 iCLIP cross-linking reads in each of our replicate, we determined the high confidence genes as those that were present in both the replicates (**Figure 3.2.5E**). A majority of the genes, 5,411, were common to both of our iCLIP data sets. We used these high confidence genes for our further analysis. Next, we sought to identify the significantly enriched mRNAs from the dMi-2-GFP iCLIP data sets. dMi-2 bound to RNAs promiscuously both *in vitro* as well as *in vivo*. Therefore, RNA concentration might be a major determinant for dMi-2-GFP binding. To account for that, we compared the iCLIP sequencing read counts for each identified RNA with the corresponding RNA sequencing read counts previously determined in S2 cells (Lenz et al., 2021 (in press)). We used the DEseq2 analysis platform with a fold change of greater than 4 (fold change > 4) and p-value less than 0.05 ( $p\text{-value} \leq 0.05$ ) to determine the ‘significantly enriched’ RNAs. Of the 5,411 high confidence genes in our iCLIP, 580 genes were detected as significantly enriched by DEseq2 analysis (**Figure 3.2.5F, green dots**). The genes that did not meet this criterion, that is, did not have 4-fold more iCLIP cross-linking counts over their RNA-seq read counts were distinguished as ‘not enriched’ (**Figure 3.2.5F, blue dots**). We selected a few genes from both - enriched as well as not enriched sets (**Figure 3.2.5F, indicated by labelling**), for validation through another approach (data not shown). Taken together, DEseq2 analysis revealed that indeed the binding of dMi-2-GFP to a large portion of RNAs is driven by their abundance, however, there are some RNAs (580) which were bound by dMi-2-GFP preferentially.

iCLIP favours cross-linking to Uridine nucleotides due to the usage of UV-C for cross-linking (Sugimoto et al., 2012). During the reverse transcription step of the iCLIP procedure, the reverse transcriptase enzyme does not readthrough the cross-linking site and dissociates due to the obstruction caused by a covalent bond created by the

UV cross-linking treatment between an amino acid and a nucleotide. Therefore, each iCLIP read lacks a single nucleotide upstream of its start (-1). It is this nucleotide at position -1, which can be identified from alignment of iCLIP reads to the respective genome, that is the actual cross-linking nucleotide and gives the single nucleotide resolution to iCLIP. In our *in vitro* RNA binding assays, we observed a preferential binding of dMi-2 to G-rich RNAs. We, therefore, asked whether dMi-2-GFP cross-links to Guanine (G) nucleotides preferentially *in vivo* as well. To this end, we used KpLogo and aligned all the iCLIP reads relative to their cross-linking sites to the genome and detected the most frequently occurring nucleotide at the cross-linking site (position zero, 0) in our iCLIP data. We extended this analysis to nine nucleotides from either side of the cross-linking site. Expectedly, the Uridine nucleotides were detected as enriched at the cross-linking site (**Figure 3.2.5F**). However, since the cross-linking favours the U-nucleotides, we disregarded that and looked for the next significantly enriched nucleotide at the cross-linking site. Interestingly, we found G-nucleotides enriched at the cross-linking site indicating that dMi-2 prefers cross-linking to G-nucleotides which supports our *in vitro* data where we observed binding preference to G-rich RNA. Analysing the frequency of the nucleotides upstream and downstream of the cross-linking site, we observe that even though the G-nucleotides are enriched at the cross-linking site itself, the sequence surrounding the cross-linking site is not G-rich. This might be due to the fact that dMi-2 is part of a multi-protein NuRD complex or associates with dMep-1 in dMec complex. Therefore, the proteins that dMi-2 associates with might be involved in modulating the RNA binding of dMi-2 and influence the sequences of RNAs that it binds. From our KpLogo analysis we can estimate -GUNNACg- as the possible sequence that may be favoured for binding by dMi-2 *in vivo*.

Taken together, our iCLIP data shows that dMi-2-GFP binds at the 3'- end of processed mature mRNAs. There are several mRNAs that are enriched over other RNAs, in our iCLIP, indicating a possible stronger interaction and dMi-2 shows a preference for cross-linking to G-nucleotides.



**Figure 3.2.5: dMi-2 binds 3'-ends of mature mRNAs**

iCLIP cross-linking sites are predominantly located at the 3'-end of transcripts (**A**) Distribution of dMi-2-GFP cross-linking sites across a metagene indicating binding towards the 3'-end of RNAs. (**B**) Venn diagram showing the overlap between dMi-2-GFP -ChIP and -iCLIP genes

with 2406 genes having both. **(C)** Metagene analysis of 2406 genes from (B) showing the distribution of dMi-2-GFP on chromatin. dMi-2-GFP predominantly binds at the 5'-end of these genes. **(D)** iCLIP read coverage across the last intron-exon junction. All annotated genes with at least two exons are aligned at the last junction between intron and exon because the reads are almost exclusively situated towards the end of transcripts. The first base of the exon is at position zero. Even distribution across this junction would indicate bound nascent RNA while a sharp increase of coverage in the exon compared to the intron indicates fully processed RNA bound to dMi-2. **(E)** Venn diagram indicating the high confidence common genes between the two iCLIP replicates. 5,411 genes were found in both the replicates. **(F)** Some transcripts are highly enriched in iCLIP experiments compared to their regular RNA-Seq abundances. In order to rule out target transcripts bound in iCLIP are not solely over-represented due to their overwhelming global abundance, we correlated the number of cross-links per gene (iCLIP counts, y-axis) with RNA-Seq abundances (RNA-seq FPKM, x-axis) in the same cell line. We used DESeq2 to determine robustly and significantly enriched mRNAs (green) across the two biological iCLIP replicates and three RNA-Seq replicates (fold change > 4 and  $p < 0.05$ ) in contrast to mostly proportionally abundant mRNAs in both experiments (blue). We chose sufficiently conservative thresholds for enrichment to account for the poor overall correlation of the two different sequencing approaches. We selected some genes with decent expression and various enrichment for validation in further experiments (highlighted with labels). **(G)** KpLogo sequence plot around the cross-linking site indicating enrichment of G-nucleotides at the cross-linking site and a possible sequence preference to 5'- GUNNACg-3' sequence. The size of nucleotide letters represents the enrichment as ( $p$ -value) above expected frequency sampled from random mRNA sequences.



### 3.3 Chapter 3

#### 3.3.1 RNA binding inhibits dMi-2 remodelling activity

In chapters 1 and 2, I established and characterized the *in vitro* and *in vivo* RNA binding properties of dMi-2. Several chromatin associated proteins (CAPs) possess RNA binding properties and interaction with RNA influences both the recruitment to the chromatin as well as expulsion from the chromatin (Beltran et al., 2016; Zhao et al., 2008). *Drosophila* MOF (males absent on the first), which is a histone acetyltransferase (HAT), is recruited to the X-chromosome in *Drosophila* males through its interaction with roX2 RNA (Akhtar et al. 2000). On the other hand, PRC2 (Polycomb repressive complex 2), a histone methyltransferase complex, is evicted from the chromatin via its interaction with RNA (Beltran et. al., 2016 and Wang et al., 2017).

I asked what function does the interaction of dMi-2 with RNA entail. To answer this, I tested the effects of RNA on the remodelling function of dMi-2 *in vitro*. I used end-positioned mononucleosomes with 80 base pairs of flanking DNA. dMi-2 binds nucleosomes and in presence of ATP, slides such an end-positioned nucleosome to a more central-position. The centrally positioned nucleosome migrates slower than the end-positioned nucleosome in the gel. However, since dMi-2 binds the nucleosome to remodel it, and the binding of dMi-2 retards the mobility of the nucleosomes to the extent that obscures the differences in the mobility of differently positioned nucleosomes. Therefore, at the end of the reaction, dMi-2 is competed off the nucleosome by competitor plasmid DNA. I tested the effects of three different RNAs in increasing amounts – (G<sub>4</sub>A<sub>4</sub>)<sub>4</sub> RNA, a G-quadruplex forming RNA (**Figure 3.3.1A**), (GA)<sub>20</sub> RNA (**Figure 3.3.1B**) and Hsp70Aa #6 RNA (**Figure 3.3.1C**), each RNA 40 nucleotide in length and each shown to bind dMi-2 in Sections 3.1.4 and 3.1.7. As expected, I observed remodelling of the nucleosome from end-position to central-position (**Figure 3.3.1A, B and C, lanes 1-2 and supplementary figure S3.3.1A, B and C, lanes 1- 2 in each, red arrow**). Interestingly, with increasing RNA concentration, I observed a gradual decrease in the remodelling efficiency of dMi-2 (**Figures 3.3.1A and B, lanes 3-10 and supplementary figure S3.3.1A and B, lanes 3-10**) to the point where no centrally-positioned nucleosome was detectable on the gel (**lanes 7-10, red asterisk marks**), indicating complete inhibition. Interestingly, I observed no drastic effects of Hsp70Aa #6 RNA on the remodelling function of dMi-2 (**Figure 3.3.1C and supplementary figure S15C, lanes 3-10**), even at the highest

concentration of RNA used (**lane 10, blue asterisk mark**), indicating no apparent inhibition of remodelling activity of dMi-2.

Next, I quantified the fraction of the centrally-positioned band against the end-positioned band in the gels in (**Figure 3.3.1A, B and C**) to calculate the apparent half-maximal concentration of each RNA at which 50% of the remodelling reaction is inhibited ( $IC_{50}$ ) (**Figure 3.3.1D**). The  $IC_{50}$  for  $(G_4A_4)_4$  RNA was  $140.1 \pm 20.93$  (nM) while that for  $(GA)_{20}$  RNA was  $152.65 \pm 52.25$  (nM) (**Figure 3.3.1D**). As was evident from the gel image, there was no apparent inhibition by Hsp70Aa #6 RNA and hence the  $IC_{50}$  could not be calculated (**Figure 3.3.1D**).

Inhibition of chromatin modifying enzymes by RNA could be due to the RNA interacting with the active site of the protein, thereby, blocking its enzymatic activity. This, for example, is the case for PRC2 (Zhang et al., 2019). Alternatively, RNA might compete off the protein from the nucleosomes. From my *in vitro* data in Chapter 1, I learned that the active site of dMi-2, that is, the ATPase region, does not bind to RNA on its own. This finding argues against inhibition of remodelling by RNA due to binding to the active site of the enzyme. Therefore, the other possible explanation is that the RNA competes dMi-2 off of the nucleosomes, thereby, causing inhibition in remodelling.

Taken together, this data supports that binding to some RNAs can inhibit the remodelling function of dMi-2 and this inhibition is possibly due to a competition for binding to dMi-2 between nucleosomes and RNA in which RNA binding evicts dMi-2 from the nucleosomes.



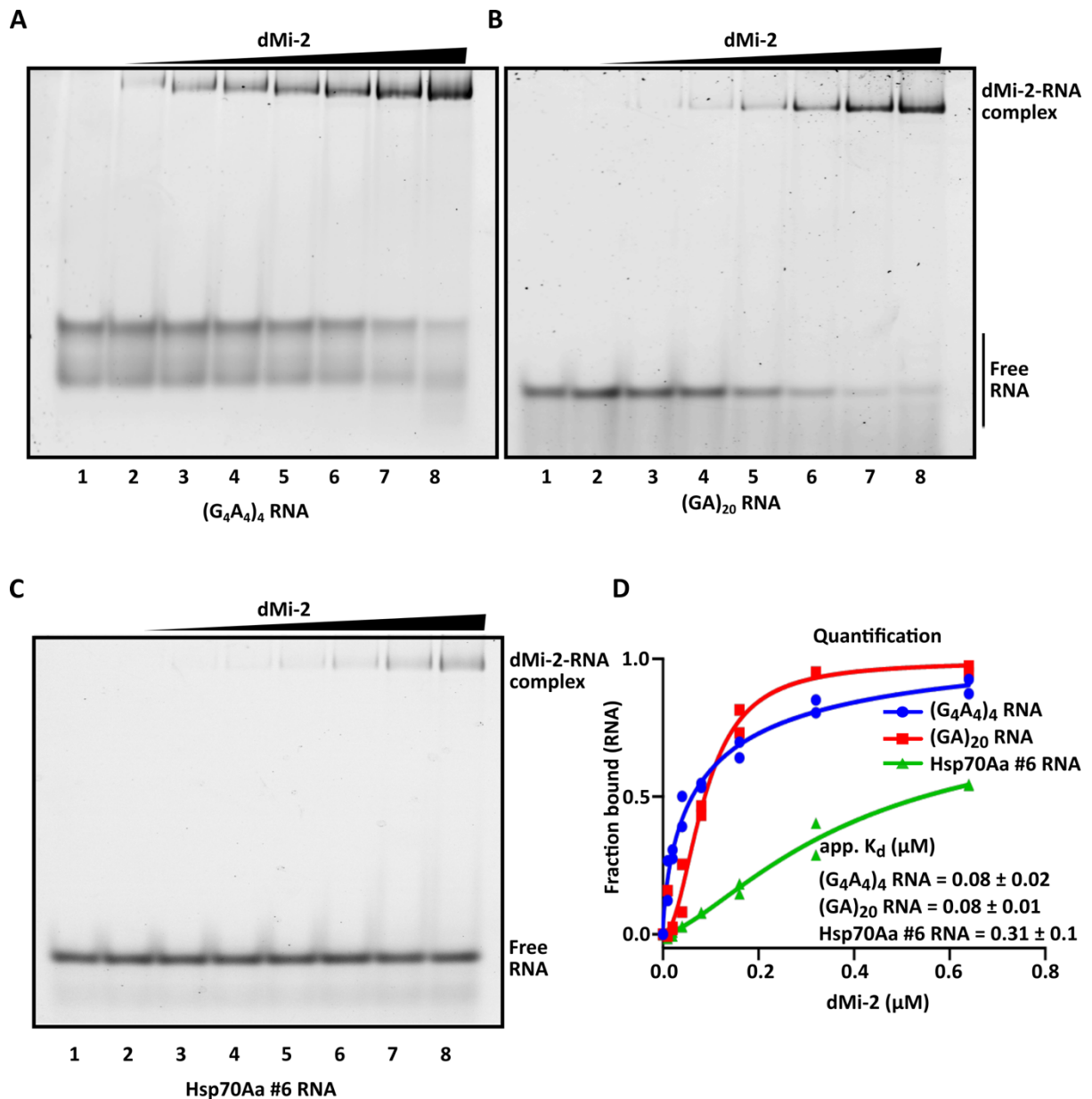
flanking DNA were used in each experiment as denoted by plus sign, except in controls (minus signs on the top). End-positioned nucleosome, lane 1 in each gel image, represented on the left (orange denotes the histone octamer positioned at the end of the nucleosomal DNA - in blue). 50 nM dMi-2 was added in each reaction (plus sign), except in controls (minus sign on the top). Remodelled nucleosome (represented on the left, orange denotes histone octamer positioned centrally on the nucleosomal DNA - in blue) can be observed in lanes 2-6 of (A) and (B) and lanes 2-10 in (C). Free RNA in each gel can be observed at the bottom. Reactions were loaded on native polyacrylamide gels, electrophoresed, stained with SYBR gold and imaged with Bio-Rad ChemiDoc touch system. Red arrows point out the remodelled nucleosome band and red asterisk marks in lanes 7-10 of (A) and (B) point out the inhibition of remodelling. Blue asterisk in lane 10 of (C) points out remodelled nucleosome even at the highest concentration of Hsp70Aa #6 RNA. Representative images of at least three independent experiments. **(D)** Quantification of gel images in (A), (B) and (C) using Bio-Rad Image-Lab software and calculation of  $IC_{50}$  using linear regression (Blue:  $(G_4A_4)_4$  RNA, Red:  $(GA)_{20}$  RNA, and Green: Hsp70Aa #6 RNA). Individual data points of two independent biological replicates are plotted.

### 3.3.2 RNAs that bind efficiently to dMi-2 inhibit remodelling

In section 3.3.1, I hypothesized that perhaps the inhibition of remodelling is possibly by way of RNA competing dMi-2 off of the nucleosomes. If this hypothesis is the possible explanation for the remodelling inhibition of dMi-2 in presence of RNA; then the RNAs that bind to dMi-2 efficiently should be able to inhibit better compared to RNAs that bind with less affinity. Perhaps that should also explain why Hsp70Aa #6 RNA failed to inhibit the remodelling. To test this hypothesis, I performed EMSA with the three RNAs used above in section 3.3.1 (**Figure 3.3.2A, B and C, and supplementary figure S3.3.2**). dMi-2 bound to all the three RNAs albeit with evident differences in binding affinity. I quantified the RNA bound to dMi-2 and the free RNA and calculated the approximate dissociation constant (app.  $K_d$ ) to estimate the differences in the binding efficiency of the three RNAs (**Figure 3.3.2D**). The  $(G_4A_4)_4$  RNA and  $(GA)_{20}$  RNA ( $K_d = 0.08 \pm 0.02 \mu\text{M}$  and  $0.08 \pm 0.001 \mu\text{M}$ , respectively) bound with at least 3-fold higher efficiency compared with Hsp70Aa #6 RNA ( $K_d = 0.31 \pm 0.1 \mu\text{M}$ ). These results support the hypothesis that binding of RNA likely competes dMi-2 off of the nucleosomes. Therefore, the RNAs that bind with high affinity -  $(G_4A_4)_4$  RNA and  $(GA)_{20}$  RNA is able to compete dMi-2 away from the nucleosomes while the RNA that binds comparatively poorly

– Hsp70Aa #6 RNA, fails to compete dMi-2 off the nucleosomes, thereby, its failure to inhibit dMi-2 remodelling activity. To further support this mechanism of inhibition, another RNA of equal length as the three RNAs above, a (CA)<sub>20</sub> RNA which also bound poorly to dMi-2 failed to inhibit its remodelling activity (data not shown).

Taken together, this data supports the hypothesis that the RNAs that bind efficiently to dMi-2 also inhibit its remodelling function likely by competing dMi-2 off of the nucleosomes.



**Figure 3.3.2: RNAs that bind efficiently to dMi-2 inhibit remodelling**

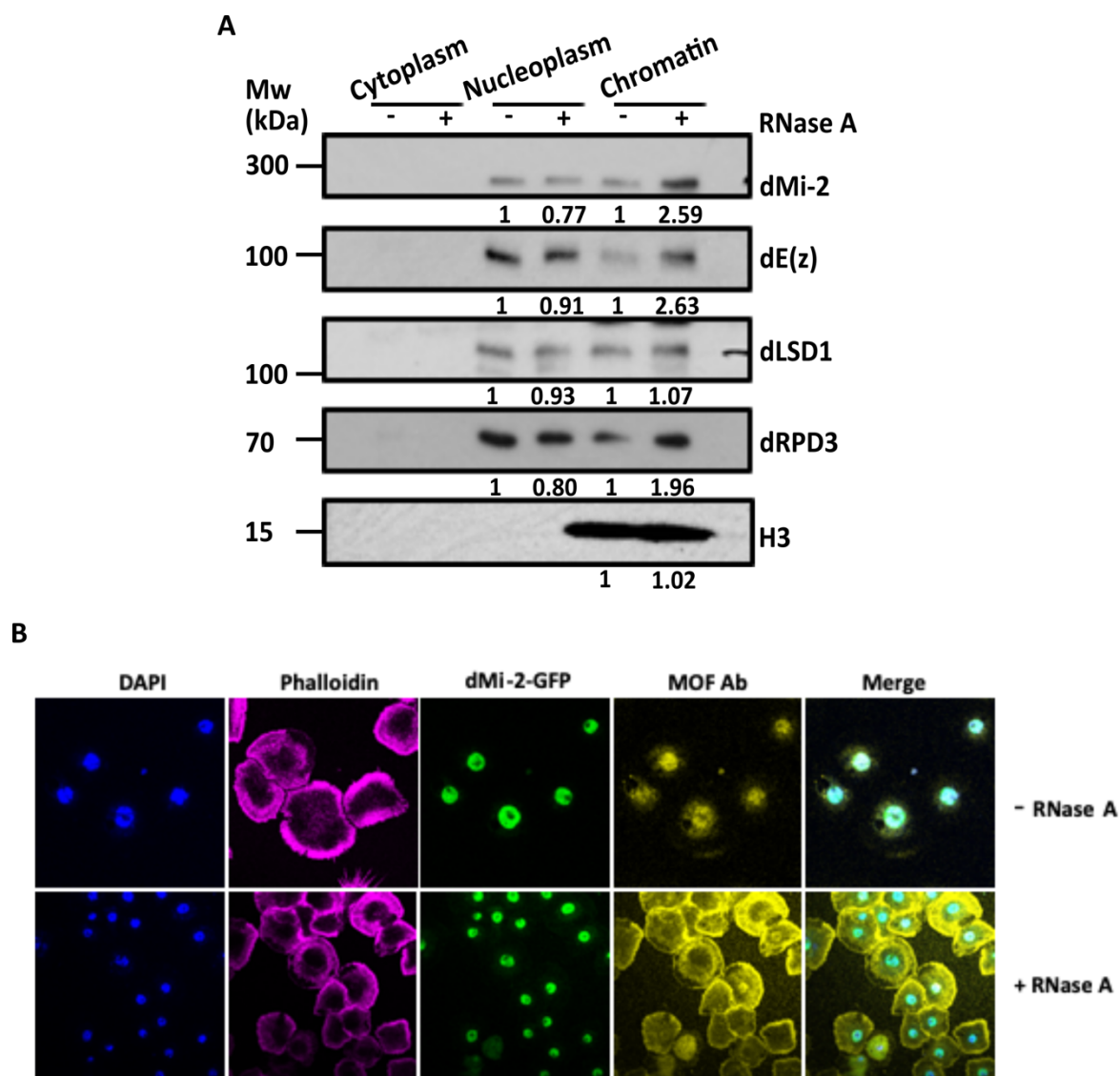
EMSA of **(A)** (G4A4)<sub>4</sub> RNA, **(B)** (GA)<sub>20</sub> RNA and **(C)** Hsp70Aa #6 RNA. 250 nM RNA was used in each reaction and dMi-2 titrated from 10 to 640 nM, lanes 2-8. Each reaction was loaded on a native polyacrylamide gel and electrophoresed. The gels were stained with SYBR gold and imaged with Bio-Rad ChemiDoc touch system. dMi-2 bound to RNA is denoted as bound complex. Free RNA is also denoted. Representative images of at least two independent experiments. **(D)** Quantification of the EMSA band intensities and calculation of (app.  $K_d$ ) using Hill slope equation (Blue: (G4A4)<sub>4</sub> RNA, Red: (GA)<sub>20</sub> repeat RNA and Green: Hsp70Aa #6 RNA) indicating poor binding of Hsp70Aa #6 compared to (G4A4)<sub>4</sub> RNA and (GA)<sub>20</sub> RNA. Data points of at least two independent experiments are plotted.

### 3.3.3 Loss of RNA causes increased association of dMi-2 with chromatin

The data in sections 3.3.1 and 3.3.2 are indicative that RNA competes off dMi-2 from the nucleosomes by binding to dMi-2. This would suggest a decreased association of dMi-2 with chromatin upon RNA binding. I hypothesized that if RNA is degraded in the nucleus, then I should observe an increased association of dMi-2 with chromatin. To test this hypothesis, I permeabilized *Drosophila* S2 cells with Tween-20 and treated them with RNase A. The control cells were similarly permeabilized but instead of RNase A treatment, the cells were treated with BSA. The cells were then fractionated into cytoplasmic, nuclear and chromatin fractions and subjected to Western blot analysis (**Figure 3.3.3A and supplementary figure S3.3.3A**). I observed no significant leaking of the nuclear proteins into the cytoplasmic fractions. Also, the chromatin fraction, as detected by H3 signal, did not spill into either the cytoplasmic or the nuclear fractions, suggesting that the fractionation worked well under these conditions. Some proteins in the cell exist in both chromatin-bound and unbound state, such as dLSD1 which was detected in both chromatin as well as nucleoplasm fractions. In a similar experiment in mammalian cells, EZH2 signal has been shown to increase in the chromatin fraction upon RNA depletion (Beltran et al. 2016). I tested the occupancy of the *Drosophila* homolog of EZH2, Enhancer of zeste (Ez), in my experiment. Indeed, my results in *Drosophila* S2 cells confirm that upon RNA depletion, Ez shows an increased association with the chromatin (Fold change = 2.63). Next, I looked at the effects of RNA depletion on dMi-2 and like Ez, dMi-2 signal decreased in the nuclear fraction (Fold change = 0.77) and increased in the chromatin fraction (Fold change = 2.59) suggesting an increased association with the chromatin upon RNase A treatment (**Figure 3.3.3A, upper panel**). RPD3, a component of the

NuRD (Nucleosome remodelling and deacetylase) complex to which dMi-2 also belongs, also showed a loss in the nuclear fraction upon RNase A treatment (fold change = 0.80) and a concomitant increase in the chromatin fraction (fold change = 1.96). This could be because RPD3 and dMi-2 interact and since dMi-2 recruits to chromatin upon RNA degradation, it could also carry RPD3 along. I detected no significant changes in the occupancy of dLSD1 and H3, suggesting that the changes I observed in the occupancy of dMi-2, Ez and RPD3 are not a general effect on all chromatin associated proteins upon RNA depletion. Taken together, this data suggests that upon depletion of RNA from the cell, dMi-2 shows an increased association with the chromatin as detected by this assay.

dMOF which is recruited to chromatin by its interaction with RNA, has been shown to leave the nucleus upon RNA depletion (Akhtar et al. 2000). On the other hand, my fractionation analysis above indicates that dMi-2 is retained within the nucleus and that more dMi-2 associates with the chromatin upon RNA depletion. To further establish that depletion of RNA results in an increased association of dMi-2 with chromatin by using an independent approach, I went on to use confocal microscopy. To this end, I used the dMi-2-GFP tagged cell lines, as used in chapter 2. As a control, I tracked the occupancy of dMOF using an antibody against dMOF. The cells were permeabilized with digitonin and treated with RNase A. The control cells were similarly permeabilized but mock treated. I used DAPI and Phalloidin conjugates to stain the nucleus and the actin filaments in the cell. The cells were fixed and imaged using a confocal microscope (**Figure 3.3.3B and supplementary figure S3.3.3B**). Indeed, dMOF was depleted from the nuclei upon RNA degradation (**Figure 3.3.3B, dMOF panel**), indicating a successful depletion of nuclear RNA. On the other hand, dMi-2 did not appear to leave the nucleus upon RNA degradation (**Figure 3.3.3B, dMi-2-GFP panel**). Whether the loss of RNA resulted in an increased association of dMi-2 with the chromatin remains uncertain in this assay due to the limited resolution afforded by the microscopy. However, the experimental setup may be applicable for future analysis with higher resolution microscopy methods such as super resolution microscopy.



**Figure 3.3.3: Loss of RNA causes increased association of dMi-2 with chromatin**

**RNase A treatment and western blot**

**(A)**  $3.5 \times 10^7$  cells were permeabilized by treatment with 0.05% of Tween-20 for 10 minutes at 4°C. Cells were treated with 10 (mg/ml) of RNase A. Control cells were similarly permeabilized but treated with equal amount of BSA. Cells were fractionated in cytoplasmic, nuclear and chromatin fraction. Each fraction was loaded on an 8-12% gradient denaturing polyacrylamide gel and electrophoresed. Western blot analysis was performed for the proteins indicated. Histone H3 was used as loading control. The western blot signals were quantified using ImageJ software. The BSA treated control signals were set to 1 and the fold change in the treated samples were calculated and numbers indicated below. **(B)**  $1 \times 10^6$  cells/mL each from dMi-2-GFP were seeded on concanavalin A coated glass coverslips,



permeabilized by treatment with 0.05 ( $\mu\text{g}/\mu\text{l}$ ) of digitonin for 10 minutes. One set was treated with RNase A and another mock treated with PBS alone and cells were fixed. For MOF immunostaining cells were blocked with BSA, incubated with anti-MOF primary antibody and Alexa-547 secondary antibody. Fixed cells were stained with DAPI and Phalloidin stains just before mounting on a glass slide using mowiol as mounting media. The GFP signal emanating from dMi-2-GFP tagged cells was used to detect dMi-2. Cells were visualized under a confocal microscope and imaged. Representative images of several independent experiments.

### **3.3.4 Binding to RNA prevents dMi-2 from interaction with chromatin**

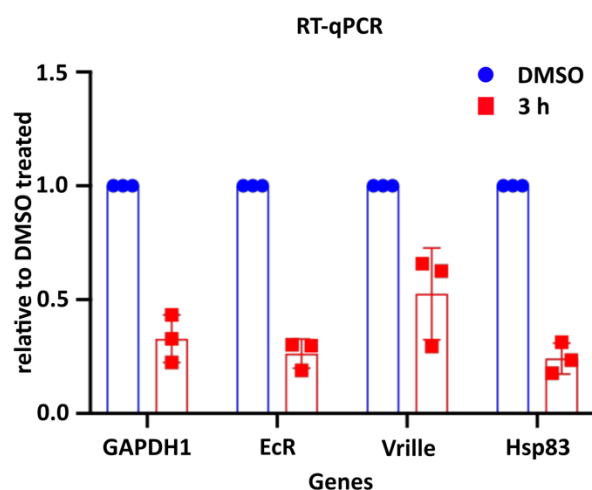
The data in section 3.3.3 indicate that interaction with RNA could prevent dMi-2 from interacting with the chromatin. To further substantiate this, I hypothesized that a global inhibition of RNA polymerase II and the resulting decrease of RNA levels in close proximity to chromatin should cause increased association of dMi-2 to the chromatin. To this end, I treated S2 cells with 10  $\mu\text{M}$  Triptolide for 3 hours. Triptolide treatment induces proteasomal degradation of RNA polymerase II (Riising et al., 2014). As controls, I treated cells with an equal volume of DMSO for 3 hours. Next, I looked at the relative transcription levels of some genes to verify the effects of Triptolide. The RNA was isolated from Triptolide treated and DMSO treated cells, reverse transcribed and amplified by qPCR (**Figure 3.3.4A**). As expected, relative to DMSO treated controls, the Triptolide treatment lead to a sharp decrease in the gene expression (**Figure 3.3.4A**) suggesting a successful downregulation of global gene expression upon RNA polymerase II degradation by Triptolide treatment.

After having established that Triptolide treatment depletes the RNA in the cell, I asked how that affects the occupancy of dMi-2 on the chromatin. To this end, I used dMi-2-GFP expressing cells and for controls I used untagged cells. As an additional control, I used a cell line expressing dLINT1-GFP from (Mačlnković et al., 2019). dLINT1 is a component of the LINT complex, a repressor complex that is also associated with chromatin (Meier et al., 2012). One set of cells from each type was treated with 10  $\mu\text{M}$  Triptolide for 3 hours and the other set was treated with DMSO for the same amount of time. To quantify the occupancy of dMi-2-GFP and dLINT1-GFP across all treatments and controls, the chromatin was prepared and I performed chromatin immunoprecipitation (ChIP) using the GFP-Trap beads. I selected a few regions in the genome at which the occupancy of dMi-2 was

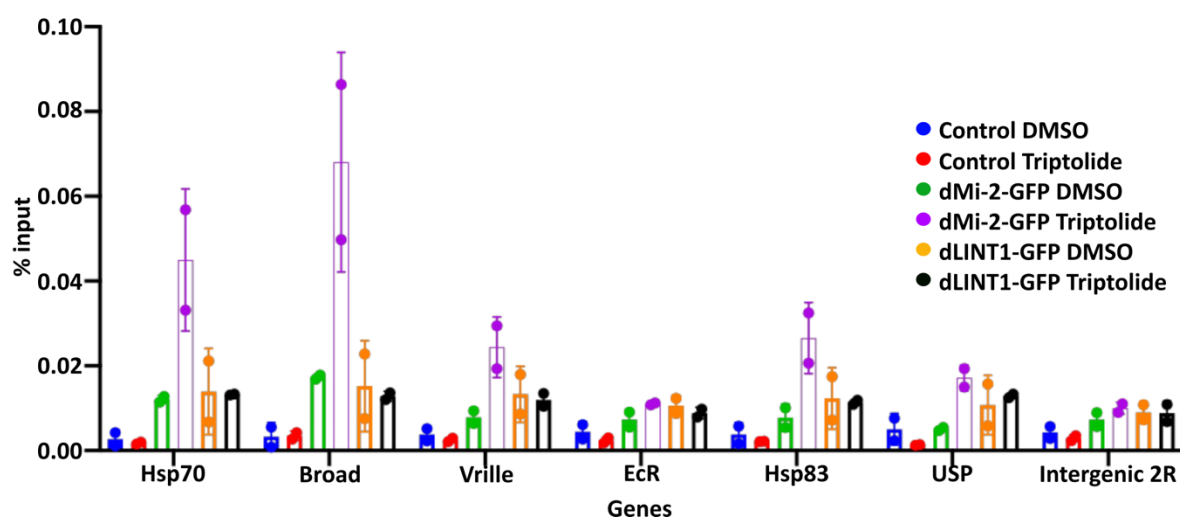
successfully quantified in ChIP assays (Kreher et al. 2017; Mathieu et al. 2012) to assess the changes in the occupancy of dMi-2-GFP and dLINT1-GFP upon inhibition of RNA Pol II transcription. As expected, since the untagged cells do not possess any GFP tag, low background signals were detected (**Figure 3.3.4B, blue and red bars**). I observed the occupancy of dLINT1-GFP at all the loci that I tested, however, I detected no significant changes in the occupancy of dLINT1-GFP upon the inhibition of transcription by Pol II degradation (**Figure 3.3.4B, orange and black bars**). Interestingly, however, there was a sharp increase in the occupancy of dMi-2-GFP at the promoters of Hsp70 and Broad genes upon Triptolide treatment. Similarly, I detected an increased occupancy of dMi-2-GFP at the promoter of Vrille gene (data not shown). I also tracked the changes at the gene body of Vrille, Hsp83 and ultraspricle (USP) genes and found an increment of dMi-2-GFP at each gene upon Triptolide treatment (**Figure 3.3.4B, green and violet bars**). On the other hand, no changes of dMi-2-GFP occupancy were detected at the Ecdysone receptor (EcR) and at an untranscribed intergenic region (2R). This could perhaps be because of inaccessibility for dMi-2 at these locations in the genome perhaps because other proteins are occupying these locations in the genome.

Taken together, this data supports the hypothesis that interaction with RNA hinders dMi-2 from interacting with chromatin and when there is a global depletion of RNA in the cell, dMi-2 becomes free to potentially associate with chromatin regions that provide binding sites for its association.

A



B



**Figure 3.3.4: Binding to RNA prevents dMi-2 from interaction with chromatin**

**(A)** S2 cells were treated with 10  $\mu\text{M}$  Triptolide or equal volume of DMSO for 3 hours. RNA was isolated from each sample and cDNA synthesized. qPCRs were performed using primers for GAPDH1, Vrille, EcR and Hsp83. The expression of each gene was measured relative to DMSO control which was set to 1. Individual data points of three independent biological replicates are plotted. Error bars indicate standard error between three biological experiments. **(B)** dMi-2-GFP, dLINT1-GFP and untagged cells were treated with 10  $\mu\text{M}$  triptolide or DMSO treated for 3 hours. Chromatin was prepared and immunoprecipitation was performed using GFP-trap beads. DNA was eluted and qPCRs performed for the genes indicated. Legend on the right indicates cell types and treatment conditions. Individual data

points of two independent experiments were plotted. Error bars indicate standard error between two independent biological experiments.

## 4 Discussion

Recent studies have implicated several chromatin associated proteins, including ATP-dependent chromatin remodellers, as RNA binding proteins. Interaction with RNA has been proposed to both help recruit some of the chromatin regulators to their target genes or expel them from chromatin. Thereby, RNA interaction has been implicated in modulating the function of chromatin regulators. In this study, I focus on dMi-2, an ATP-dependent chromatin remodeling enzyme which I show to possess broad RNA binding activity. I report the *in vitro* and *in vivo* characterization and the potential functional significance of the RNA binding property of dMi-2.

Some of the RNA binding proteins (RBPs) specifically bind to RNAs that have distinct structural and sequence features (Hendrickson et al., 2016). For example, the phage MS2 coat protein (MS2-CP) interacts with a stem-loop RNA called MS2-SL (He et al., 2016). Similarly, HuR (also called Elavl1), an RBP that is involved in stabilizing the RNAs has been established to preferentially bind AU-rich elements in mRNAs (Siang et al., 2020). Likewise, P40, a RBP that is encoded by human LINE-1 retrotransposons binds to L1Hs RNA in a sequence specific manner where it recognizes two regions within L1Hs RNA called site-A and site- B (Hohjoh and Singer, 1997). On the other hand, several of the RBPs bind to RNA with a more relaxed sequence specificity, that is, they can bind to a wide range of RNAs with different sequence and structural features. Most of such RBPs lack a canonical RNA binding domain. For instance, the heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) which binds to numerous RNAs and is involved in several RNA processing events such as splicing, stability, nuclear export and translation (Levengood and Tolbert, 2019). Similarly, polypyrimidine tract-binding protein (PTB), also called as hnRNP1, a well characterized RBP that regulates splicing, polyadenylation and initiation of translation, has been shown to bind numerous RNAs both in the nucleus as well as in cytoplasm (Xue et al., 2009). In *in vitro* experiments it was shown to bind to several RNAs indicating a relaxed sequence specificity, yet it bound to U-tract containing RNA with higher affinity (Pérez et al., 1997). *In vivo* it employs this degree of specificity to bind to U-tract containing mRNAs. In this study, I demonstrate that dMi-2 binds to RNA with a relaxed sequence or structural specificity. dMi-2 was previously reported to bind to a 230-mer RNA fragment from the Hsp70Aa transcript (Murawska et al., 2011). Using electrophoretic mobility shift assays, I illustrate that dMi-2 can bind to several other regions within the Hsp70Aa RNA. I reveal its binding to several RNA

fragments transcribed from an unrelated gene - Pgant35a, and several other physiological and non-physiological RNAs of varying length, sequence and structure, in vitro. However, I observed quantifiable differences in the efficiency between some RNAs for binding to dMi-2. For example, a 100-nucleotide long RNA from Pgant35a gene bound with less affinity than the RNA of equal length from Hsp70Aa gene. This points to dMi-2 possessing ability to bind several RNAs and yet possess some degree of specificity to bind certain RNAs better, just like PTB.

Several protein-RNA interactions are highly salt sensitive, such proteins often lose or have a highly diminished RNA binding in higher salt conditions. Protein-RNA complexes of P40 (Hohjoh and Singer, 1997), Pentatricopeptide repeat protein 10 (PPR10) (McDermott et al., 2018) and Fox1 (Auweter et al., 2006) are highly salt sensitive. Similarly, PRC2 RNA interaction is sensitive to salt with loss in crosslinking to HOTAIR RNA at 400 mM salt (Wang et al., 2017). In contrast, I found that dMi-2 interacted with RNA in the presence of significantly higher salt concentrations and in the presence of high concentrations of detergent. These findings suggest that dMi-2 is a robust RNA binding protein and that stronger non-ionic interactions between the base or sugar-phosphate backbone of RNA and dMi-2 might play the major role in its RNA binding (detailed below).

In this study, I map the major RNA binding region of dMi-2 to the first 376 amino acids. The ATPase domain, the chromodomains, the PHD fingers or the C-terminal portion revealed no significant RNA binding in my assays. The HMG-box domain in the N-terminal portion which lies within the first 376 amino acids also did not bind RNA on its own in my assay. This study presents the regions flanking the HMG-box on either side, 1-117 and 238-376 amino acids, as major RNA binding regions in dMi-2. The amino acid sequence between 1-117 and 238-376 is rich in positively charged amino acids such as lysine (K) and arginine (R). Perhaps, the positive charge in these regions plays some role in RNA interaction, however, it is important to note that the HMG-box which did not show any RNA binding on its own, also contains few of the K and R stretches. Therefore, it is improbable that the charge alone drives the RNA binding properties of 1-117 and 238-376. This hypothesis is supported by the ability of dMi-2 to bind RNA in the presence of high salt concentrations (see above). In several non-canonical RBPs that do not possess any *bona fide* RNA binding domain (RBD) like an RNA recognition motif (RRM), K homology (KH) domain or a cold shock (CD) domain, the intrinsically disordered regions (IDRs) have been observed to be key determinants in RNA interaction (Castello et al., 2013; He et al., 2016). Both of these regions in dMi-2, 1-117 and

239-376 amino acids, are predicted to be intrinsically disordered. In fact, some reports suggest that the IDRs which are often rich in K, R and G (Glycine) amino acids, could directly interact with RNA (Castello et al., 2016). The mechanisms of such interaction, however, remains elusive, partly due to the challenge that the IDRs confer in X-ray crystallography and other structural studies. In case of dMi-2, however, it would not be wrong to dismiss the ionic interactions as the major mechanism of its interaction with RNA because, I observe strong RNA binding at salt concentrations at which the ionic interactions are largely inhibited. Therefore, perhaps the IDRs in dMi-2 directly interact with the RNA through more stronger hydrogen bond interactions or  $\pi$  - stacking interactions. The dMi-2 could directly contact the sugar-phosphate backbone or bases of the RNA through hydrogen bond formation leading to a stable complex. Or perhaps the R-residues, which are known to form the strong  $\pi$  - stacking interactions, within the IDRs contact the bases in the RNA to form a highly stable complex. There is also a possibility that the IDRs of dMi-2, upon interaction with RNA, fold in a best-fit pocket for the RNA and the combined effect of the charge and spatial arrangement allows the interaction of dMi-2 with RNA to be strong to resist high salt conditions. Needless to say, these models are speculative and would need thorough studies such as Nuclear Magnetic Resonance (NMR) or X-ray crystallographic analysis for substantiation.

Using iCLIP, this study illustrates the RNA binding characteristics of dMi-2 *in vivo* in an unbiased manner for the first time. The *in vivo* data from iCLIP unveils interaction of dMi-2 with a large number of RNAs, in agreement with the promiscuous binding character of dMi-2 that was found *in vitro*. In a heat shock state, dMi-2 interacts with the nascent RNA fragments transcribed from the heat shock genes (Murawska et al., 2011). The genome-wide RNA binding data of CHD4 has reported its interaction with processed mRNAs (Hendrickson et al., 2016). The iCLIP data from this study agrees with the CHD4 RNA binding profile and shows that dMi-2 interacts with the mature mRNAs, almost exclusively. This indicates an evolutionarily conserved RNA binding property between *Drosophila* dMi-2 and human CHD4. Furthermore, the iCLIP data indicates that the interaction takes place at the 3'-end of the mRNAs. This is opposite to the interaction of dMi-2 to its target genes where it largely localizes towards the promoter region. This indicates that the interaction of dMi-2 with RNA might not lead to its recruitment to the genome and that the recruitment mechanism on heat shock genes, in heat shock condition, might not be a general role of the RNA binding function of dMi-2. Perhaps, the interaction of dMi-2 at the 3'-end of the RNAs is due to the prevalence of highly folded structures at the 3'-ends of RNAs or conceivably dMi-2 plays a role in the processing of the 3'- ends of RNAs during polyadenylation. The polyadenylated mRNAs are

exported out of the nucleus for translation to take place. For such a process, the mRNA is made competent for export by disassembly of the polyadenylation factors (Casañal et al., 2017; Qu et al., 2009). It is possible that dMi-2 either helps in the disassembly of the polyadenylation factors or prepares the mRNAs for the nuclear export. However, in depth analysis into the role of dMi-2 at the 3'-end of RNAs need to be carried to ascertain the exact function. A logical experiment to address this question would be to monitor the mRNA nuclear export under dMi-2 depleted condition.

I found that dMi-2 preferentially binds to Guanine-rich RNAs *in vitro*. The iCLIP data presents the Guanine nucleotide as the most frequently cross-linked nucleotide after Uridine, however, the Uridine enrichment at the cross-linking sites could very likely arise from the inherent positive bias due to the use of UV-C in the iCLIP procedure (Sugimoto et al., 2012). Therefore, the *in vivo* results complement the *in vitro* results in this study. However, even though Guanine nucleotide appeared enriched at the cross-linking site, the RNA reads bound by dMi-2 did not appear to be significantly rich in Guanine content *in vivo*. This could possibly arise from the differences between the physiological environment and the *in vitro* conditions. For the *in vitro* experiments, I used purified dMi-2, however, *in vivo*, dMi-2 exists in two complexes – the multi-subunit Nucleosome remodeling and deacetylase (NuRD) and dMep-1 containing, dMec complex. It is plausible that the other components in the NuRD complex or dMep-1 in the dMec complex, exert some influence in determining the RNA fragments that the dMi-2 binds to. This could be either by influencing the spatial arrangement of dMi-2 or because some of the other proteins with which dMi-2 associates, bind to RNA themselves, thereby, cooperatively determining the RNAs that the complex interacts with. One approach to determine this would be to express each of the NuRD components and dMep-1 individually and test their RNA binding abilities independently. The second approach would be to express all the dMi-2 associated proteins and reconstitute the NuRD complex and dMec complex *in vitro*. The RNA binding profile of such a complex would presumably be closer to what appears in our iCLIP data.

Furthermore, this study reveals that interaction with RNA inhibits the remodeling activity of dMi-2. I used *in vitro* remodeling assays to detect the effects of different types of RNA on the remodeling function of dMi-2. While two Guanine-rich RNA fragments ((G<sub>4</sub>A<sub>4</sub>)<sub>4</sub> and (GA)<sub>20</sub>) inhibited the remodeling of dMi-2, another RNA fragment (Hsp70Aa #6) of equal length but with less Guanine content (30% compared with 40% and 50% in (G<sub>4</sub>A<sub>4</sub>)<sub>4</sub> and (GA)<sub>20</sub>, respectively), had no apparent effect on the remodeling activity. Further analysis revealed



that Hsp70Aa #6 RNA bound dMi-2 with much less affinity (in EMSAs) compared with the RNAs that inhibited the remodeling. Perhaps, the inhibition of remodeling activity is brought about by a mechanism that depends on strong RNA binding. The affinity of binding to dMi-2 might depend on the Guanine-richness of the RNA. However, the differences in Guanine content are similar (10%) between (GA)<sub>20</sub> RNA and (G<sub>4</sub>A<sub>4</sub>)<sub>4</sub> RNA, and between (G<sub>4</sub>A<sub>4</sub>)<sub>4</sub> RNA and Hsp70Aa #6 RNA (10%). Yet, Hsp70Aa #6 RNA bound poorly compared with the other two RNAs and had no apparent effect on remodeling. This could arise from three possibilities - a) dMi-2 requires a minimum threshold of Guanine nucleotides which could be more than 30% for it to bind an RNA fragment with high affinity, b) the pattern of Guanine nucleotides in an RNA might influence its interaction with dMi-2. In (G<sub>4</sub>A<sub>4</sub>)<sub>4</sub> RNA, Guanine nucleotides are regularly interspersed as blocks of four nucleotides separated by four non-Guanine nucleotides. Similarly, in (GA)<sub>20</sub> RNA, Guanine nucleotides are regularly interspersed with every second nucleotide being a Guanine nucleotide. This is not the case for Hsp70Aa #6 RNA where Guanine nucleotides are more randomly distributed, and/or c) there are structural differences between the RNAs and that dMi-2 binds better to the secondary structures formed by (G<sub>4</sub>A<sub>4</sub>)<sub>4</sub> RNA and (GA)<sub>20</sub> RNA but not to the structure formed by Hsp70Aa #6 RNA. Although, there is a clear impossibility of secondary structures formed through canonical base-pairing due to unavailability of Cytosine (C) and Uracil (U) nucleotides in (GA)<sub>20</sub> RNA, however, a similarly Guanine-rich RNA with same Guanine-content and length - (GU)<sub>20</sub> RNA, has been shown to fold into its own secondary structures (Wang et al., 2017). In fact, the (GA)<sub>20</sub> RNA used in this study, indeed appeared to fold in its own distinct structure, different from a G-quadruplex structure of (G<sub>4</sub>A<sub>4</sub>)<sub>4</sub> RNA in the CD spectroscopy analysis. Just, like the stretches of Guanine-nucleotides in (G<sub>4</sub>A<sub>4</sub>)<sub>4</sub> RNA exert their influence on the folding, perhaps the Guanine-nucleotides in (GA)<sub>20</sub> RNA also exert a folding pressure resulting in a structure that is distinct, yet has features that are recognized by dMi-2 preferentially.

The RNA binding of several chromatin modifying proteins has been proposed to either aid their recruitment to the chromatin or antagonize their association with chromatin. However, depending upon the context, RNA binding can exert both effects as well. This is vividly true for PRC2 as its interaction with RNA has been proposed to recruit it to the PRC2 target genes while the RNA binding evicts PRC2 from PRC2 non-target genes (Beltran et al., 2016; Long et al., 2020; Riising et al., 2014). In this study, I found that binding to RNA possibly counteracts the association of dMi-2 to chromatin. The RNase A mediated degradation of cellular RNA leads to increased association of dMi-2 with chromatin *in vitro*. To support this view, I illustrate that inhibition of transcription also leads to an increased occupancy of dMi-

2 with certain chromatin loci *in vivo*. It seems that the RNAs in the cell act as a sponge for dMi-2 to bind to, preventing unwanted dMi-2 interaction with chromatin and thereby aberrant gene repression. It is likely that this function of RNA binding works in cue with other gene expression signals. I propose that in a repressed state when there is no signal for transcription at a locus, therefore, low to no mRNA is available for dMi-2 to interact with, allowing dMi-2 to interact with the chromatin and maintaining the repressed state. However, when there is a gene activation signal for example, recruitment of TATA-binding protein (TBP) to the promoters of respective genes leading to gene expression cue and subsequent production of RNA from the locus (Coleman et al., 1999), dMi-2 is competed off from the chromatin by its interaction with RNA and an activated state is maintained (**Figure 4.0, Model**). It is pertinent to mention that RNA competes off dMi-2 from DNA *in vitro* (Murawska et al., 2011). In addition, CHD4, the human homolog of dMi-2, is competed off from the nucleosome by RNA *in vitro* (data not shown), supporting the proposed model.

This study characterizes the RNA binding of dMi-2 and mainly uses EMSA to detect the binding. While as such EMSA is a robust assay, the chemical equilibrium of binding reactions is not maintained during electrophoresis (Hellman and Fried, 2007). Such a caveat could miss protein-RNA complexes that dissociate rapidly. Therefore, the approximate dissociation constants determined in this study might differ from the dissociation constants in the solution. An alternative to electrophoresis dependent binding quantifications is the use of fluorescently labelled probes (RNA) and measurement of the dissociation constants in-solution by using techniques such as fluorescence polarization (FP) assays. Also, the *in vitro* RNA binding assays lack the exact physiological conditions, therefore, inside a cell, the binding patterns of dMi-2 could be slightly different. This was evident in dMi-2 preference for binding to Guanine-rich RNA *in vitro* and its lack thereof *in vivo*. Therefore, as stated above, reconstituting the complexes in which dMi-2 exists in the cell would recapitulate its physiological state and provide a more accurate RNA binding profile of dMi-2. Additionally, I have stained the EMSA gels with SYBR gold stain. The sensitivity of SYBR gold for in-gel detection is at least 10 times more than ethidium bromide, however, it was not able to stain Poly (G), Poly (U) and (CA)<sub>20</sub> RNAs properly. Therefore, in such cases a different stain or the detection mechanism, such as radio-labelling could be done to improve the detection as well as accuracy of the dissociation constants. Furthermore, this study indicates the RNA binding regions as 1-117 and 238-376 amino acids in the N-terminal region of dMi-2. Further truncations within these regions must be performed to determine the minimal RNA binding motifs. Also experiments such as Hydrogen deuterium exchange mass spectrometry (HDX- MS) or Nuclear magnetic

resonance (NMR) of dMi-2-RNA complexes could be conducted to detect the precise RNA-protein interacting surfaces. The details of the precise amino acids involved in RNA binding that are revealed through such data should be used to dissect the exact RNA binding function of dMi-2. This can be achieved by creating the dMi-2 loss-of-RNA-binding mutant flies and analysis on the phenotypes. Such RNA binding mutant dMi-2 can also be created in *Drosophila* cells and analysis of gene expression through RNA-sequencing could be done to inform on its role in gene expression regulation.

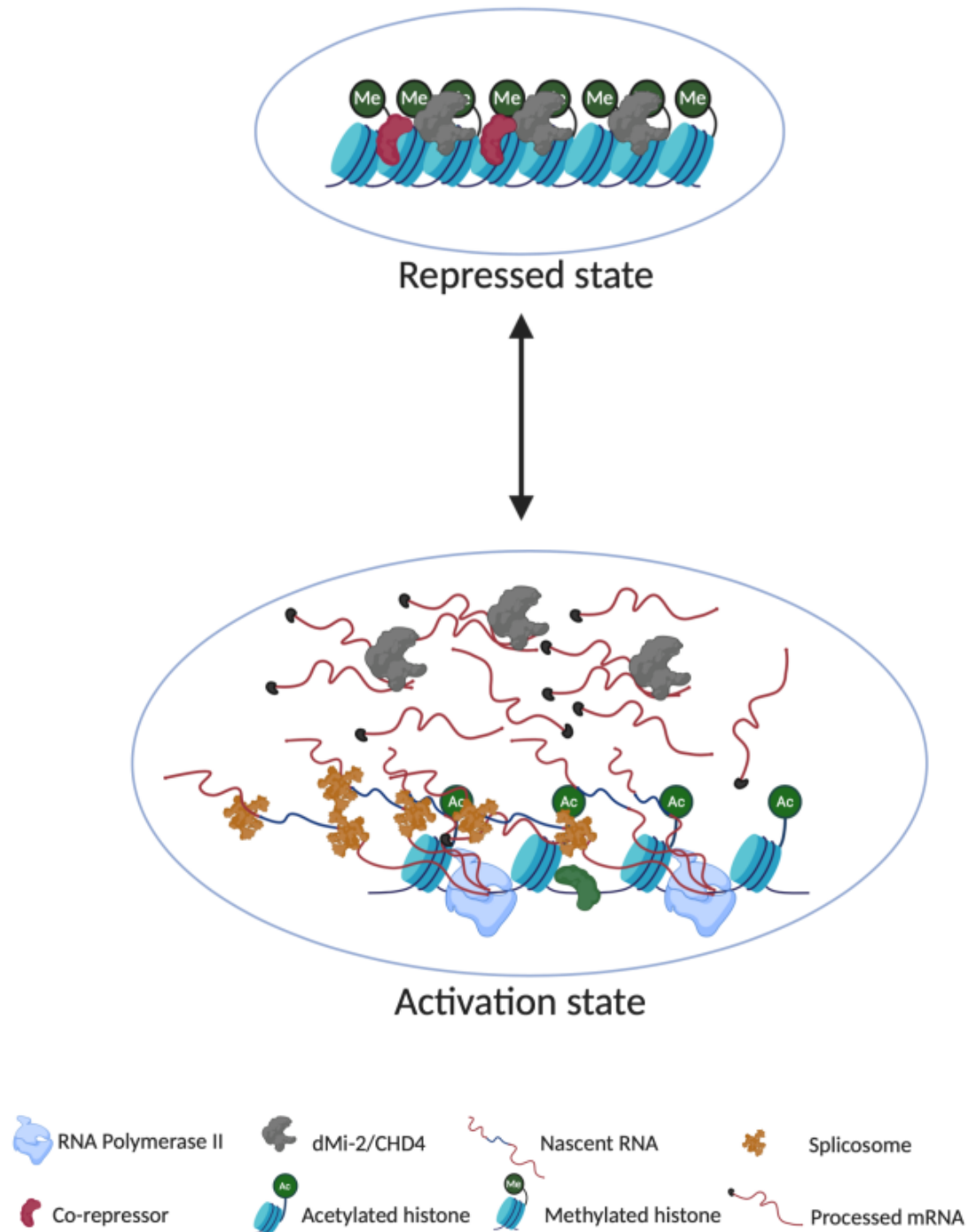
In-depth studies into the dMi-2 interaction at the 3'- end of mRNAs should be conducted to unveil the role of this interaction. As suggested above, perhaps dMi-2 plays a role either in the 3'-end processing or the nuclear export of mRNAs. Therefore, to substantiate these possibilities two strategies could be adopted: a) monitoring the 3'-end processing and nuclear transport in a dMi-2 knockdown situation or a dMi-2 RNA binding mutant condition and b) establishing whether dMi-2 physically interacts either with the 3'-end processing machinery or the nuclear export machinery by performing protein-protein interaction analysis such as mass spectrometry. Protein-protein interaction assays based on immunoprecipitation experiments are generally conducted after depletion of DNA and RNA that could mediate protein-protein interactions. However, if dMi-2 plays a role in 3'-end processing, then its interaction with the 3'-end processing machinery may be mediated by mRNA. Therefore, care must be taken to not deplete RNA while conducting such an experiment.

The RNAs that were identified as bound by dMi-2 *in vivo* in this study should also be studied in detail to dissect how dMi-2 interaction affects their biological functions. For example, if dMi-2 binds an RNA which codes for an enzyme involved in a cellular process such as glycolysis. Loss of dMi-2 RNA binding function may impact the processing, stability or translation of this RNA. Such effects of dMi-2 on its RNA partners must be studied to provide a more in-depth role of dMi-2 RNA binding function. This can be achieved by a) monitoring the changes in the expression levels of these RNAs between a wild type condition and a dMi-2 RNA binding mutant condition using RNA sequencing analysis, b) monitoring the protein levels coded by the RNAs found in dMi-2 iCLIP in a wild type and dMi-2 RNA binding mutant condition by proteomic studies.

This study proposes that the RNAs that bind to dMi-2 with high affinity also inhibit its remodeling activity. This is based on our *in vitro* remodeling assays. However, it is important

to note that a) the *in vitro* environment does not recapitulate the exact physiological state and b) I use mononucleosomes in my assays. In the physiological state, the dMi-2 deals with arrays of nucleosomes and this might affect its *in vivo* behavior to be slightly different from the *in vitro* observations. However, despite these caveats, the inhibition of remodeling by RNA interaction provides a mechanism for maintaining gene expression of actively transcribed genes. The depletion of RNA led to an increased association of dMi-2 to chromatin. A thorough genome-wide analysis of that aspect could be conducted to detect the exact location of the newly chromatin-associated dMi-2, perhaps by ChIP-seq of an RNA-binding mutant version of dMi-2. It is likely, that such a mutant dMi-2 causes an aberrant gene expression repression of the loci to which it binds upon loss of its RNA interaction properties.

In conclusion, this thesis establishes the RNA binding role of the canonical chromatin remodeler – dMi-2. This study provides evidence that the intrinsically disordered regions (IDRs) in the N-terminal regions of dMi-2 mediate its RNA binding. It establishes the promiscuous RNA binding nature and identifies thousands of RNAs that dMi-2 binds in *Drosophila* S2 cells. The findings here establish, for the first time, an inhibitory role of RNA binding on the remodeling activity of dMi-2. Finally, this thesis proposes a model that suggests RNA binding antagonizes dMi-2 association with chromatin which may provide a mechanism to maintain the expression of actively transcribed genes.



**Figure 4.0: Model**

dMi-2 interacts with chromatin at transcriptionally repressed loci. Upon transcriptional activation, dMi-2 interacts with the mRNA produced from the actively transcribed genes. RNA interaction competes dMi-2 off from the chromatin, acting as a 'sponge'. At actively transcribed genes, RNA interaction with dMi-2 provides a mechanism of maintaining the gene expression.

## 5 Summary

### 5.1 Summary

Chromatin is maintained in a dynamic relaxed or repressed state such that DNA binding sites for the transcription machinery are either accessible or occluded. This way the gene expression is regulated. There are several chromatin regulators including ATP-dependent chromatin remodelers that help change or modulate the conformation of the chromatin to repress or to facilitate gene expression. Recent studies have implicated several chromatin remodelers in RNA binding. However, the precise function of this RNA binding property is intensively debated due to our limited understanding of its function. In this thesis, dMi-2, an ATP-dependent chromatin remodeler, has been studied for its interaction with RNA and a model is proposed for the effects of RNA binding on its function that supports the hypothesis that RNA binding modulates the function of chromatin remodelers.

In the first part of this thesis, the RNA binding properties of dMi-2 were characterized *in vitro*. dMi-2 formed biochemically stable complexes with several RNAs of varying sequences. This suggested a binding of promiscuous nature. However, dMi-2 bound some RNAs with higher affinity compared with others. Further, the RNA binding regions in dMi-2 were mapped. The major RNA binding regions of dMi-2 were located in its N-terminal part. Analysis of the type of sequences that dMi-2 bound to revealed a preference for G-rich RNA.

In the second part of this study, *in vivo* RNA binding of dMi-2 was characterized in an unbiased genome-wide manner. The individual nucleotide-resolution crosslinking and immunoprecipitation (iCLIP) experiment recapitulated the promiscuous nature of RNA binding that was found in *in vitro* assays. iCLIP revealed thousands of different dMi-2-binding RNAs. Further analysis revealed that dMi-2 interacted almost exclusively near the 3'-end of the mRNAs. At the cross-linking site, G-nucleotides were enriched.

In the third part of this thesis, the effects of RNA binding on the function of dMi-2 were studied. It was shown that RNA binding inhibits the remodeling activity of dMi-2. Further analysis illustrated that the RNAs that bound dMi-2 with higher affinity also inhibited its remodeling better compared with the RNAs that bound with lower affinity. Biochemical analysis revealed that the depletion of RNA led to an increased chromatin occupancy of dMi-2. Likewise, inhibition of transcription *in vivo* also resulted to an increased chromatin occupancy of dMi-2.

In summary, the results of this thesis support a model which suggests that dMi-2-RNA interaction causes inhibition of the remodeling function of dMi-2. This may be caused by eviction of dMi-2 from chromatin upon RNA interaction. At actively transcribing genes, RNA competes dMi-2 away from the chromatin and at repressed genes, dMi-2 has increased chromatin occupancy due to lack of RNA. This thesis contributes to the understanding of the role of RNA interaction in modulating the function of chromatin remodelers.

## 5.2 Zusammenfassung

Chromatin wird in einem dynamischen entspannten oder unterdrückten Zustand gehalten, so dass DNA-Bindungsstellen für die Transkriptionsmaschinerie entweder zugänglich oder verschlossen sind. Auf diese Weise wird die Genexpression reguliert. Es gibt mehrere Chromatinregulatoren, darunter ATP-abhängige Chromatin-Remodeler, die dazu beitragen, die Konformation des Chromatins zu verändern oder zu modulieren, um die Genexpression zu unterdrücken oder zu ermöglichen. Neuere Studien haben mehrere Chromatin-Remodeler in die RNA-Bindung einbezogen. Die genaue Funktion dieser RNA-Bindungseigenschaft wird jedoch aufgrund unseres begrenzten Verständnisses der Funktion intensiv diskutiert. In dieser Arbeit wurde dMi-2, ein ATP-abhängiger Chromatin-Remodeler, auf seine Interaktion mit RNA untersucht und ein Modell für die Auswirkungen der RNA-Bindung auf seine Funktion vorgeschlagen, das die Hypothese unterstützt, dass die RNA-Bindung die Funktion von Chromatin-Remodelern moduliert.

Im ersten Teil dieser Arbeit wurden die RNA-Bindungseigenschaften von dMi-2 *in vitro* charakterisiert. dMi-2 bildete biochemisch stabile Komplexe mit mehreren RNAs unterschiedlicher Sequenz. Dies deutete auf eine Bindung promiskuitiver Natur hin. Allerdings band dMi-2 einige RNAs mit höherer Affinität im Vergleich zu anderen. Weiterhin wurden die RNA-Bindungsregionen in dMi-2 kartiert. Die wichtigsten RNA-Bindungsregionen von dMi-2 befanden sich in seinem N-terminalen Teil. Die Analyse der Art der Sequenzen, an die dMi-2 gebunden hat, ergab eine Präferenz für G-reiche RNA.

Im zweiten Teil dieser Studie wurde die *in vivo* RNA-Bindung von dMi-2 genomweit charakterisiert. Das iCLIP-Experiment (individual nucleotide-resolution crosslinking and immunoprecipitation) rekapitulierte die promiskuitive Natur der RNA-Bindung, die in den *in vitro*-assays gefunden wurde. iCLIP enthüllte tausende von verschiedenen dMi-2-bindenden RNAs. Weitere Analysen zeigten, dass dMi-2 fast ausschließlich in der Nähe des 3'-Endes der mRNAs interagiert. An der Vernetzungsstelle waren G-Nukleotide angereichert.

Im dritten Teil dieser Arbeit wurden die Auswirkungen der RNA-Bindung auf die Funktion von dMi-2 untersucht. Es wurde gezeigt, dass die RNA-Bindung die Remodeling-Aktivität von dMi-2 hemmt. Weitere Analysen zeigten, dass die RNAs, die dMi-2 mit höherer Affinität banden, auch dessen Remodeling-Aktivität besser hemmten als die RNAs, die mit geringerer Affinität banden. Die biochemische Analyse zeigte, dass der Abbau der RNA zu einer erhöhten



Chromatinbesetzung von dMi-2 führte. Ebenso führte die Inhibition der Transkription *in vivo* zu einer erhöhten Chromatinbesetzung von dMi-2.

Zusammenfassend unterstützen die Ergebnisse dieser Arbeit ein Modell, das nahelegt, dass die dMi-2-RNA-Interaktion eine Hemmung der Remodeling-Funktion von dMi-2 bewirkt. Dies könnte durch die Verdrängung von dMi-2 aus dem Chromatin bei RNA-Interaktion verursacht werden. An aktiv transkribierten Genen löst die RNA dMi-2 vom Chromatin ab und an repressiven Genen hat dMi-2 eine erhöhte Chromatinbesetzung aufgrund des Mangels an RNA. Diese Arbeit trägt zum Verständnis der Rolle der RNA-Interaktion bei der Modulation der Funktion von Chromatin-Remodelern bei.

## 6 References

Akhtar, A., Zink, D., and Becker, P.B. (2000). Chromodomains are protein-RNA interaction modules. *Nature*.

Akopian, D., Shen, K., Zhang, X., and Shan, S.O. (2013). Signal recognition particle: An essential protein-targeting machine.

Alendar, A., Lambooj, J.P., Bhaskaran, R., Lancini, C., Song, J.Y., Van Vugt, H., Snoek, M., and Berns, A. (2020). Gene expression regulation by the Chromodomain helicase DNA-binding protein 9 (CHD9) chromatin remodeler is dispensable for murine development. *PLoS ONE*.

ALLFREY, V.G., FAULKNER, R., and MIRSKY, A.E. (1964). ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. *Proc. Natl. Acad. Sci. USA*.

Alver, B.H., Kim, K.H., Lu, P., Wang, X., Manchester, H.E., Wang, W., Haswell, J.R., Park, P.J., and Roberts, C.W.M. (2017). The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nature Communications*.

Amir, M., Kumar, V., Dohare, R., Islam, A., Ahmad, F., and Hassan, M.I. (2018). Sequence, structure and evolutionary analysis of cold shock domain proteins, a member of OB fold family. *Journal of Evolutionary Biology*.

Annunziato, A.T. (2008). DNA Packaging: Nucleosomes and Chromatin. In *Nature Education*, p.

Arents, G., and Moudrianakis, E.N. (1993). Topography of the histone octamer surface: Repeating structural motifs utilized in the docking of nucleosomal DNA. *Proceedings of the National Academy of Sciences of the United States of America*.

Asad, Z., Pandey, A., Babu, A., Sun, Y., Shevade, K., Kapoor, S., Ullah, I., Ranjan, S., Scaria, V., Bajpai, R., et al. (2016). Rescue of neural crest-derived phenotypes in a zebrafish CHARGE model by Sox10 downregulation. *Human Molecular Genetics*.

Ascano, M., Mukherjee, N., Bandaru, P., Miller, J.B., Nusbaum, J.D., Corcoran, D.L., Langlois, C., Munschauer, M., Dewell, S., Hafner, M., et al. (2012). FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature*.

Auweter, S.D., Oberstrass, F.C., and Allain, F.H.T. (2006). Sequence-specific binding of single-stranded RNA: Is there a code for recognition? *Nucleic Acids Research*.

Badenhorst, P., Voas, M., Rebay, I., and Wu, C. (2002). Biological functions of the ISWI chromatin remodeling complex NURF. *Genes and Development*.

Bae, H.J., Dubarry, M., Jeon, J., Soares, L.M., Dargemont, C., Kim, J., Geli, V., and Buratowski, S. (2020). The Set1 N-terminal domain and Swd2 interact with RNA polymerase II CTD to recruit COMPASS. *Nature Communications*.

- Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017). Biomolecular condensates: Organizers of cellular biochemistry.
- Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*.
- Basta, J., and Rauchman, M. (2015). The nucleosome remodeling and deacetylase complex in development and disease.
- Bavetsias, V., and Linardopoulos, S. (2015). Aurora kinase inhibitors: Current status and outlook.
- Bedford, M.T., and Clarke, S.G. (2009). Protein Arginine Methylation in Mammals: Who, What, and Why.
- Beltran, M., Yates, C.M., Skalska, L., Dawson, M., Reis, F.P., Viiri, K., Fisher, C.L., Sibley, C.R., Foster, B.M., Bartke, T., et al. (2016). The interaction of PRC2 with RNA or chromatin s mutually antagonistic. *Genome Research*.
- Beltran, M., Tavares, M., Justin, N., Khandelwal, G., Ambrose, J., Foster, B.M., Worlock, K.B., Tvardovskiy, A., Kunzelmann, S., Herrero, J., et al. (2019). G-tract RNA removes Polycomb repressive complex 2 from genes. *Nature Structural and Molecular Biology*.
- Beusch, I., Barraud, P., Moursy, A., Cléry, A., and Allain, F.H.T. (2017). Tandem hnRNP A1 RNA recognition motifs act in concert to repress the splicing of survival motor neuron exon 7. *ELife*.
- Bhatia, S., Pawar, H., Dasari, V., Mishra, R.K., Chandrashekar, S., and Brahmachari, V. (2010). Chromatin remodeling protein INO80 has a role in regulation of homeotic gene expression in *Drosophila*. *Genes to Cells*.
- Bhaumik, S.R., Smith, E., and Shilatifard, A. (2007). Covalent modifications of histones during development and disease pathogenesis (*Nat Struct Mol Biol*).
- Bonnici, J., Tumber, A., Kawamura, A., and Schofield, C.J. (2018). Inhibitors of both the N-methyl lysyl- and arginyl-demethylase activities of the JmjC oxygenases. *Philosophical Transactions of the Royal Society B: Biological Sciences*.
- Bornelöv, S., Reynolds, N., Xenophontos, M., Gharbi, S., Johnstone, E., Floyd, R., Ralser, M., Signolet, J., Loos, R., Dietmann, S., et al. (2018). The Nucleosome Remodeling and Deacetylation Complex Modulates Chromatin Structure at Sites of Active Transcription to Fine-Tune Gene Expression. *Molecular Cell*.
- Bouazoune, K., and Brehm, A. (2006). ATP-dependent chromatin remodeling complexes in *Drosophila*.
- Bouazoune, K., Mitterweger, A., Längst, G., Imhof, A., Akhtar, A., Becker, P.B., and Brehm, A. (2002). The dMi-2 chromodomains are DNA binding modules important for ATP-dependent nucleosome mobilization. *EMBO Journal*.

- Bracken, A.P., Brien, G.L., and Verrijzer, C.P. (2019). Dangerous liaisons: Interplay between SWI/SNF, NURD, and polycomb in chromatin regulation and cancer. *Genes and Development* 33, 936–959.
- Brahma, S., Udugama, M.I., Kim, J., Hada, A., Bhardwaj, S.K., Hailu, S.G., Lee, T.H., and Bartholomew, B. (2017). INO80 exchanges H2A.Z for H2A by translocating on DNA proximal to histone dimers. *Nature Communications*.
- Brehm, A., Längst, G., Kehle, J., Clapier, C.R., Imhof, A., Eberharter, A., Müller, J., and Becker, P.B. (2000). DMI-2 and ISWI chromatin remodelling factors have distinct nucleosome binding and mobilization properties. *EMBO Journal*.
- Bugga, L., McDaniel, I.E., Engie, L., and Armstrong, J.A. (2013). The *Drosophila melanogaster* CHD1 Chromatin Remodeling Factor Modulates Global Chromosome Structure and Counteracts HP1a and H3K9me2. *PLoS ONE*.
- Casañal, A., Kumar, A., Hill, C.H., Easter, A.D., Emsley, P., Degliesposti, G., Gordiyenko, Y., Santhanam, B., Wolf, J., Wiederhold, K., et al. (2017). Architecture of eukaryotic mRNA 3'-end processing machinery. *Science*.
- Castello, A., Fischer, B., Hentze, M.W., and Preiss, T. (2013). RNA-binding proteins in Mendelian disease. *Trends in Genetics*.
- Castello, A., Fischer, B., Frese, C.K., Horos, R., Alleaume, A.M., Foehr, S., Curk, T., Krijgsveld, J., and Hentze, M.W. (2016). Comprehensive Identification of RNA-Binding Domains in Human Cells. *Molecular Cell*.
- Ceccacci, E., and Minucci, S. (2016). Inhibition of histone deacetylases in cancer therapy: Lessons from leukaemia. *British Journal of Cancer* 114, 605–611.
- Centore, R.C., Sandoval, G.J., Soares, L.M.M., Kadoch, C., and Chan, H.M. (2020). Mammalian SWI/SNF Chromatin Remodeling Complexes: Emerging Mechanisms and Therapeutic Strategies.
- Chang, B., Chen, Y., Zhao, Y., and Bruick, R.K. (2007). JMJD6 is a histone arginine demethylase. *Science*.
- Chen, P.B., Hung, J.H., Hickman, T.L., Coles, A.H., Carey, J.F., Weng, Z., Chu, F., and Fazio, T.G. (2013). Hdac6 regulates Tip60-p400 function in stem cells. *ELife*.
- Clapier, C.R., and Cairns, B.R. (2012). Regulation of ISWI involves inhibitory modules antagonized by nucleosomal epitopes. *Nature*.
- Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes.
- Cléry, A., and Allain, F.H.-T. (2012). FROM STRUCTURE TO FUNCTION OF RNA BINDING DOMAINS. *Landes Biosciences* 137–158.
- Coleman, R.A., Taggart, A.K.P., Burma, S., Chicca, J.J., and Pugh, B.F. (1999). TFIIA regulates TBP and TFIID dimers. *Molecular Cell*.

- Collins, R.T., Furukawa, T., Tanese, N., and Treisman, J.E. (1999). Osa associates with the Brahma chromatin remodeling complex and promotes the activation of some target genes. *EMBO Journal*.
- Conn, K.L., Hendzel, M.J., and Schang, L.M. (2013). The Differential Mobilization of Histones H3.1 and H3.3 by Herpes Simplex Virus 1 Relates Histone Dynamics to the Assembly of Viral Chromatin. *PLoS Pathogens*.
- Conrad, T., Cavalli, F.M.G., Holz, H., Hallacli, E., Kind, J., Ilik, I., Vaquerizas, J.M., Luscombe, N.M., and Akhtar, A. (2012). The MOF Chromobarrel Domain Controls Genome-wide H4K16 Acetylation and Spreading of the MSL Complex. *Developmental Cell*.
- Corley, M., Burns, M.C., and Yeo, G.W. (2020). How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms.
- Corona, D.F.V., and Tamkun, J.W. (2004). Multiple roles for ISWI in transcription, chromosome organization and DNA replication.
- Crick, F. (1970). Central dogma of molecular biology. *Nature*, 227(5258). *Nature Biotechnology*.
- Daubresse, G., Deuring, R., Moore, L., Papoulas, O., Zakrajsek, I., Waldrip, W.R., Scott, M.P., Kennison, J.A., and Tamkun, J.W. (1999). The *Drosophila* kismet gene is related to chromatin-remodeling factors and is required for both segmentation and segment identity. *Development*.
- Davidovich, C., Zheng, L., Goodrich, K.J., and Cech, T.R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. *Nature Structural and Molecular Biology*.
- Davidovich, C., Wang, X., Cifuentes-Rojas, C., Goodrich, K.J., Gooding, A.R., Lee, J.T., and Cech, T.R. (2015). Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. *Molecular Cell*.
- De Ruijter, A.J.M., Van Gennip, A.H., Caron, H.N., Kemp, S., and Van Kuilenburg, A.B.P. (2003). Histone deacetylases (HDACs): Characterization of the classical HDAC family. *Biochemical Journal* 370, 737–749.
- Dekker, J., and Mirny, L. (2016). The 3D Genome as Moderator of Chromosomal Communication.
- Dekker, J., and Misteli, T. (2015). Long-range chromatin interactions. *Cold Spring Harbor Perspectives in Biology*.
- DeLisle, A.J., Graves, R.A., Marzluff, W.F., and Johnson, L.F. (1983). Regulation of histone mRNA production and stability in serum-stimulated mouse 3T6 fibroblasts. *Molecular and Cellular Biology*.
- Delmas, V., Stokes, D.G., and Perry, R.P. (1993). A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain. *Proceedings of the National Academy of Sciences of the United States of America*.

- Diehl, K.L., Ge, E.J., Weinberg, D.N., Jani, K.S., Allis, C.D., and Muir, T.W. (2019). PRC2 engages a bivalent H3K27M-H3K27me3 dinucleosome inhibitor. *Proceedings of the National Academy of Sciences of the United States of America*.
- Dill, K.A., and MacCallum, J.L. (2012). The protein-folding problem, 50 years on.
- Dunleavy, E.M., Roche, D., Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., Daigo, Y., Nakatani, Y., and Almouzni-Pettinotti, G. (2009). HJURP Is a Cell-Cycle-Dependent Maintenance and Deposition Factor of CENP-A at Centromeres. *Cell*.
- Duthie, S.M., Nesterova, T.B., Formstone, E.J., Keohane, A.M., Turner, B.M., Zakian, S.M., and Brockdorff, N. (1999). Xist RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in cis. *Human Molecular Genetics*.
- Ebbert, R., Birkmann, A., and Schüller, H.J. (1999). The product of the SNF2/SWI2 paralogue INO80 of *Saccharomyces cerevisiae* required for efficient expression of various yeast structural genes is part of a high-molecular-weight protein complex. *Molecular Microbiology*.
- Ee, L.S., McCannell, K.N., Tang, Y., Fernandes, N., Hardy, W.R., Green, M.R., Chu, F., and Fazio, T.G. (2017). An Embryonic Stem Cell-Specific NuRD Complex Functions through Interaction with WDR5. *Stem Cell Reports*.
- Elfring, L.K., Deuring, R., McCallum, C.M., Peterson, C.L., and Tamkun, J.W. (1994). Identification and characterization of *Drosophila* relatives of the yeast transcriptional activator SNF2/SWI2. *Molecular and Cellular Biology* 14, 2225–2234.
- Euskirchen, G., Auerbach, R.K., and Snyder, M. (2012). SWI/SNF chromatin-remodeling factors: Multiscale analyses and diverse functions. *Journal of Biological Chemistry* 287, 30897–30905.
- Eustermann, S., Schall, K., Kostrewa, D., Lakomek, K., Strauss, M., Moldt, M., and Hopfner, K.P. (2018). Structural basis for ATP-dependent chromatin remodelling by the INO80 complex. *Nature*.
- Farnung, L., Ochmann, M., and Cramer, P. (2020). Nucleosome-CHD4 chromatin remodeler structure maps human disease mutations. *ELife*.
- Fasulo, B., Deuring, R., Murawska, M., Gause, M., Dorigi, K.M., Schaaf, C.A., Dorsett, D., Brehm, A., and Tamkun, J.W. (2012). The *Drosophila* Mi-2 Chromatin-Remodeling Factor Regulates Higher-Order Chromatin Structure and Cohesin Dynamics In Vivo. *PLoS Genetics*.
- Fischle, W., Wang, Y., and Allis, C.D. (2003). Histone and chromatin cross-talk. *Current Opinion in Cell Biology*.
- Foltz, D.R., Jansen, L.E.T., Bailey, A.O., Yates, J.R., Bassett, E.A., Wood, S., Black, B.E., and Cleveland, D.W. (2009). Centromere-Specific Assembly of CENP-A Nucleosomes Is Mediated by HJURP. *Cell*.
- Fong, N., Saldi, T., Sheridan, R.M., Cortazar, M.A., and Bentley, D.L. (2017). RNA Pol II Dynamics Modulate Co-transcriptional Chromatin Modification, CTD Phosphorylation, and Transcriptional Direction. *Molecular Cell*.

- Francesco Fiorentino, A.M., and Rotili, D. (2020). *Histone Modifications in Therapy* (Elsevier).
- Franke, A., and Baker, B.S. (1999). The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in *Drosophila*. *Molecular Cell*.
- Gambogi, C.W., and Black, B.E. (2019). The nucleosomes that mark centromere location on chromosomes old and new.
- Garraway, L.A., and Lander, E.S. (2013). *Lessons from the cancer genome*.
- Gaspar-Maia, A., and Sevilla, A. (2016). *Chromatin Dynamics and Epigenetics of Stem Cells and Stem-Like Cancer Cells* (Elsevier Inc.).
- Ghasemi, M., Pawar, H., Mishra, R.K., and Brahmachari, V. (2015). The functional diversity of *Drosophila* Ino80 in development. *Mechanisms of Development*.
- Gierer, A., and Schramm, G. (1956). Infectivity of ribonucleic acid from tobacco mosaic virus [3].
- Gräff, J., and Tsai, L.H. (2013). Histone acetylation: Molecular mnemonics on the chromatin.
- Grant, P.A., Duggan, L., Côté, J., Roberts, S.M., Brownell, J.E., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C.D., Winston, F., et al. (1997). Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: Characterization of an ada complex and the saga (spt/ada) complex. *Genes and Development*.
- Griffith, F. (1928). The Significance of Pneumococcal Types. *Journal of Hygiene* 27, 113–159.
- Han, K., and Nepal, C. (2007). PRI-Modeler: Extracting RNA structural elements from PDB files of protein-RNA complexes. *FEBS Letters*.
- Hans, F., and Dimitrov, S. (2001). Histone H3 phosphorylation and cell division.
- Hasegawa, Y., Brockdorff, N., Kawano, S., Tsutui, K., Tsutui, K., and Nakagawa, S. (2010). The matrix protein hnRNP U is required for chromosomal localization of xist RNA. *Developmental Cell*.
- Hauk, G., McKnight, J.N., Nodelman, I.M., and Bowman, G.D. (2010). The Chromodomains of the Chd1 Chromatin Remodeler Regulate DNA Access to the ATPase Motor. *Molecular Cell*.
- He, C., Sidoli, S., Warneford-Thomson, R., Tatomer, D.C., Wilusz, J.E., Garcia, B.A., and Bonasio, R. (2016). High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. *Molecular Cell*.
- He, J., Xuan, T., Xin, T., An, H., Wang, J., Zhao, G., and Li, M. (2014). Evidence for chromatin-remodeling complex PBAP-controlled maintenance of the *Drosophila* ovarian germline stem cells. *PLoS ONE*.
- Hellman, L.M., and Fried, M.G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature Protocols*.

- Hendrickson, D., Kelley, D.R., Tenen, D., Bernstein, B., and Rinn, J.L. (2016). Widespread RNA binding by chromatin-associated proteins. *Genome Biology*.
- nHentze, M.W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A brave new world of RNA-binding proteins.
- Hodawadekar, S.C., and Marmorstein, R. (2007). Chemistry of acetyl transfer by histone modifying enzymes: Structure, mechanism and implications for effector design.
- Hohjoh, H., and Singer, M.F. (1997). Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. *EMBO Journal*.
- Howman, E.V., Fowler, K.J., Newson, A.J., Redward, S., MacDonald, A.C., Kalitsis, P., and Choo, K.H.A. (2000). Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. *Proceedings of the National Academy of Sciences of the United States of America*.
- Hsu, J.Y., Sun, Z.W., Li, X., Reuben, M., Tatchell, K., Bishop, D.K., Grushcow, J.M., Brame, C.J., Caldwell, J.A., Hunt, D.F., et al. (2000). Mitotic phosphorylation of histone H3 is governed by Ip11/aurora kinase and Glc7/PP1 phosphatase in budding yeast and nematodes. *Cell*.
- Huang, H., Sabari, B.R., Garcia, B.A., David Allis, C., and Zhao, Y. (2014). SnapShot: Histone Modifications. *CELL 159*, 458-458.e1.
- Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L.E., Sibley, C.R., Sugimoto, Y., Tajnik, M., König, J., and Ule, J. (2014). iCLIP: Protein-RNA interactions at nucleotide resolution. *Methods*.
- Hyun, K., Jeon, J., Park, K., and Kim, J. (2017). Writing, erasing and reading histone lysine methylations.
- Inoue, A., Jiang, L., Lu, F., and Zhang, Y. (2017). Genomic imprinting of Xist by maternal H3K27me3. *Genes and Development*.
- Ito, T., Bulger, M., Pazin, M.J., Kobayashi, R., and Kadonaga, J.T. (1997). ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell*.
- Jain, K., Fraser, C.S., Marunde, M.R., Parker, M.M., Sagum, C., Burg, J.M., Hall, N., Popova, I.K., Rodriguez, K.L., Vaidya, A., et al. (2020). Characterization of the plant homeodomain (PHD) reader family for their histone tail interactions. *Epigenetics and Chromatin*.
- Järvelin, A.I., Noerenberg, M., Davis, I., and Castello, A. (2016). The new (dis)order in RNA regulation.
- Jha, S., and Dutta, A. (2009). RVB1/RVB2: Running Rings around Molecular Biology.
- Jiang, Y.-K., Zou, J.-W., Wu, Y.-Q., Zhang, N., Yu, Q.-S., and Jiang, Y.-J. (2009). Molecular dynamics simulation on HP1 protein binding by histone H3 tail methylation and phosphorylation. *International Journal of Quantum Chemistry 109*, 746–755.



- Jónsson, Z.O., Jha, S., Wohlschlegel, J.A., and Dutta, A. (2004). Rvb1p/Rvb2p recruit Arp5p and assemble a functional Ino80 chromatin remodeling complex. *Molecular Cell*.
- Kandel, E.R. (2001). *The molecular biology of memory storage: A dialogue between genes and synapses*.
- Kargapolova, Y., Rehim, R., Kayserili, H., Brühl, J., Zirkel, A., Li, Y., Yigit, G., Hoischen, A., Frank, S., Russ, N., et al. (2020). Overarching control of autophagy and DNA damage response by CHD6 revealed by modeling a rare human pathology. *BioRxiv*.
- Kebede, A.F., Nieborak, A., Shahidian, L.Z., Le Gras, S., Richter, F., Gómez, D.A., Baltissen, M.P., Meszaros, G., De Fatima Magliarelli, H., Taudt, A., et al. (2017). Histone propionylation is a mark of active chromatin. *Nature Structural and Molecular Biology*.
- Kehle, J., Beuchle, D., Treuheit, S., Christen, B., Kennison, J.A., and Bienz, M. (1998). dMi-2, a hunchback-interacting protein that functions in Polycomb repression. *Science*.
- Keller, C.I., and Akhtar, A. (2015). The MSL complex: Juggling RNA-protein interactions for dosage compensation and beyond.
- Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Morales, D.R., Thomas, K., Presser, A., Bernstein, B.E., Van Oudenaarden, A., et al. (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 United States of America*.
- Khan, S.A. (2015). Global histone post-translational modifications and cancer: Biomarkers for diagnosis, prognosis and treatment? *World Journal of Biological Chemistry*.
- Khan, S.A., Amnekar, R., Khade, B., Barreto, S.G., Ramadwar, M., Shrikhande, S.V., and Gupta, S. (2016). p38-MAPK/MSK1-mediated overexpression of histone H3 serine 10 phosphorylation defines distance-dependent prognostic value of negative resection margin in gastric cancer. *Clinical Epigenetics*.
- Khorasanizadeh, S. (2004). *The Nucleosome: From Genomic Organization to Genomic Regulation* (Cell Press).
- Kim, T.W., Kang, B.H., Jang, H., Kwak, S., Shin, J., Kim, H., Lee, S.E., Lee, S.M., Lee, J.H., Kim, J.H., et al. (2015). Ctbp2 Modulates NuRD-Mediated Deacetylation of H3K27 and Facilitates PRC2-Mediated H3K27me3 in Active Embryonic Stem Cell Genes during Exit from Pluripotency. *Stem Cells*.
- Kind, J., Vaquerizas, J.M., Gebhardt, P., Gentzel, M., Luscombe, N.M., Bertone, P., and Akhtar, A. (2008). Genome-wide Analysis Reveals MOF as a Key Regulator of Dosage Compensation and Gene Expression in *Drosophila*. *Cell*.
- Klattenhoff, C.A., Scheuermann, J.C., Surface, L.E., Bradley, R.K., Fields, P.A., Steinhauser, M.L., Ding, H., Butty, V.L., Torrey, L., Haas, S., et al. (2013). Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell*.
- Kljashorny, V., Nikonov, S., Ovchinnikov, L., Lyabin, D., Vodovar, N., Curmi, P., and Manivet, P. (2015). The Cold Shock Domain of YB-1 segregates RNA from DNA by non-bonded interactions. *PLoS ONE*.

Komar, D., and Juszczynski, P. (2020). Rebelled epigenome: histone H3S10 phosphorylation and H3S10 kinases in cancer biology and therapy.

Kovač, K., Sauer, A., Mačinković, I., Awe, S., Finkernagel, F., Hoffmeister, H., Fuchs, A., Müller, R., Rathke, C., Längst, G., et al. (2018). Tumour-associated missense mutations in the dMi-2 ATPase alters nucleosome remodelling properties in a mutation-specific manner. *Nature Communications*.

Kraushaar, D.C., Chen, Z., Tang, Q., Cui, K., Zhang, J., and Zhao, K. (2018). The gene repressor complex NuRD interacts with the histone variant H3.3 at promoters of active genes. *Genome Research*.

Kreher, J., Kovač, K., Bouazoune, K., Mačinković, I., Ernst, A.L., Engelen, E., Pahl, R., Finkernagel, F., Murawska, M., Ullah, I., et al. (2017). EcR recruits dMi-2 and increases efficiency of dMi-2-mediated remodelling to constrain transcription of hormone-regulated genes. *Nature Communications*.

Kretz, M., and Meister, G. (2014). RNA Binding of PRC2: Promiscuous or Well Ordered?

Kunert, N., Wagner, E., Murawska, M., Klinker, H., Kremmer, E., and Brehm, A. (2009). dMec: A novel Mi-2 chromatin remodelling complex involved in transcriptional repression. *EMBO Journal*.

Lai, A.Y., and Wade, P.A. (2011). Cancer biology and NuRD: A multifaceted chromatin remodelling complex.

Lai, W.K.M., and Pugh, B.F. (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication.

Lan, F., and Shi, Y. (2009). Epigenetic regulation: Methylation of histone and non-histone proteins.

Latcheva, N.K., Viveiros, J.M., and Marena, D.R. (2019). The Drosophila Chromodomain Protein Kismet Activates Steroid Hormone Receptor Transcription to Govern Axon Pruning and Memory In Vivo. *IScience*.

Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression.

Lee, J.S., Smith, E., and Shilatifard, A. (2010). The Language of Histone Crosstalk.

Leulliot, N., and Varani, G. (2001). Current topics in RNA-protein recognition: Control of specificity and biological function through induced fit and conformational capture. *Biochemistry*.

Levengood, J.D., and Tolbert, B.S. (2019). Idiosyncrasies of hnRNP A1-RNA recognition: Can binding mode influence function. *Seminars in Cell and Developmental Biology*.

Li, S., Ali, S., Duan, X., Liu, S., Du, J., Liu, C., Dai, H., Zhou, M., Zhou, L., Yang, L., et al. (2018). JMJD1B Demethylates H4R3me2s and H3K9me2 to Facilitate Gene Expression for Development of Hematopoietic Stem and Progenitor Cells. *Cell Reports*.

- Liang, Z., Brown, K.E., Carroll, T., Taylor, B., Vidal, I.F., Hendrich, B., Rueda, D., Fisher, A.G., and Merckenschlager, M. (2017). A high-resolution map of transcriptional repression. *ELife*.
- Long, Y., Bolanos, B., Gong, L., Liu, W., Goodrich, K.J., Yang, X., Chen, S., Gooding, A.R., Maegley, K.A., Gajiwala, K.S., et al. (2017). Conserved RNA-binding specificity of polycomb repressive complex 2 is achieved by dispersed amino acid patches in EZH2. *ELife*.
- Long, Y., Hwang, T., Gooding, A.R., Goodrich, K.J., Rinn, J.L., and Cech, T.R. (2020). RNA is essential for PRC2 chromatin occupancy and function in human pluripotent stem cells. *Nature Genetics*.
- Low, J.K.K., Silva, A.P.G., Sharifi Tabar, M., Torrado, M., Webb, S.R., Parker, B.L., Sana, M., Smits, C., Schmidberger, J.W., Brillault, L., et al. (2020). The Nucleosome Remodeling and Deacetylase Complex Has an Asymmetric, Dynamic, and Modular Architecture. *Cell Reports*.
- Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*.
- MačInković, I., Theofel, I., Hundertmark, T., Kovač, K., Awe, S., Lenz, J., Forné, I., Lamp, B., Nist, A., Imhof, A., et al. (2019). Distinct CoREST complexes act in a cell-type-specific manner. *Nucleic Acids Research*.
- Maehara, K., Harada, A., Sato, Y., Matsumoto, M., Nakayama, K.I., Kimura, H., and Ohkawa, Y. (2015). Tissue-specific expression of histone H3 variants diversified after species separation. *Epigenetics and Chromatin*.
- Mak, W., Baxter, J., Silva, J., Newall, A.E., and Otte, A.P. (2002). Mitotically stable association of polycomb group proteins Eed and Enx1 with the inactive X chromosome in trophoblast stem cells. *Current Biology*.
- Malik, H.S., and Henikoff, S. (2003). Phylogenomics of the nucleosome.
- Manelyte, L., and Langst, G. (2013). Chromatin Remodelers and Their Way of Action. In *Chromatin Remodelling*, p.
- Marques, J.G., Gryder, B.E., Pavlovic, B., Chung, Y., Ngo, Q.A., Frommelt, F., Gstaiger, M., Song, Y., Benischke, K., Laubscher, D., et al. (2020). NURD subunit CHD4 regulates super-enhancer accessibility in rhabdomyosarcoma and represents a general tumor dependency. *ELife*.
- Martin, C., and Zhang, Y. (2005). The diverse functions of histone lysine methylation.
- Martire, S., and Banaszynski, L.A. (2020a). The roles of histone variants in fine-tuning chromatin organization and function.
- Martire, S., and Banaszynski, L.A. (2020b). The roles of histone variants in fine-tuning chromatin organization and function. *Nature Reviews Molecular Cell Biology*.

- Mathieu, E.L., Finkernagel, F., Murawska, M., Scharfe, M., Jarek, M., and Brehm, A. (2012). Recruitment of the ATP-dependent chromatin remodeler dMi-2 to the transcribed region of active heat shock genes. *Nucleic Acids Research*.
- Matthews, M.M., Thomas, J.M., Zheng, Y., Tran, K., Phelps, K.J., Scott, A.I., Havel, J., Fisher, A.J., and Beal, P.A. (2016). Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nature Structural and Molecular Biology*.
- McDaniel, I.E., Lee, J.M., Berger, M.S., Hanagami, C.K., and Armstrong, J.A. (2008). Investigations of CHD1 function in transcription and development of *Drosophila melanogaster*. *Genetics*.
- McDermott, J.J., Civic, B., and Barkan, A. (2018). Effects of RNA structure and salt concentration on the affinity and kinetics of interactions between pentatricopeptide repeat proteins and their RNA ligands. *PLoS ONE*.
- Mendel, G. (1901). *Versuche über Pflanzenhybriden. Zwei abhandlungen. (1865 und 1869.)* (Leipzig: W. Engelmann).
- Miller, J.L., and Grant, P.A. (2013). The role of DNA methylation and histone modifications in transcriptional regulation in humans. *Subcellular Biochemistry* 61, 289–317.
- Misteli, T. (2020). *The Self-Organizing Genome: Principles of Genome Architecture and Function*.
- Mittal, P., and Roberts, C.W.M. (2020). The SWI/SNF complex in cancer — biology, biomarkers and therapy.
- Mohrmann, L., Langenberg, K., Krijgsveld, J., Kal, A.J., Heck, A.J.R., and Verrijzer, C.P. (2004). Differential Targeting of Two Distinct SWI/SNF-Related *Drosophila* Chromatin-Remodeling Complexes. *Molecular and Cellular Biology*.
- Morettini, S., Tribus, M., Zeilner, A., Sebald, J., Campo-Fernandez, B., Scheran, G., Wörle, H., Podhraski, V., Fyodorov, D.V., and Lusser, A. (2011). The chromodomains of CHD1 are critical for enzymatic activity but less important for chromatin localization. *Nucleic Acids Research*.
- Morozova, N., Myers, J., and Shamooy, Y. (2006). Protein-RNA interactions: Exploring binding patterns with a three-dimensional superposition analysis of high resolution structures. *Bioinformatics*.
- Murawska, M., and Brehm, A. (2011). CHD chromatin remodelers and the transcription cycle.
- Murawska, M., Kunert, N., van Vugt, J., Längst, G., Kremmer, E., Logie, C., and Brehm, A. (2008). dCHD3, a Novel ATP-Dependent Chromatin Remodeler Associated with Sites of Active Transcription. *Molecular and Cellular Biology*.
- Murawska, M., Hassler, M., Renkawitz-Pohl, R., Ladurner, A., and Brehm, A. (2011). Stress-induced PARP activation mediates recruitment of *Drosophila* Mi-2 to promote heat shock gene expression. *PLoS Genetics*.

- Musselman, C.A., and Kutateladze, T.G. (2011). Handpicking epigenetic marks with PHD fingers.
- Narlikar, G.J., Sundaramoorthy, R., and Owen-Hughes, T. (2013). Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes.
- Neugeborn, L., and Carlson, M. (1984). Genes affecting the regulation of SUC2 gene expression by glucose repression in *Saccharomyces cerevisiae*. *Genetics*.
- Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Molecular Cell*.
- Ng, S.S., Yue, W.W., Oppermann, U., and Klose, R.J. (2009). Dynamic protein methylation in chromatin biology.
- Oki, M., Aihara, H., and Ito, T. (2007). Role of histone phosphorylation in chromatin dynamics and its implications in diseases. *Sub-Cellular Biochemistry*.
- Onofrio, A., Parisi, G., Punzi, G., Todisco, S., Di Noia, M.A., Bossis, F., Turi, A., De Grassi, A., and Pierrri, C.L. (2014). Distance-dependent hydrophobic-hydrophobic contacts in protein folding simulations. *Physical Chemistry Chemical Physics*.
- Papamichos-Chronakis, M., Watanabe, S., Rando, O.J., and Peterson, C.L. (2011). Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. *Cell*.
- Park, S.Y., and Kim, J.S. (2020). A short guide to histone deacetylases including recent progress on class II enzymes.
- Park, D., Shivram, H., and Iyer, V.R. (2014). Chd1 co-localizes with early transcription elongation factors independently of H3K36 methylation and releases stalled RNA polymerase II at introns. *Epigenetics and Chromatin*.
- Parthun, M.R. (2007). Hat1: The emerging cellular roles of a type B histone acetyltransferase.
- Pelletier, J., Graff, J., Ruggero, D., and Sonenberg, N. (2015). Targeting the eIF4F translation initiation complex: A critical nexus for cancer development. *Cancer Research*.
- Pérez, I., Lin, C.H., McAfee, J.G., and Patton, J.G. (1997). Mutation of PTB binding sites causes misregulation of alternative 3' splice site selection in vivo. *RNA*.
- Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD.
- Platt, J.L., Rogers, B.J., Rogers, K.C., Harwood, A.J., and Kimmel, A.R. (2013). Different CHD chromatin remodelers are required for expression of distinct gene sets and specific stages during development of *dictyostelium discoideum*. *Development (Cambridge)*.
- Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Tong, I.L., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., et al. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell*.

- Puglisi, J.D., Chen, L., Blanchard, S., and Frankel, A.D. (1995). Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science*.
- Puvvula, P.K., Desetty, R.D., Pineau, P., Marchio, A., Moon, A., Dejean, A., and Bischof, O. (2014). Long noncoding RNA PANDA and scaffold-attachment-factor SAFA control senescence entry and exit. *Nature Communications*.
- Qu, X., Lykke-Andersen, S., Nasser, T., Saguez, C., Bertrand, E., Jensen, T.H., and Moore, C. (2009). Assembly of an Export-Competent mRNP Is Needed for Efficient Release of the 3'-End Processing Complex after Polyadenylation. *Molecular and Cellular Biology*.
- Rakow, S., Pullamsetti, S.S., Bauer, U.M., and Bouchard, C. (2020). Assaying epigenome functions of PRMTs and their substrates. *Methods*.
- Reynolds, N., Latos, P., Hynes-Allen, A., Loos, R., Leaford, D., O'Shaughnessy, A., Mosaku, O., Signolet, J., Brennecke, P., Kalkan, T., et al. (2012). NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. *Cell Stem Cell*.
- Ribeiro-Silva, C., Vermeulen, W., and Lans, H. (2019). SWI/SNF: Complex complexes in genome stability and cancer.
- Ricci, M.A., Manzo, C., García-Parajo, M.F., Lakadamyali, M., and Cosma, M.P. (2015). Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell*.
- Riising, E.M., Comet, I., Leblanc, B., Wu, X., Johansen, J.V., and Helin, K. (2014). Gene silencing triggers polycomb repressive complex 2 recruitment to CpG Islands genome wide. *Molecular Cell*.
- Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Bruggmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell*.
- Rivera, C., Saavedra, F., Alvarez, F., Díaz-Celis, C., Ugalde, V., Li, J., Forné, I., Gurard-Levin, Z.A., Almouzni, G., Imhof, A., et al. (2015). Methylation of histone H3 lysine 9 occurs during translation. *Nucleic Acids Research*.
- Rosenfeld, J.A., Wang, Z., Schones, D.E., Zhao, K., DeSalle, R., and Zhang, M.Q. (2009). Determination of enriched histone modifications in non-genic portions of the human genome. *BMC Genomics*.
- Rossetto, D., Avvakumov, N., and Côté, J. (2012). Histone phosphorylation: A chromatin modification involved in diverse nuclear events.
- Ruthenburg, A.J., Allis, C.D., and Wysocka, J. (2007). Methylation of Lysine 4 on Histone H3: Intricacy of Writing and Reading a Single Epigenetic Mark.
- Ryan, D.P., Sundaramoorthy, R., Martin, D., Singh, V., and Owen-Hughes, T. (2011). The DNA-binding domain of the Chd1 chromatin-remodelling enzyme contains SANT and SLIDE domains. *EMBO Journal*.

- Sakabe, K., Wang, Z., and Hart, G.W. (2010).  $\beta$ -N-acetylglucosamine (O-GlcNAc) is part of the histone code. *Proceedings of the National Academy of Sciences of the United States of America* *107*, 19915–19920.
- Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C.T., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. *Nature*.
- Saunders, L.R., and Verdin, E. (2007). Sirtuins: Critical regulators at the crossroads between cancer and aging.
- Sawicka, A., and Seiser, C. (2014). Sensing core histone phosphorylation - A matter of perfect timing.
- Scacchetti, A., Brueckner, L., Jain, D., Schauer, T., Zhang, X., Schnorrer, F., van Steensel, B., Straub, T., and Becker, P.B. (2018). CHRAC/ACF contribute to the repressive ground state of chromatin. *Life Science Alliance* *1*.
- Scacchetti, A., Schauer, T., Reim, A., Apostolou, Z., Sparr, A.C., Krause, S., Heun, P., Wierer, M., and Becker, P.B. (2020). Drosophila SWR1 and NuA4 complexes are defined by DOMINO isoforms. *ELife*.
- Shepard, P.J., and Hertel, K.J. (2009). The SR protein family.
- Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell*.
- Shoemaker, C.B., and Chalkley, R. (1978). An H3 histone-specific kinase isolated from bovine thymus chromatin. *Journal of Biological Chemistry*.
- Siang, D.T.C., Lim, Y.C., Kyaw, A.M.M., Win, K.N., Chia, S.Y., Degirmenci, U., Hu, X., Tan, B.C., Walet, A.C.E., Sun, L., et al. (2020). The RNA-binding protein HuR is a negative regulator in adipogenesis. *Nature Communications*.
- Simic, R., Lindstrom, D.L., Tran, H.G., Roinick, K.L., Costa, P.J., Johnson, A.D., Hartzog, G.A., and Arndt, K.M. (2003). Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. *EMBO Journal*.
- Sims, J.K., and Wade, P.A. (2011). SnapShot: Chromatin remodeling: CHD. *Cell* *144*, 626-626.e1.
- Singh, G., Pratt, G., Yeo, G.W., and Moore, M.J. (2015). The clothes make the mRNA: Past and present trends in mRNP fashion.
- Siomi, H., Matunis, M.J., Michael, W.M., and Dreyfuss, G. (1993). The pre-mRNA binding K protein contains a novel evolutionary conserved motif. *Nucleic Acids Research*.
- Srinivasan, S., Armstrong, J.A., Deuring, R., Dahlsveen, I.K., McNeill, H., and Tamkun, J.W. (2005). The Drosophila trithorax group protein kismet facilitates an early step in transcriptional elongation by RNA polymerase II. *Development*.

- Stielow, B., Sapetschnig, A., Krüger, I., Kunert, N., Brehm, A., Boutros, M., and Suske, G. (2008). Identification of SUMO-Dependent Chromatin-Associated Transcriptional Repression Components by a Genome-wide RNAi Screen. *Molecular Cell*.
- Strein, C., Alleaume, A.M., Rothbauer, U., Hentze, M.W., and Castello, A. (2014). A versatile assay for RNA-binding proteins in living cells. *RNA*.
- Stroud, H., Otero, S., Desvoyes, B., Ramírez-Parra, E., Jacobsen, S.E., and Gutierrez, C. (2012). Genome-wide analysis of histone H3.1 and H3.3 variants in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*.
- Sugimoto, Y., König, J., Hussain, S., Zupan, B., Curk, T., Frye, M., and Ule, J. (2012). Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biology*.
- Sun, Z., Zhang, Y., Jia, J., Fang, Y., Tang, Y., Wu, H., and Fang, D. (2020). H3K36me<sub>3</sub>, message from chromatin to DNA damage repair. *Cell and Bioscience*.
- Szerlong, H.J., and Hansen, J.C. (2011). Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure.
- Tachiwana, H., Osakabe, A., Shiga, T., Miya, Y., Kimura, H., Kagawa, W., and Kurumizaka, H. (2011). Structures of human nucleosomes containing major histone H3 variants. *Acta Crystallographica Section D: Biological Crystallography*.
- Tagami, H., Ray-Gallet, D., Almouzni, G., and Nakatani, Y. (2004). Histone H3.1 and H3.3 Complexes Mediate Nucleosome Assembly Pathways Dependent or Independent of DNA Synthesis. *Cell*.
- Talbert, P.B., and Henikoff, S. (2010). Histone variants ancient wrap artists of the epigenome.
- Tencer, A.H., Cox, K.L., Di, L., Bridgers, J.B., Lyu, J., Wang, X., Sims, J.K., Weaver, T.M., Allen, H.F., Zhang, Y., et al. (2017). Covalent Modifications of Histone H3K9 Promote Binding of CHD3. *Cell Reports*.
- Teplova, M., Malinina, L., Darnell, J.C., Song, J., Lu, M., Abagyan, R., Musunuru, K., Teplov, A., Burley, S.K., Darnell, R.B., et al. (2011). Protein-RNA and protein-protein recognition by dual KH1/2 domains of the neuronal splicing factor Nova-1. *Structure*.
- Tjeertes, J.V., Miller, K.M., and Jackson, S.P. (2009). Screen for DNA-damage-responsive histone modifications identifies H3K9Ac and H3K56Ac in human cells. *EMBO Journal*.
- Tong, J.K., Hassig, C.A., Schnitzler, G.R., Kingston, R.E., and Schreiber, S.L. (1998). Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature*.
- Torchy, M.P., Hamiche, A., and Klaholz, B.P. (2015). Structure and function insights into the NuRD chromatin remodeling complex.
- Tsukiyama, T., and Wu, C. (1995). Purification and Properties of an ATP-Dependent Nucleosome Remodeling Factor.



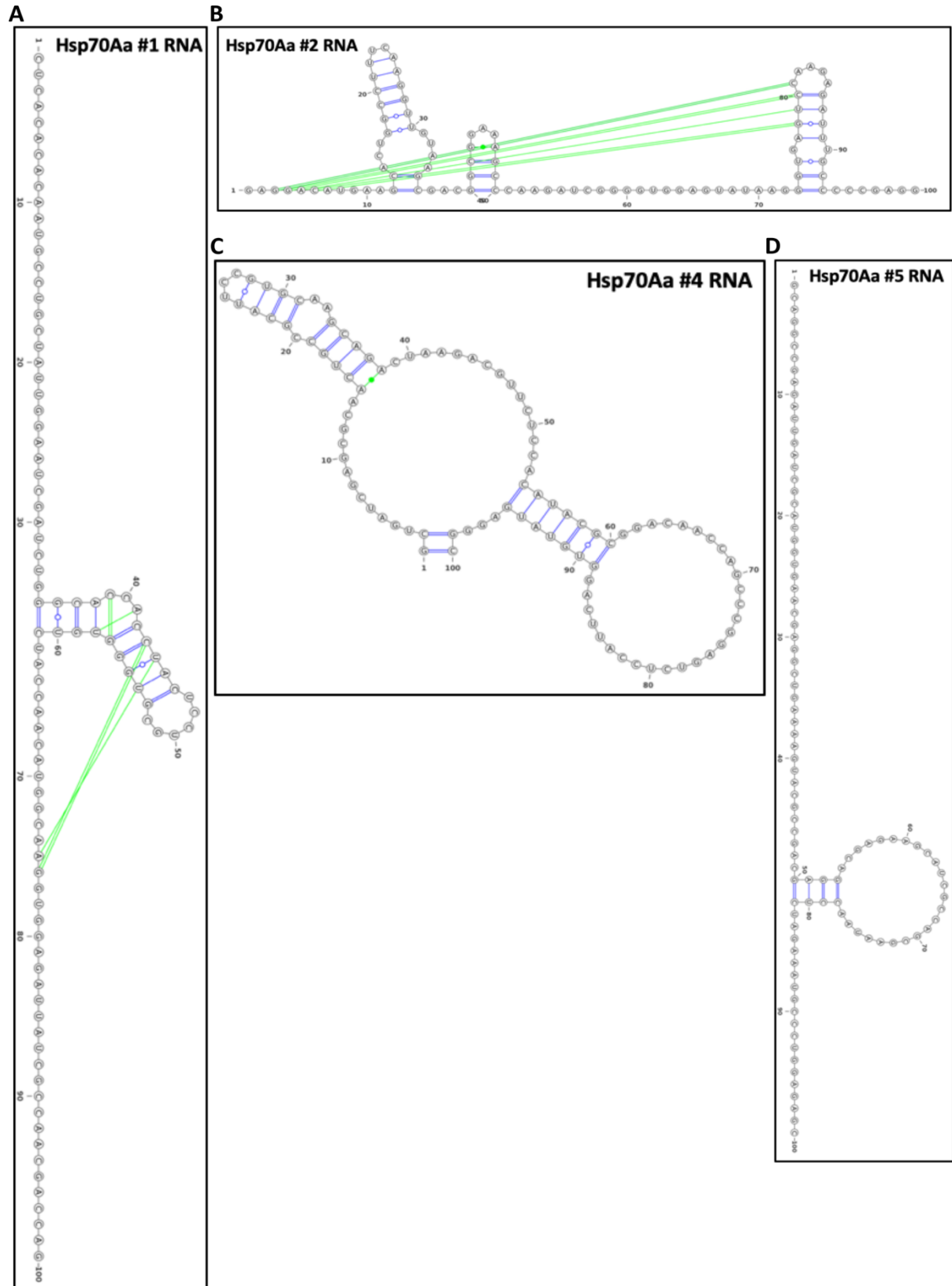
- Tsukiyama, T., Daniel, C., Tamkun, J., and Wu, C. (1995). /SW/, a Member of the SliW2KNF2 ATPase Family, Encodes the 140 kDa Subunit of the Nucleosome Remodeling Factor.
- Tyagi, M., Imam, N., Verma, K., and Patel, A.K. (2016). Chromatin remodelers: We are the drivers!!. *Nucleus* 7, 388–404.
- Udugama, M., Sabri, A., and Bartholomew, B. (2011). The INO80 ATP-Dependent Chromatin Remodeling Complex Is a Nucleosome Spacing Factor. *Molecular and Cellular Biology*.
- Valverde, R., Edwards, L., and Regan, L. (2008). Structure and function of KH domains.
- Varga-Weisz, P.D., Wilm, M., Bonte, E., Dumas, K., Mann, M., and Becker, P.B. (1997). Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. *Nature* 388, 598–602.
- Wade, P.A., Jones, P.L., Vermaak, D., and Wolffe, A.P. (1998). A multiple subunit Mi-2 histone deacetylase from *Xenopus laevis* cofractionates with an associated Snf2 superfamily ATPase. *Current Biology*.
- Wahab, S., Saettone, A., Nabeel-Shah, S., Dannah, N., and Fillingham, J. (2020). Exploring the Histone Acetylation Cycle in the Protozoan Model *Tetrahymena thermophila*. *Frontiers in Cell and Developmental Biology*.
- Wallrath, L.L., and Elgin, S.C.R. (1995). Position effect variegation in *Drosophila* is associated with an altered chromatin structure. *Genes and Development*.
- Walport, L.J., Hopkinson, R.J., Chowdhury, R., Schiller, R., Ge, W., Kawamura, A., and Schofield, C.J. (2016). Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. *Nature Communications*.
- Wang, W., Huper, G., Guo, Y., Murphy, S.K., Olson, J.A., and Marks, J.R. (2005). Analysis of methylation-sensitive transcriptome identifies GADD45a as a frequently methylated gene in breast cancer. *Oncogene*.
- Wang, X., Goodrich, K.J., Gooding, A.R., Naeem, H., Archer, S., Paucek, R.D., Youmans, D.T., Cech, T.R., and Davidovich, C. (2017). Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines. *Molecular Cell*.
- Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q., et al. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genetics*.
- Watanabe, S., Tan, D., Lakshminarasimhan, M., Washburn, M.P., Hong, E.J.E., Walz, T., and Peterson, C.L. (2015). Structural analyses of the chromatin remodelling enzymes INO80-C and SWR-C. *Nature Communications*.
- Wiedemann, S.M., Mildner, S.N., Bönisch, C., Israel, L., Maiser, A., Matheisl, S., Straub, T., Merkl, R., Leonhardt, H., Kremmer, E., et al. (2010). Identification and characterization of two novel primate-specific histone H3 variants, H3.X and H3.Y. *Journal of Cell Biology*.
- Wiles, E.T., and Selker, E.U. (2017). H3K27 methylation: a promiscuous repressive chromatin mark.

- Wilkinson, B., Grepo, N., Thompson, B.L., Kim, J., Wang, K., Evgrafov, O.V., Lu, W., Knowles, J.A., and Campbell, D.B. (2015). The autism-associated gene chromodomain helicase DNA-binding protein 8 (CHD8) regulates noncoding RNAs and autism-related genes. *Translational Psychiatry*.
- Wilson, B.G., and Roberts, C.W.M. (2011). SWI/SNF nucleosome remodellers and cancer. *Nature Reviews Cancer* *11*, 481–492.
- Wilson, K.A., Holland, D.J., and Wetmore, S.D. (2016). Topology of RNA-protein nucleobase-amino acid  $\pi$ - $\pi$  interactions and comparison to analogous DNA-protein  $\pi$ - $\pi$  contacts. *RNA*.
- Wong, L.H., McGhie, J.D., Sim, M., Anderson, M.A., Ahn, S., Hannan, R.D., George, A.J., Morgan, K.A., Mann, J.R., and Choo, K.H.A. (2010). ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. *Genome Research*.
- Wood, A., Schneider, J., Dover, J., Johnston, M., and Shilatifard, A. (2003). The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. *Journal of Biological Chemistry*.
- Woodcock, C.L. (2006). Chromatin architecture (*Curr Opin Struct Biol*).
- Wu, L., and Winston, F. (1997). Evidence that Snf-Swi controls chromatin structure over both the TATA and UAS regions of the SUC2 promoter in *Saccharomyces cerevisiae*. *Nucleic Acids Research*.
- Xhemalce, B., Dawson, M.A., and Bannister, A.J. (2011). Histone Modifications. In *Encyclopedia of Molecular Cell Biology and Molecular Medicine*, (Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA), p.
- Xiang, S., Gapsys, V., Kim, H.Y., Bessonov, S., Hsiao, H.H., Möhlmann, S., Klaukien, V., Ficner, R., Becker, S., Urlaub, H., et al. (2013). Phosphorylation drives a dynamic switch in serine/arginine-rich proteins. *Structure*.
- Xiao, B., Freedman, B.S., Miller, K.E., Heald, R., and Marko, J.F. (2012). Histone H1 compacts DNA under force and during chromatin assembly. *Molecular Biology of the Cell*.
- Xiong, C., Wen, Z., Yu, J., Chen, J., Liu, C.P., Zhang, X., Chen, P., Xu, R.M., and Li, G. (2018). UBN1/2 of HIRA complex is responsible for recognition and deposition of H3.3 at cis-regulatory elements of genes in mouse ES cells. *BMC Biology* *16*, 110.
- Xue, Y., Wong, J., Moreno, G.T., Young, M.K., Côté, J., and Wang, W. (1998). NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Molecular Cell*.
- Xue, Y., Zhou, Y., Wu, T., Zhu, T., Ji, X., Kwon, Y.S., Zhang, C., Yeo, G., Black, D.L., Sun, H., et al. (2009). Genome-wide Analysis of PTB-RNA Interactions Reveals a Strategy Used by the General Splicing Repressor to Modulate Exon Inclusion or Skipping. *Molecular Cell*.
- Yan, J., Dutta, B., Hee, Y.T., and Chng, W.J. (2019a). Towards understanding of PRC2 binding to RNA.

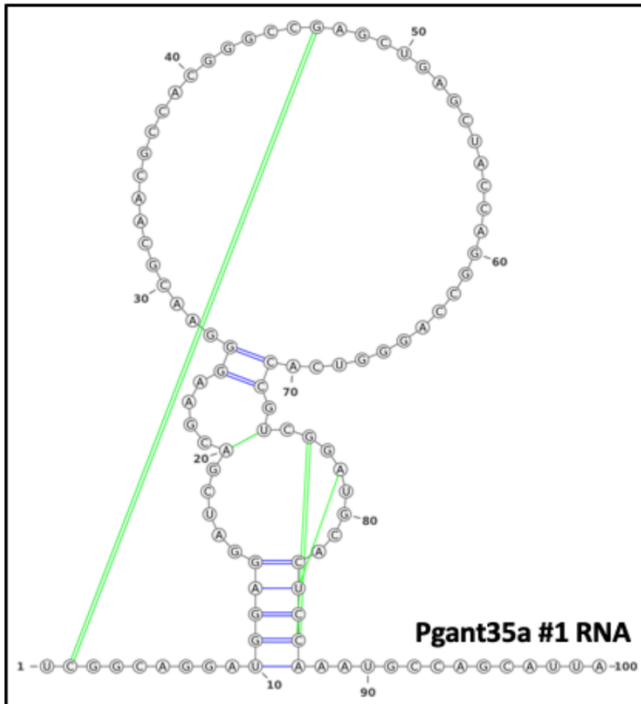
- Yan, L., Wu, H., Li, X., Gao, N., and Chen, Z. (2019b). Structures of the ISWI–nucleosome complex reveal a conserved mechanism of chromatin remodeling. *Nature Structural and Molecular Biology*.
- Yang, X.J., and Seto, E. (2007). HATs and HDACs: From structure, function and regulation to novel strategies for therapy and prevention.
- Yang, X.J., and Seto, E. (2008). The Rpd3/Hda1 family of lysine deacetylases: From bacteria and yeast to mice and men.
- Yang, L., Wang, C., Li, F., Zhang, J., Nayab, A., Wu, J., Shi, Y., and Gong, Q. (2017). The human RNA-binding protein and E3 ligase MEX-3C binds the MEX-3–recognition element (MRE) motif with high affinity. *Journal of Biological Chemistry*.
- Zeng, X., Lin, X., and Hou, S.X. (2013). The Osa-containing SWI/SNF chromatin-remodeling complex regulates stem cell commitment in the adult *Drosophila* intestine. *Development (Cambridge)*.
- Zhang, H., Bishop, B., Ringenberg, W., Muir, W.M., and Ogas, J. (2012). The CHD3 remodeler PICKLE associates with genes enriched for trimethylation of histone H3 lysine 27. *Plant Physiology*.
- Zhang, Q., McKenzie, N.J., Warneford-Thomson, R., Gail, E.H., Flanigan, S.F., Owen, B.M., Lauman, R., Levina, V., Garcia, B.A., Schittenhelm, R.B., et al. (2019). RNA exploits an exposed regulatory site to inhibit the enzymatic activity of PRC2. *Nature Structural and Molecular Biology*.
- Zhang, W., Aubert, A., Gomez de Segura, J.M., Karuppasamy, M., Basu, S., Murthy, A.S., Diamante, A., Drury, T.A., Balmer, J., Cramard, J., et al. (2016). The Nucleosome Remodeling and Deacetylase Complex NuRD Is Built from Preformed Catalytically Active Sub-modules. *Journal of Molecular Biology*.
- Zhang, Y., LeRoy, G., Seelig, H.P., Lane, W.S., and Reinberg, D. (1998). The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell*.
- Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes and Development*.
- Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science*.
- Zhao, Z., Sentürk, N., Song, C., and Grummt, I. (2018). lncRNA PAPAS tethered to the rDNA enhancer recruits hypophosphorylated CHD4/NuRD to repress rRNA synthesis at elevated temperatures. *Genes and Development*.
- Zhou, C.Y., Johnson, S.L., Gamarra, N.I., and Narlikar, G.J. (2016). Mechanisms of ATP-Dependent Chromatin Remodeling Motors. *Annual Review of Biophysics* 45, 153–181.

# 7 Appendix

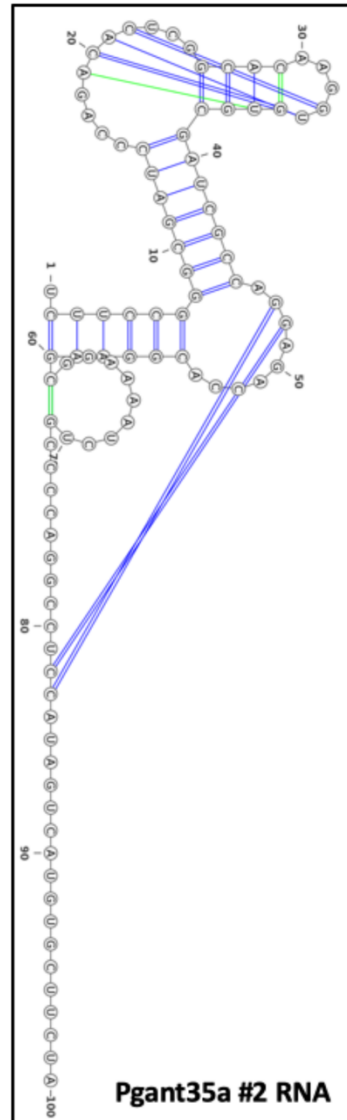
## Supplementary data



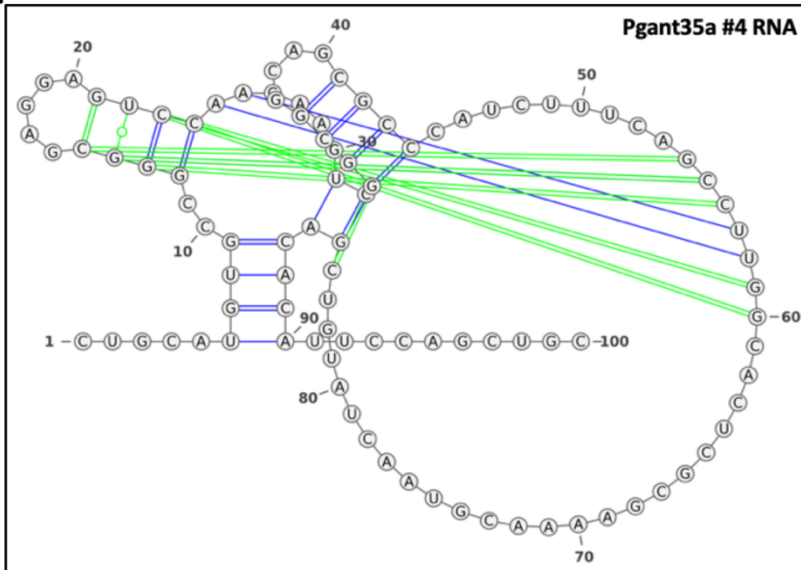
E



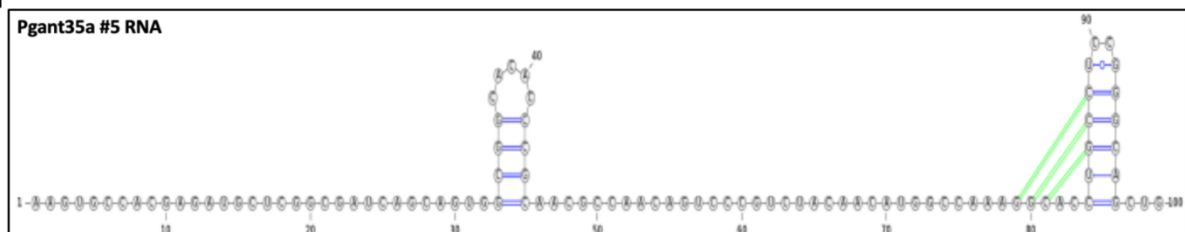
F



G

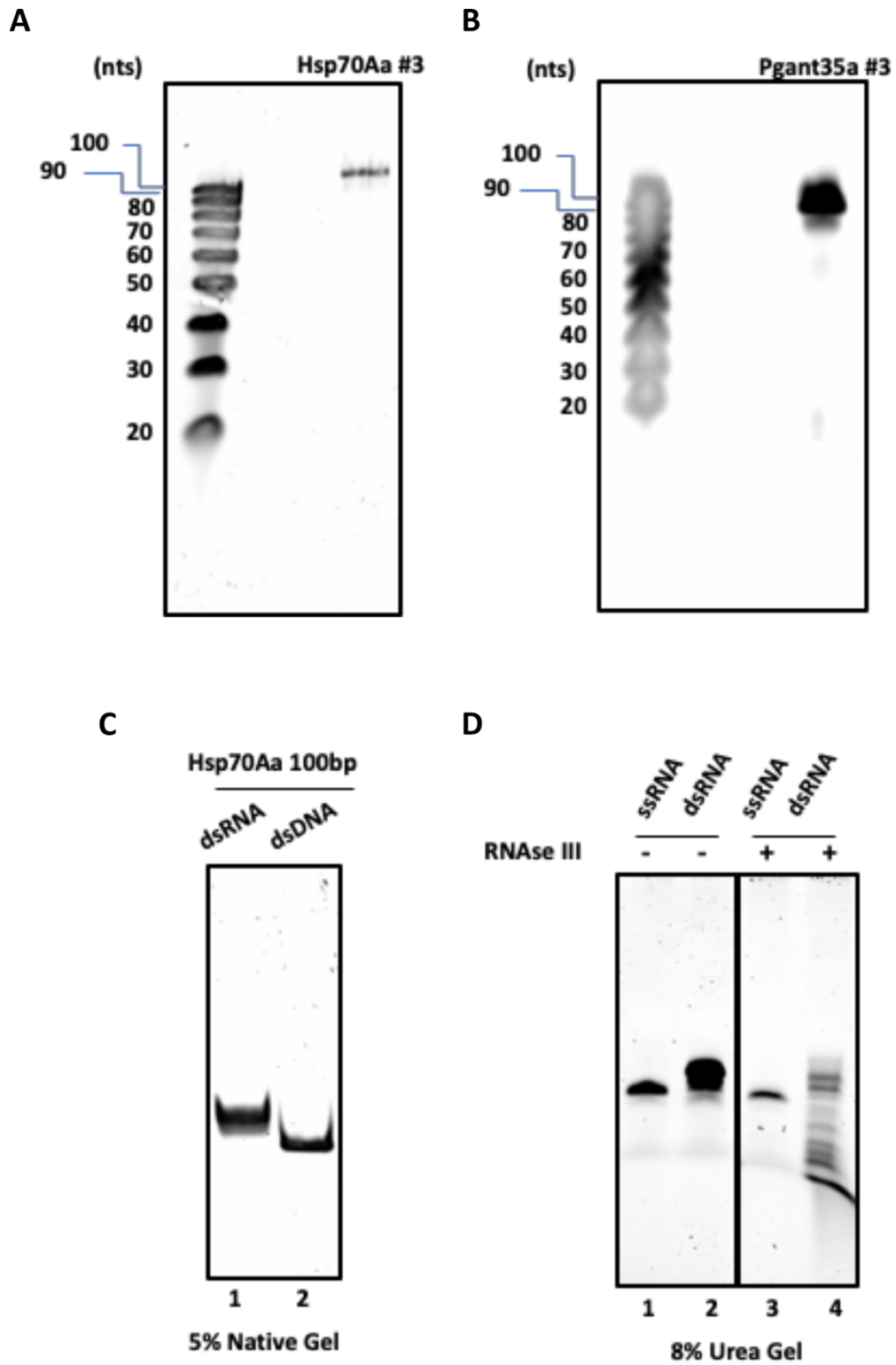


H



Supplementary Figure S3.1.2

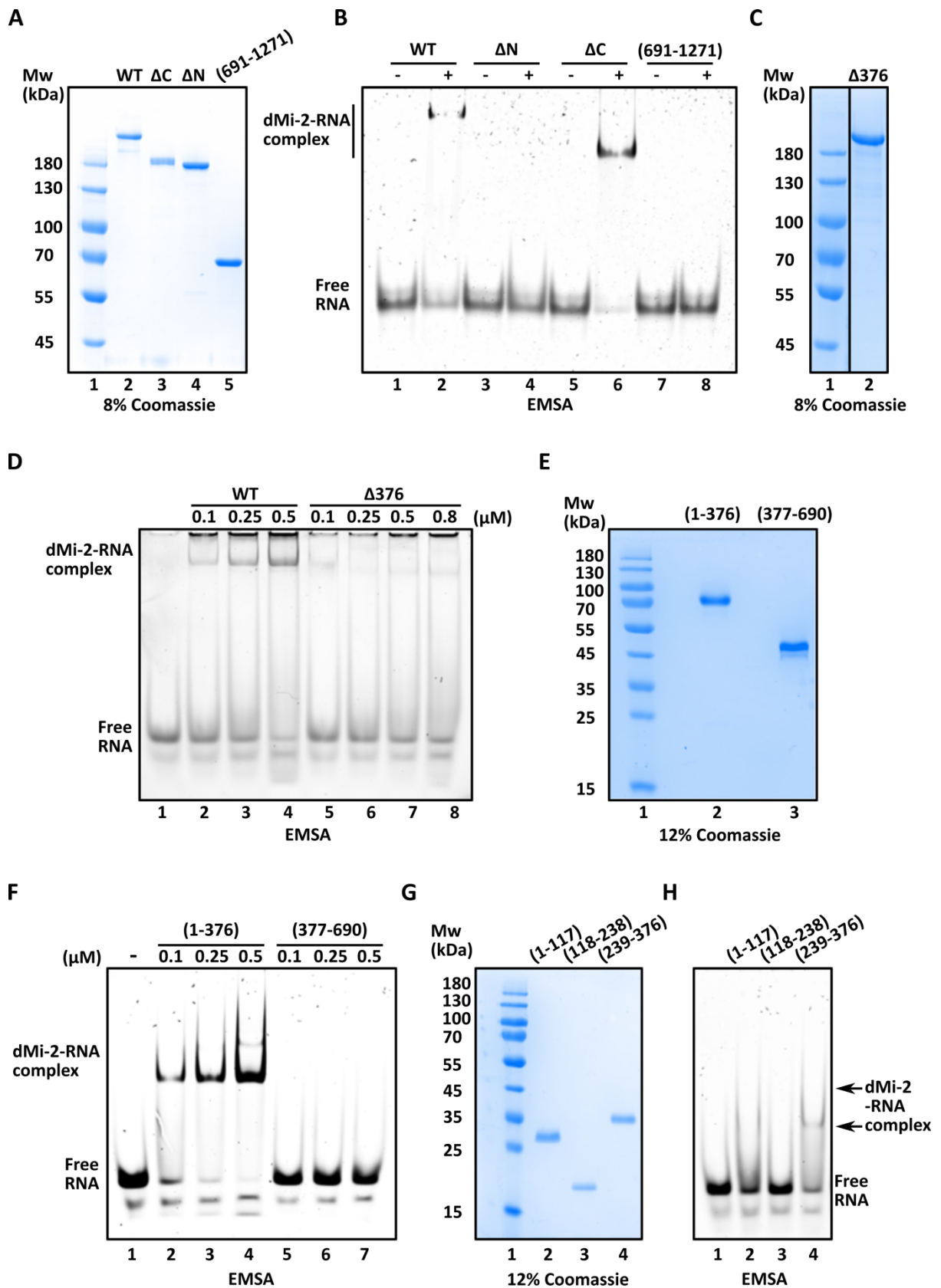
Predicted secondary structures of the RNAs (A-H). The structures were predicted with SPOT-RNA: RNA Secondary Structure Prediction Software with optimal default settings. The canonical base pairs predicted are in blue and non-canonical lone pairs and triplets in Green.



Supplementary Figure S3.1.3:

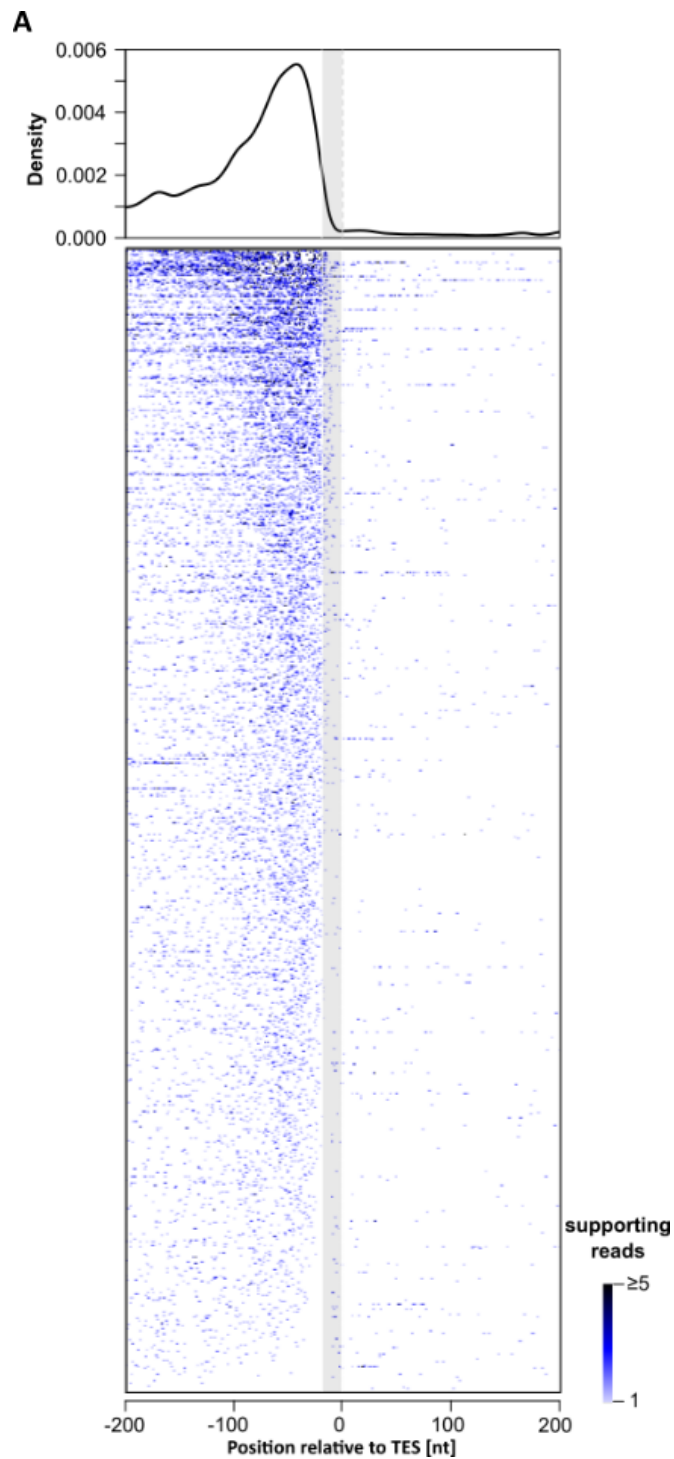
#3 RNAs (100-mer single stranded RNAs) of Hsp70Aa and Pgant35a along with single stranded RNA size marker (A and B) are loaded on a urea denaturing polyacrylamide gel and electrophoresed. C) 100 base pair (bp) double stranded RNA and DNA generated from Hsp70Aa gene (Lanes 1 and 2, respectively) were loaded on a native polyacrylamide gel and electrophoresed D) ssRNA and dsRNA were untreated (Lane 1 and 2, respectively) or treated with RNase III (Lane 3 and 4, respectively). Each gel was stained with 1X SYBR gold and imaged with Bio-Rad Image Lab.





**Supplementary Figure S3.1.5:**

N-terminally FLAG-tagged dMi-2 Wild type (WT) and its truncation mutants were purified by FLAG immunopurification and eluted with FLAG peptide. The protein samples were loaded on SDS gels in **(A)**, **(C)**, **(E)** and **(G)** and electrophoresed. Molecular weight markers were loaded in lane 1 for each. The gels were stained with Coomassie Brilliant Blue and imaged with Bio-Rad Image Lab. EMSAs of dMi-2 WT and its truncation mutants in **(B)**, **(D)**, **(F)** and **(H)**. In **(B)**, 50 nanograms of Hsp70Aa #3 RNA (Figure 3A, schematics) was used for each reaction and 173 nanograms each of dMi-2 WT (lane 2), dMi-2  $\Delta$ N (lane 4), dMi-2  $\Delta$ C (lane 6) and dMi-2 (691-1271) (lane 8). In **(D)**, 250 nanomolar Hsp70Aa #6 RNA was used for each reaction with titrations of dMi-2 WT or dMi-2  $\Delta$ 376, as indicated. In **(F)**, 250 nanomolar Hsp70Aa #6 RNA was used in each reaction and dMi-2 (1-376), lanes 2-4 or dMi-2 (377-690), lanes 5-7, were titrated from 0.5 to 2 micromolar ( $\mu$ M). In **(H)**, 250 nanomolar Hsp70Aa #6 RNA was used in each reaction and 1.5 micromolar ( $\mu$ M) protein each from dMi-2 (1-117) lane 2, (118-238) lane 3 and (239-376) lane 4. RNA only was loaded for each EMSA in lane 1. EMSA reactions were loaded on native polyacrylamide gels and electrophoresed. The gels were stained with 1X SYBR gold and imaged with Bio-Rad Image Lab.



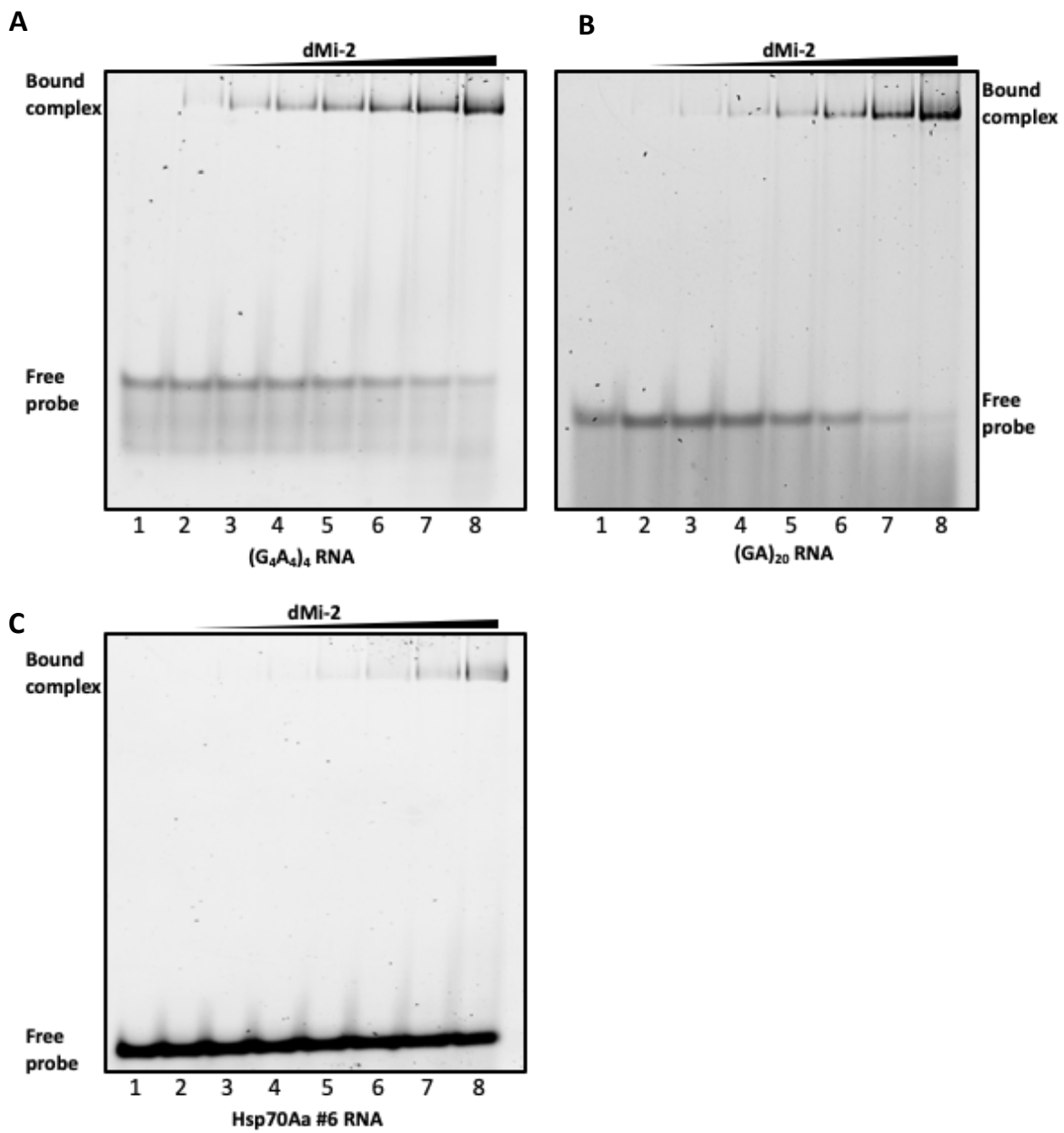
**Supplementary Figure S3.2.5:**

(A) Genes with iCLIP reads (from both replicates) are aligned at the transcription end site (TES = 0). In blue, cross-linking sites are indicated. On top, the cumulative cross-linking density for all genes is plotted. The grey shaded area from position -19 to -1 indicates a part of transcript where reads are extremely unlikely to be mapped due to the minimum length of trimmed reads of 18 nucleotides. Cross-links in this zone are mostly due to partial alignments

containing a poly(A) signal, very rarely read-through events or false positive mappings. Due to this technical constraint, iCLIP is not able to predict exact cross-linking positions in these last nucleotides. The colour code on the right indicates number of supporting reads for each cross-linking site from 1 to  $\geq 5$ .



of the nucleosomal DNA denoted in blue). 50 nanomolar (nM) dMi-2 was added in each reaction (plus sign), except in controls (minus sign on the top). Remodelled nucleosome (represented on the left, orange denotes histone octamer positioned centrally on the nucleosomal DNA denoted in blue) can be observed in lanes 2-6 of (A) and (B) and lanes 2-10 in (C). Free RNA in each gel can be observed at the bottom. Reactions were loaded on native polyacrylamide gels, electrophoresed, stained with SYBR gold and imaged with Bio-Rad ChemiDoc system. Red arrows point out the remodelled nucleosome band and red asterisk marks in lanes 7-10 of (A) and (B) point out the inhibition of remodelling. Blue asterisk in lane 10 of (C) points out remodelled nucleosome even at the highest concentration of Hsp70Aa #6 RNA. Representative images of at least three independent experiments.

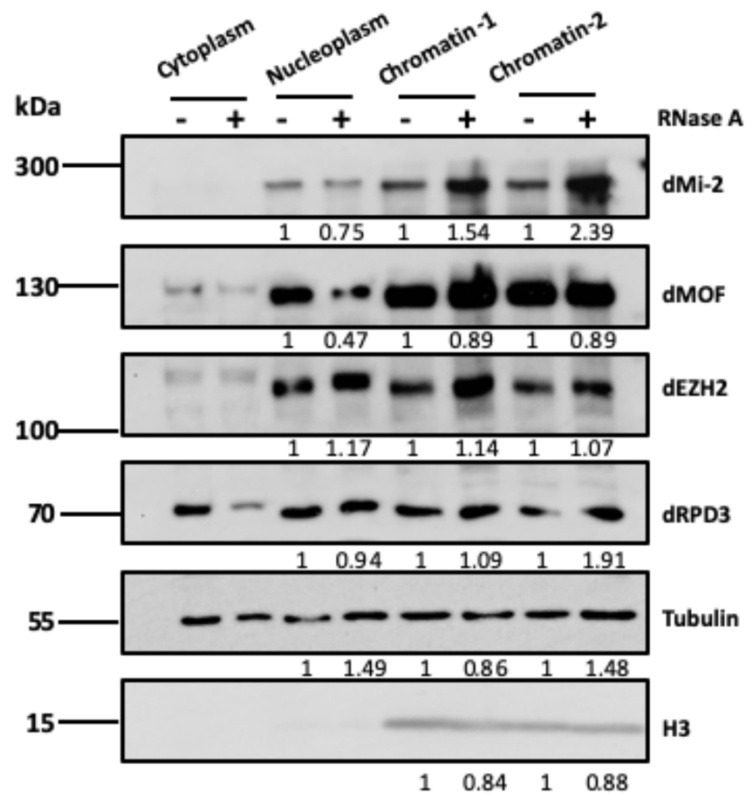


**Supplementary figure S3.3.2: RNAs that bind efficiently to dMi-2 inhibit remodelling**

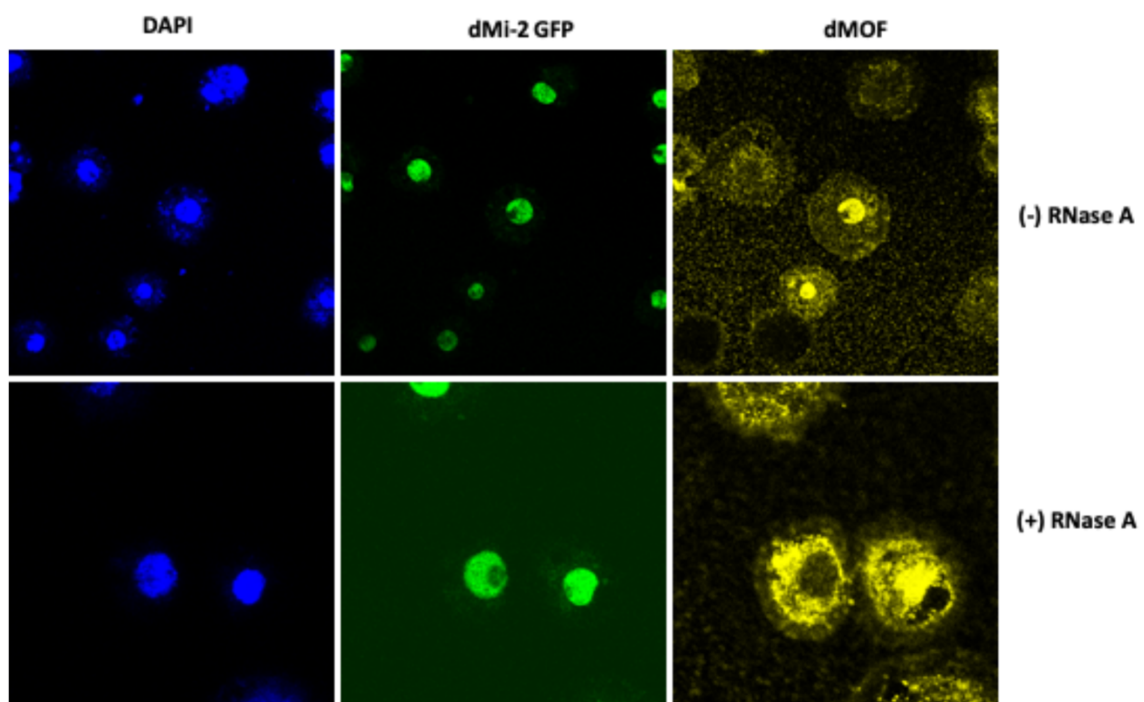
Electrophoretic mobility shift assays (EMSA) of **(A)**  $(G_4A_4)_4$  RNA, **(B)**  $(GA)_{20}$  RNA and **(C)** Hsp70Aa #6 RNA. 250 nanomolar (nM) RNA was used in each reaction and dMi-2 titrated from 10 to 640 nanomolar (nM), lanes 2-8. Each reaction was loaded on a native polyacrylamide gel and electrophoresed. The gels were stained with SYBR gold and imaged with Bio-Rad ChemiDoc system. dMi-2 bound to RNA is denoted as bound complex. Free RNA is also denoted. Representative images of at least two independent experiments.



A



B



### Supplementary figure S3.3.3: Loss of RNA causes increased association of dMi-2 with chromatin

RNase A treatment and western blot (A)  $3.5 \times 10^7$  cells were permeabilized by treatment with 0.05% of Tween-20 for 10 minutes at 4°C. Cells were treated with 10 milligrams per millilitre

(mg/ml) of RNase A. Control cells were similarly permeabilized but treated with equal amount of BSA (bovine serum albumin). Cells were fractionated in cytoplasmic, nuclear and chromatin fraction. Each fraction was loaded on an 8-12% gradient denaturing polyacrylamide gel and electrophoresed. Western blot analysis was performed for the proteins indicated. Histone H3 was used as loading control. The western blot signals were quantified using ImageJ software. The BSA treated control signals were set to 1 and the fold change in the treated samples were calculated and numbers indicated below. **(B)**  $1 \times 10^6$  cells per millilitre each from dMi-2-GFP and S2 cells were seeded on concanavalin A coated glass coverslips, permeabilized by treatment with 0.05 micrograms per microliter ( $\mu\text{g}/\mu\text{l}$ ) of digitonin for 10 minutes. One set was treated with RNase A and another mock treated with PBS alone and cells were fixed. S2 cells were blocked with BSA, incubated with anti-MOF primary antibody and Alexa-547 secondary antibody. Fixed cells were stained with DAPI and Phalloidin stains just before mounting on a glass slide using mowiol as mounting media. Cells were visualized under a confocal microscope and imaged. Representative images of several independent experiments (This figure will be replaced with the better ones).

Table 3: dMi-2-GFP iCLIP-seq list of enriched mRNAs with Fold change > 4 and ( $p < 0.05$ )

FlyBase Gene Report	name	log2FC
FBgn0053868	His2B:CG33868	10.8002796376212
FBgn0052580	Muc14A	9.45703266969185
FBgn0052320	CG32320	7.98431559774432
FBgn0037811	CG12592	7.34205351830939
FBgn0261836	Msp300	6.90512108738892
FBgn0033141	CG12831	6.83376678488231
FBgn0259110	mmd	6.48009829238807
FBgn0034126	jtb	6.31564321070758
FBgn0003016	osp	6.15032241570444
FBgn0267033	mamo	6.13947123092771
FBgn0013275	Hsp70Aa	6.01423597289303
FBgn0042131	CG18808	5.82723197292279
FBgn0285926	Imp	5.51261175360771
FBgn0030914	CG6106	5.35333678436727
FBgn0042627	FASN2	5.34698447517476
FBgn0032683	kon	5.33454638168472
FBgn0026206	mei-P26	5.28106560202889
FBgn0016977	spen	5.12465549706164
FBgn0265487	mbl	5.09354674179842
FBgn0284408	trol	5.02973784581865
FBgn0016976	stnA	5.02202858797867
FBgn0016975	stnB	5.02202858797867
FBgn0034140	Lst	5.01429266653495
FBgn0264490	Eip93F	5.01429266653495
FBgn0010113	hdc	4.82922645855303
FBgn0038826	Syp	4.80554742361066
FBgn0004652	fru	4.73165515675901
FBgn0040765	luna	4.70726440615755
FBgn0053554	Nipped-A	4.62241953495861
FBgn0003165	pum	4.58493032464889
FBgn0266696	Svil	4.56984780051999

FBgn0261552	ps	4.54412038731827
FBgn0086758	chinmo	4.50048111975893
FBgn0052113	Vps13D	4.4066220711289
FBgn0013733	shot	4.3883913027958
FBgn0027339	jim	4.38457515666963
FBgn0004656	fs(1)h	4.3649173105306
FBgn0032136	Apoltp	4.36486760035777
FBgn0283451	br	4.33824319934438
FBgn0033661	CG13185	4.32920287048232
FBgn0011230	poe	4.28356352372469
FBgn0261574	kug	4.2827128208743
FBgn0262656	Myc	4.27141752469343
FBgn0260400	elav	4.23457122505278
FBgn0000479	dnc	4.20062776178223
FBgn0265523	Smr	4.18250497919764
FBgn0036851	CG14082	4.17569410266339
FBgn0052296	Mrtf	4.16936958085499
FBgn0025741	PlexA	4.12430657847216
FBgn0010300	brat	4.07380164125748
FBgn0036398	upSET	4.06134713863072
FBgn0261793	Trf2	4.0403012026309
FBgn0261797	Dhc64C	4.03134331769893
FBgn0265045	Strn-Mlck	3.96752078846758
FBgn0039959	CG17514	3.95073446887442
FBgn0264975	Nrg	3.94381527946911
FBgn0030674	CG8184	3.90684931717943
FBgn0283521	lola	3.88370229293706
FBgn0035142	Hipk	3.88267307312318
FBgn0085430	CG34401	3.88106441668734
FBgn0261617	nej	3.87889444131911
FBgn0052767	CG32767	3.877811390303
FBgn0011666	msi	3.85600196686996
FBgn0260003	Dys	3.8502939063812
FBgn0283499	InR	3.84300467611203

FBgn0263396	sqd	3.84026621842002
FBgn0259170	alpha-Man-la	3.8302661807562
FBgn0259789	zld	3.82192351680268
FBgn0266557	kis	3.82151639555282
FBgn0031107	HERC2	3.81224333164984
FBgn0002526	LanA	3.805099444757
FBgn0035424	Larp4B	3.8021684823626
FBgn0085407	Pvf3	3.78230261448468
FBgn0026869	Thd1	3.76329605650254
FBgn0011224	heph	3.75833576673696
FBgn0085446	CG34417	3.74416715758126
FBgn0004435	Galphaq	3.73700919148138
FBgn0259176	bun	3.7325444173771
FBgn0052479	Usp10	3.7106283740648
FBgn0029979	mahe	3.70083745707653
FBgn0052529	Hers	3.69450431101918
FBgn0028397	Tob	3.68980501695715
FBgn0016694	Pdp1	3.68476402351277
FBgn0285917	sbb	3.67453804230069
FBgn0261802	CG42748	3.67074461112003
FBgn0043884	mask	3.66411181002783
FBgn0000546	EcR	3.66042646641071
FBgn0050069	CG30069	3.64908467905196
FBgn0040297	Nhe2	3.64055360918131
FBgn0010905	Spn	3.6394433747121
FBgn0263352	Unr	3.63765257248846
FBgn0000568	Eip75B	3.63389715711136
FBgn0040752	Prosap	3.63019268796045
FBgn0002945	nkd	3.62847814774335
FBgn0267912	CanA-14F	3.62012387182123
FBgn0053100	eIF4EHP	3.61374655497384
FBgn0085478	CG34449	3.59778213391869
FBgn0000567	Eip74EF	3.58569829686856
FBgn0032600	BuGZ	3.58098124511512

<b>FBgn0004607</b>	zfh2	3.5682779897614
<b>FBgn0003862</b>	trx	3.56813599914016
<b>FBgn0263289</b>	scrib	3.56332797796981
<b>FBgn0264707</b>	RhoGEF3	3.55815203137006
<b>FBgn0283657</b>	Tlk	3.55695406810382
<b>FBgn0267336</b>	Glut4EF	3.53084643846658
<b>FBgn0023479</b>	teq	3.52818050338941
<b>FBgn0036165</b>	chrb	3.52163484398033
<b>FBgn0260634</b>	eIF4G2	3.52128170385548
<b>FBgn0286778</b>	CG46385	3.5142655680346
<b>FBgn0259745</b>	wech	3.49010643440631
<b>FBgn0029114</b>	Tollo	3.48130830860426
<b>FBgn0086674</b>	Tpst	3.47650915420401
<b>FBgn0004198</b>	ct	3.47225869320717
<b>FBgn0051998</b>	CG31998	3.46989323416501
<b>FBgn0053087</b>	LRP1	3.45453633687723
<b>FBgn0262160</b>	CG9932	3.45092768873571
<b>FBgn0015269</b>	Nf1	3.44876936407616
<b>FBgn0001122</b>	Galphao	3.44845291492252
<b>FBgn0266717</b>	Bruce	3.44492005125665
<b>FBgn0250754</b>	CG42232	3.44320665275096
<b>FBgn0052423</b>	shep	3.44319538672698
<b>FBgn0033159</b>	Dscam1	3.44183985016598
<b>FBgn0261934</b>	dikar	3.43707054163435
<b>FBgn0052133</b>	Ptip	3.42728351637155
<b>FBgn0039920</b>	CG11360	3.42007160617459
<b>FBgn0039928</b>	Cals	3.4145470485376
<b>FBgn0000635</b>	Fas2	3.40403417781517
<b>FBgn0266801</b>	CG45263	3.39219887597165
<b>FBgn0002643</b>	mam	3.38968029893869
<b>FBgn0026086</b>	Adar	3.38883409311273
<b>FBgn0036004</b>	Jarid2	3.38308793052713
<b>FBgn0004395</b>	unk	3.38295557036386
<b>FBgn0261671</b>	tweek	3.38226224328963

FBgn0085434	NaCP60E	3.38199769231437
FBgn0261549	rdgA	3.38199769231437
FBgn0003415	skd	3.37900372704959
FBgn0041092	tai	3.37010573765372
FBgn0033010	Atf6	3.35706130141735
FBgn0014388	sty	3.35458588850229
FBgn0286516	aqz	3.35292354750985
FBgn0004167	kst	3.35050831229279
FBgn0004449	Ten-m	3.34606086706945
FBgn0261822	Bsg	3.34348060293133
FBgn0260794	ctrip	3.34281364883831
FBgn0025740	PlexB	3.33426992284789
FBgn0264953	Piezo	3.33052226811085
FBgn0262737	mub	3.32572560063108
FBgn0003391	shg	3.32448490211995
FBgn0029504	CHES-1-like	3.30768140409356
FBgn0263490	mld	3.30414734457373
FBgn0002673	twe	3.30305909078457
FBgn0031698	Ncoa6	3.29128342704834
FBgn0021764	sdk	3.28901098269
FBgn0086686	l(3)L1231	3.2646294718284
FBgn0001078	ftz-f1	3.26417477578184
FBgn0259735	mtgo	3.2561152947444
FBgn0039633	CG11873	3.25499268303909
FBgn0052062	Rbfox1	3.24925612910232
FBgn0262739	AGO1	3.24587077721261
FBgn0000163	baz	3.24020373967179
FBgn0262743	Fs(2)Ket	3.23950100186323
FBgn0040388	boi	3.23656496660836
FBgn0014163	fax	3.23627305400339
FBgn0039209	REPTOR	3.22224426178399
FBgn0004396	CrebA	3.22144748849293
FBgn0033636	tou	3.21037135373029
FBgn0039831	CG12054	3.20807220954901

<b>FBgn0263929</b>	jvl	3.20234815023948
<b>FBgn0029676</b>	HIP-R	3.19393183019868
<b>FBgn0262975</b>	cnc	3.18485371316302
<b>FBgn0265623</b>	Su(z)2	3.1733913193136
<b>FBgn0039727</b>	Vps13B	3.16963220685496
<b>FBgn0259168</b>	mnb	3.16503976795187
<b>FBgn0085436</b>	Not1	3.1527052537297
<b>FBgn0000541</b>	E(bx)	3.13447764610696
<b>FBgn0261244</b>	inaE	3.13102301524477
<b>FBgn0261811</b>	pico	3.12093344475092
<b>FBgn0032633</b>	Lrch	3.11701059560866
<b>FBgn0003301</b>	rut	3.10769586102493
<b>FBgn0263706</b>	CG43658	3.10471562946762
<b>FBgn0030243</b>	CG2186	3.10095665296189
<b>FBgn0003525</b>	stg	3.09940269412615
<b>FBgn0260965</b>	CG42588	3.09869818486504
<b>FBgn0262582</b>	cic	3.09322351281376
<b>FBgn0020306</b>	dom	3.09115766849551
<b>FBgn0030266</b>	CG11122	3.08189061616957
<b>FBgn0039590</b>	CG10011	3.08029928355784
<b>FBgn0085423</b>	GramD1B	3.07473982893751
<b>FBgn0003137</b>	Ppn	3.0723859502579
<b>FBgn0264307</b>	orb2	3.06586608908757
<b>FBgn0263346</b>	smash	3.06506306903183
<b>FBgn0025936</b>	Eph	3.04927954974251
<b>FBgn0004227</b>	nonA	3.03510370990698
<b>FBgn0260486</b>	Ziz	3.03464784352583
<b>FBgn0264324</b>	spg	3.03091887610756
<b>FBgn0003048</b>	pcx	3.03042550216857
<b>FBgn0266580</b>	Gp210	3.0175721334852
<b>FBgn0026083</b>	tyf	3.01615631438175
<b>FBgn0052000</b>	anne	3.01385864470053
<b>FBgn0011817</b>	nmo	3.0126332727389
<b>FBgn0036735</b>	Edc3	3.00039518386434



FBgn0262614	pyd	2.99965147976344
FBgn0003870	ttk	2.99758069839968
FBgn0036374	Spt20	2.99248434088424
FBgn0011656	Mef2	2.98779671394399
FBgn0002645	Map205	2.98167777396056
FBgn0259214	PMCA	2.9809277190444
FBgn0003371	sgg	2.97996051170067
FBgn0011837	Tis11	2.97487211307862
FBgn0026427	Su(var)2-HP2	2.97432374037429
FBgn0031738	CG9171	2.97369016149345
FBgn0261854	aPKC	2.96540913913289
FBgn0032817	CG10631	2.95539765239398
FBgn0031118	RhoGAP19D	2.95527648174142
FBgn0004876	cdi	2.94939203997566
FBgn0039466	CG5521	2.94410455065222
FBgn0000382	csw	2.93670978762071
FBgn0036446	CG9384	2.93605771520376
FBgn0260934	par-1	2.93306625401769
FBgn0030766	mthl1	2.93061247024479
FBgn0263987	spoon	2.92717972704771
FBgn0020224	Cbl	2.91662054381818
FBgn0036746	Crtc	2.90537352750203
FBgn0035106	rno	2.90460279384508
FBgn0010051	ltp-r83A	2.90273704447419
FBgn0086655	jing	2.90201761594318
FBgn0022764	Sin3A	2.8861920193383
FBgn0261444	CG3638	2.8827488548919
FBgn0261285	Ppcs	2.87971687010065
FBgn0262730	dtn	2.87812818239281
FBgn0001297	kay	2.86859479846172
FBgn0262617	Nuak1	2.86490291641935
FBgn0263006	SERCA	2.86423609190985
FBgn0051635	CG31635	2.86403681268674
FBgn0259212	cno	2.86154470129852

FBgn0005624	Psc	2.85883324097677
FBgn0039214	puf	2.8502109345035
FBgn0053208	Mical	2.84891808140212
FBgn0003028	ovo	2.83065947875137
FBgn0025726	unc-13	2.82077350836129
FBgn0002781	mod(mdg4)	2.81813406152301
FBgn0003345	sd	2.81601978946513
FBgn0029903	pod1	2.81547257223379
FBgn0085370	Pde11	2.81536571581785
FBgn0085412	CG34383	2.80792030918454
FBgn0015589	Apc	2.80451313895737
FBgn0259171	Pde9	2.79648308869151
FBgn0265998	Doa	2.79605604571792
FBgn0030512	Nadsyn	2.78841435848043
FBgn0034570	CG10543	2.78162934907266
FBgn0260442	rhea	2.77761946673531
FBgn0005631	robo1	2.7769636097824
FBgn0264493	rdx	2.77626319561565
FBgn0264495	gpp	2.77403839959632
FBgn0036518	RhoGAP71E	2.76869941727135
FBgn0051224	CG31224	2.76448256044201
FBgn0000581	E(Pc)	2.76283691839791
FBgn0034240	MESR4	2.76240177983413
FBgn0263993	CG43736	2.75994931351646
FBgn0001169	H	2.75265183573361
FBgn0263392	Tet	2.74889286482095
FBgn0036844	Mkp3	2.74178206088994
FBgn0262579	Ect4	2.74164315594264
FBgn0036576	CG5151	2.74139315638568
FBgn0262103	Sik3	2.73460943545281
FBgn0024921	Tnpo	2.73401091114471
FBgn0261397	didum	2.73136070132917
FBgn0010382	CycE	2.72974163868659
FBgn0003507	srp	2.72093330075867

<b>FBgn0286831</b>	sov	2.71996880851096
<b>FBgn0024734</b>	PRL-1	2.71857630468388
<b>FBgn0004370</b>	Ptp10D	2.71685556126891
<b>FBgn0004838</b>	Hrb27C	2.71502169320622
<b>FBgn0037705</b>	mura	2.71442075186685
<b>FBgn0003502</b>	Btk29A	2.70728325553527
<b>FBgn0259984</b>	kuz	2.70218573454958
<b>FBgn0037902</b>	CG5281	2.6960386883981
<b>FBgn0265988</b>	mv	2.69549738552988
<b>FBgn0020412</b>	JIL-1	2.693800416969
<b>FBgn0285955</b>	cv-c	2.6917436091765
<b>FBgn0036734</b>	CG7564	2.69087996488417
<b>FBgn0040324</b>	Ephrin	2.6896366929364
<b>FBgn0023458</b>	Rbcn-3A	2.6869344999041
<b>FBgn0266019</b>	rudhira	2.6860473218105
<b>FBgn0267861</b>	Maf1	2.68523612828502
<b>FBgn0016070</b>	smg	2.68350823790166
<b>FBgn0000119</b>	arr	2.68297840408335
<b>FBgn0061200</b>	Nup153	2.6738236003526
<b>FBgn0000578</b>	ena	2.6667129227645
<b>FBgn0260970</b>	Ubr3	2.66594800603171
<b>FBgn0266084</b>	Fhos	2.66511672835558
<b>FBgn0020251</b>	sfl	2.66403695852226
<b>FBgn0263102</b>	psq	2.6584051364151
<b>FBgn0011818</b>	oaf	2.65667401747727
<b>FBgn0024555</b>	ffl	2.65545777199525
<b>FBgn0031150</b>	bves	2.652465254001
<b>FBgn0086690</b>	Plp	2.65177408603601
<b>FBgn0034734</b>	CG4554	2.64743996853497
<b>FBgn0259927</b>	CG42450	2.64512369157988
<b>FBgn0000247</b>	ca	2.63746346130591
<b>FBgn0023407</b>	B4	2.63696519188326
<b>FBgn0003483</b>	spn-E	2.63650998004523
<b>FBgn0265974</b>	ttv	2.63536274112336

FBgn0026179	siz	2.6305123740467
FBgn0263131	CG43373	2.62989293314375
FBgn0266410	CG45050	2.62362727460374
FBgn0278608	Dsp1	2.62117296723614
FBgn0052654	Sec16	2.61928641494117
FBgn0030049	Trf4-1	2.6155556000095
FBgn0003416	sl	2.61534764521833
FBgn0000464	Lar	2.61092808278158
FBgn0003117	pnr	2.60859696767451
FBgn0004648	svr	2.60239052390867
FBgn0261710	nocte	2.59794022696811
FBgn0023213	eIF4G1	2.59232114877232
FBgn0085450	Snoo	2.59080276135639
FBgn0262733	Src64B	2.59059455650457
FBgn0041585	olf186-F	2.59048197961029
FBgn0031981	Megf8	2.58720184872033
FBgn0039419	CG12290	2.58600056303399
FBgn0010825	Gug	2.58300976952829
FBgn0265630	sno	2.58110539895081
FBgn0038197	foxo	2.58101111516715
FBgn0002905	mus308	2.57910465099039
FBgn0051320	HEATR2	2.57883017429885
FBgn0015778	rin	2.56931396354219
FBgn0005640	Eip63E	2.5585194868648
FBgn0285892	tea	2.55783249557159
FBgn0015396	jumu	2.55534492596839
FBgn0001624	dlg1	2.55505073146627
FBgn0013343	Syx1A	2.5546444663924
FBgn0028863	CG4587	2.55362847357971
FBgn0004509	Fur1	2.55274572070552
FBgn0000017	Abl	2.54819289960723
FBgn0000719	fog	2.5471645343893
FBgn0003256	rl	2.54650205299112
FBgn0027492	wdb	2.54479447769425

FBgn0026376	Rgl	2.5408249384042
FBgn0023097	bon	2.53974131454693
FBgn0030505	NFAT	2.53535444246016
FBgn0020309	crol	2.53285460792943
FBgn0004875	enc	2.53037984556701
FBgn0011826	Pp2B-14D	2.52994717269263
FBgn0004369	Ptp99A	2.52718960839937
FBgn0011274	Dif	2.52622928276393
FBgn0086779	step	2.52336323716752
FBgn0039923	MED26	2.50990106139535
FBgn0039904	Hcf	2.50706206838391
FBgn0026401	Nipped-B	2.50278175970482
FBgn0000097	aop	2.50208135258897
FBgn0024277	trio	2.49594705243386
FBgn0087008	e(y)3	2.49484920682123
FBgn0013263	Trl	2.47943508634996
FBgn0260941	app	2.47812516625264
FBgn0032821	CdGAPr	2.47595732733249
FBgn0261556	CG42674	2.47264191444187
FBgn0026575	hang	2.46899307225253
FBgn0031126	Cyp6v1	2.46406990187567
FBgn0262473	Tl	2.46093927376994
FBgn0033194	Vps13	2.45921579032316
FBgn0015772	Nak	2.45919160179705
FBgn0263395	hppy	2.45772707196989
FBgn0032395	Wdr81	2.45759753557472
FBgn0261545	CG42663	2.45542236011173
FBgn0001105	Gbeta13F	2.4517895390429
FBgn0035400	CG11537	2.44217333855291
FBgn0036451	CG9425	2.44206538210106
FBgn0036522	CG7372	2.44041972376882
FBgn0023510	Rbcn-3B	2.44020782129427
FBgn0037021	CG11399	2.43649622489732
FBgn0037248	srl	2.43635760122446

FBgn0036621	roq	2.42955082149738
FBgn0004860	ph-d	2.42903756534114
FBgn0052350	Vps11	2.42896249587558
FBgn0041094	scyl	2.42661298409011
FBgn0011725	twin	2.42563498489733
FBgn0025802	Sbf	2.42397895015361
FBgn0265082	Cdep	2.41861685099219
FBgn0261873	sdt	2.41407758789775
FBgn0003447	sn	2.41249859207593
FBgn0010316	dap	2.41134446229863
FBgn0263112	Mitf	2.40780166769646
FBgn0039302	Nup358	2.40606703947924
FBgn0267964	CG46244	2.40404964375311
FBgn0002431	hyd	2.40281841637812
FBgn0015520	nonA-l	2.40191424018986
FBgn0037344	CG2926	2.39992487571929
FBgn0004655	wapl	2.39989089784983
FBgn0027603	Ulp1	2.39710915992238
FBgn0032223	GATAd	2.39634451649019
FBgn0016081	fry	2.39344457225599
FBgn0259243	Pka-R1	2.39232550308458
FBgn0000307	chif	2.39231308871872
FBgn0264491	how	2.39118665169054
FBgn0026375	RhoGAPp190	2.39102094244172
FBgn0001324	kto	2.38596539533345
FBgn0265042	lrk1	2.38297263091889
FBgn0000273	Pka-C1	2.37696218390045
FBgn0004390	RasGAP1	2.37486462758511
FBgn0030680	CG8944	2.37340714143728
FBgn0020386	Pdk1	2.3727985103154
FBgn0050183	CG30183	2.37121829783637
FBgn0003392	shi	2.36263125229868
FBgn0035397	CG11486	2.36177342549171
FBgn0036043	CG8177	2.35854757997247

FBgn0035959	CG4911	2.35836490466317
FBgn0262124	uex	2.35817837980896
FBgn0267795	Frl	2.3576825855114
FBgn0002873	mud	2.35656922771361
FBgn0003209	raw	2.35007959980049
FBgn0267348	LanB2	2.34985035375732
FBgn0004606	zfh1	2.34806397191747
FBgn0030809	Ubr1	2.34724074399706
FBgn0032796	CG10188	2.34437811370054
FBgn0284225	CG46309	2.34290978191281
FBgn0261954	east	2.33596010857812
FBgn0039431	plum	2.33094658374251
FBgn0283427	FASN1	2.33004623210842
FBgn0033504	CAP	2.32523646705443
FBgn0014141	cher	2.3219777816587
FBgn0264895	RapGAP1	2.31861998857016
FBgn0262734	eIF4H1	2.31825158632215
FBgn0036032	CG16711	2.31492141777987
FBgn0021796	Tor	2.31471774911897
FBgn0011481	Ssdp	2.31366884885179
FBgn0033155	Br140	2.30583949606892
FBgn0026147	TLL4A	2.29747302803618
FBgn0031969	pes	2.29319594340789
FBgn0086451	l(2)k09022	2.28979262552197
FBgn0263232	Nxf3	2.28526622496391
FBgn0051158	Efa6	2.28422412767222
FBgn0032341	Reps	2.28127980299909
FBgn0243516	Vrp1	2.28127817852858
FBgn0036741	anchor	2.280969446207
FBgn0029666	CG10803	2.27860660288234
FBgn0053217	CG33217	2.27395155547263
FBgn0031759	lid	2.27321662528559
FBgn0032456	MRP	2.27044580631546
FBgn0266111	ana3	2.26939379175172

FBgn0031885	Mnn1	2.26813095649748
FBgn0030400	Pits	2.26460689042668
FBgn0003118	pnt	2.26258012116944
FBgn0037760	FBXO11	2.26117870867679
FBgn0052756	CG32756	2.26078840174657
FBgn0035842	CG7504	2.26078840174657
FBgn0052103	SCaMC	2.25863951748893
FBgn0026432	Grip163	2.25641233546752
FBgn0033802	CG17724	2.25459669574268
FBgn0284223	CG46307	2.24686457234522
FBgn0050015	CG30015	2.24442575141235
FBgn0042185	MCU	2.2427688595078
FBgn0029123	SoxN	2.23954047974951
FBgn0261383	IntS6	2.23657110238323
FBgn0031990	PAPLA1	2.23542383772161
FBgn0262740	Evi5	2.23438410598615
FBgn0030412	Tomosyn	2.23401869685696
FBgn0034087	clu	2.23187014960679
FBgn0264270	Sxl	2.22632554379118
FBgn0031187	Usp2	2.22394931577402
FBgn0259224	CG42324	2.22133349648103
FBgn0020257	Ppa	2.22077693249726
FBgn0028476	Usp1	2.21974854291646
FBgn0026189	prominin-like	2.2192287460897
FBgn0262872	milt	2.21691085967949
FBgn0034975	enok	2.21460531066863
FBgn0039944	CG17162	2.21287549675706
FBgn0015278	Pi3K68D	2.21213047307464
FBgn0038968	CG12499	2.21086051514842
FBgn0003463	sog	2.20832887634621
FBgn0000721	for	2.20778344343556
FBgn0031006	rictor	2.20585106788442
FBgn0003721	Tm1	2.20206271537516
FBgn0005632	faf	2.20135566073757

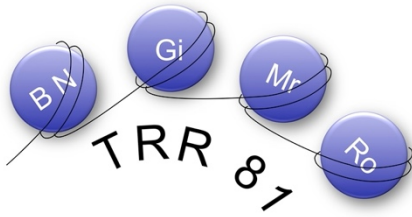


<b>FBgn0042693</b>	wrd	2.20078729456201
<b>FBgn0034158</b>	CG5522	2.19639281222816
<b>FBgn0039141</b>	spas	2.19538872246171
<b>FBgn0034083</b>	lbk	2.19295808676032
<b>FBgn0015513</b>	mbc	2.18909815396437
<b>FBgn0036534</b>	DCP2	2.18898764101792
<b>FBgn0032938</b>	CG8671	2.18733289497642
<b>FBgn0025455</b>	CycT	2.18493006626074
<b>FBgn0004889</b>	tws	2.18311973418013
<b>FBgn0029687</b>	Vap33	2.18231528793915
<b>FBgn0031896</b>	CG4502	2.17920385113533
<b>FBgn0263391</b>	hts	2.17870935704716
<b>FBgn0000317</b>	ck	2.17396869743185
<b>FBgn0058002</b>	ND-AGGG	2.17318212077496
<b>FBgn0000524</b>	dx	2.17129668624419
<b>FBgn0003396</b>	shn	2.17086724359754
<b>FBgn0259228</b>	C3G	2.17003674627658
<b>FBgn0031952</b>	cdc14	2.16993074893325
<b>FBgn0005386</b>	ash1	2.16962247455508
<b>FBgn0035237</b>	CG13917	2.16835868148038
<b>FBgn0260938</b>	tay	2.1658575456593
<b>FBgn0026181</b>	Rok	2.16565195875127
<b>FBgn0015371</b>	chn	2.16524386372645
<b>FBgn0261885</b>	osa	2.16491991225475
<b>FBgn0004391</b>	shtd	2.16353536538267
<b>FBgn0283741</b>	prage	2.1627287689669
<b>FBgn0261239</b>	Hr39	2.16198891111363
<b>FBgn0029825</b>	CG12728	2.16133510381566
<b>FBgn0041111</b>	lilli	2.16106057603673
<b>FBgn0051992</b>	gw	2.15474998721364
<b>FBgn0038504</b>	Sur-8	2.14635996182577
<b>FBgn0058263</b>	MFS17	2.14419470102632
<b>FBgn0046692</b>	Stlk	2.14380167504649
<b>FBgn0032120</b>	CG33298	2.13264937043225

<b>FBgn0039528</b>	dsd	2.13166722662859
<b>FBgn0264607</b>	CaMKII	2.13110556965004
<b>FBgn0261570</b>	CG42684	2.13109214969947
<b>FBgn0050011</b>	gem	2.13053274233609
<b>FBgn0030812</b>	wcy	2.13052392827129
<b>FBgn0003310</b>	S	2.12635222191387
<b>FBgn0011747</b>	Ank	2.12513324330182
<b>FBgn0027548</b>	nito	2.1242793674139
<b>FBgn0003892</b>	ptc	2.12054722297233
<b>FBgn0003277</b>	RplI215	2.11968763870511
<b>FBgn0035023</b>	ITP	2.1190042754919
<b>FBgn0085443</b>	spri	2.11854207792077
<b>FBgn0027948</b>	msps	2.1181843969848
<b>FBgn0039945</b>	CG17159	2.1175895390335
<b>FBgn0025866</b>	CalpB	2.11711652863622
<b>FBgn0032006</b>	Pvr	2.11254030945481
<b>FBgn0040230</b>	dbo	2.11158906049993
<b>FBgn0261618</b>	larp	2.11058275262565
<b>FBgn0284235</b>	CG46319	2.11045849283883
<b>FBgn0040010</b>	CG17493	2.10794201754853
<b>FBgn0026577</b>	CG8677	2.1033120362932
<b>FBgn0063485</b>	Lasp	2.10208243781143
<b>FBgn0264605</b>	CG43954	2.10125365441681
<b>FBgn0015286</b>	Rala	2.096522405419
<b>FBgn0003041</b>	pbl	2.09056590123978
<b>FBgn0265297</b>	pAbp	2.08901635476039
<b>FBgn0260789</b>	mxc	2.08825587776901
<b>FBgn0001168</b>	h	2.0865058772932
<b>FBgn0039907</b>	lgs	2.08587200697008
<b>FBgn0266487</b>	CG45086	2.08542670851974
<b>FBgn0035688</b>	fmt	2.08494827868727
<b>FBgn0032363</b>	Dlg5	2.08342146918111
<b>FBgn0031681</b>	Pgant5	2.08308887874727
<b>FBgn0265778</b>	PDZ-GEF	2.07757110483131

<b>FBgn0031030</b>	Tao	2.0757835738441
<b>FBgn0000183</b>	BicD	2.07492374540185
<b>FBgn0029881</b>	pigs	2.07441618807715
<b>FBgn0030344</b>	Nrd1	2.0719067755965
<b>FBgn0266450</b>	Kr-h1	2.0716438763146
<b>FBgn0033913</b>	CG8468	2.07128296365492
<b>FBgn0004101</b>	bs	2.06937035205546
<b>FBgn0022768</b>	Pp2C1	2.06844328886748
<b>FBgn0052850</b>	CG32850	2.06370671723186
<b>FBgn0014075</b>	Uggt	2.06201824934049
<b>FBgn0033569</b>	CG12942	2.05910660703589
<b>FBgn0261823</b>	Asx	2.05775802305879
<b>FBgn0013756</b>	Mtor	2.05608880140724
<b>FBgn0034046</b>	tun	2.05606508771281
<b>FBgn0263108</b>	BtbVII	2.05567475858338
<b>FBgn0001316</b>	klar	2.05378415309378
<b>FBgn0037240</b>	Cont	2.05319027054982
<b>FBgn0037720</b>	CG8312	2.04760291043297
<b>FBgn0002921</b>	Atpalpha	2.04654126515635
<b>FBgn0025633</b>	CG13366	2.04524519260379
<b>FBgn0032123</b>	Oatp30B	2.04459278854709
<b>FBgn0005617</b>	msl-1	2.03751144023282
<b>FBgn0016917</b>	Stat92E	2.03560644896501
<b>FBgn0037770</b>	Art4	2.03212653592837
<b>FBgn0003410</b>	sina	2.02953402706872
<b>FBgn0030466</b>	CG15744	2.02715697745512
<b>FBgn0264560</b>	garz	2.02588702575747
<b>FBgn0010309</b>	pigeon	2.02031322334521
<b>FBgn0034657</b>	LBR	2.01628140677471
<b>FBgn0011766</b>	E2f1	2.01459943305984
<b>FBgn0052676</b>	stx	2.01354758724351
<b>FBgn0002716</b>	mei-W68	2.01141652315342
<b>FBgn0037831</b>	Cap-H2	2.00752263397315
<b>FBgn0021760</b>	chb	2.00709774689602

<b>FBgn0033246</b>	ACC	2.00592253890408
<b>FBgn0026620</b>	tacc	2.00397501700747
<b>FBgn0014007</b>	Ptp69D	2.00168898842332
<b>FBgn0015477</b>	Rme-8	2.00021075832301



## PhD Portfolio - MGK

### Summary of PhD training

**Name of PhD student: Ikram Ullah**

**Affiliation:**

BMFZ – Institute of Molecular Biology und Tumor research (IMT)  
Philipps-Universität Marburg  
Hans-Meerwein-Str. 2  
D-35043 Marburg  
Tel: +49-6421-28-66854  
Fax: +49-6421-28-66842  
e-mail: ikram.ullah@imt.uni-Marburg.de

**Start of PhD work:** (02-12-2014)

**Topic of PhD thesis work:**

Characterization of RNA interactions of dMi-2

**Thesis committee:**

1. Prof. Dr. Alexander Brehm (Supervisor)
2. Prof. Dr. Tillman Borggrefe
3. Prof. Dr. Uta-Maria Bauer

**Meetings of the thesis committee:**

- 09-09-2015
  - 27-02-2017
- Other presentations were the thesis committee members were present:
- Kleinwalsertal retreat, 25-02-2019
  - Kleinwalsertal retreat, 31-03-2017
  - TRR81 internal seminar, 01-10-2020

**PhD training**

**General courses**

-

**Specific courses**

-

**Seminars and workshops**

Transregio 81 seminar: RNA antagonizes dMi-2/CHD4 interaction with chromatin (Marburg, via zoom. 01-10-2020).

**Presentations:**

Journal Club:

- 1) Mutant p53 cooperates with SWI/SNF chromatin remodelling complex to regulate VEGFR2 in breast cancer cells. Pfister et. al., Genes & Development 2015 (IMT, 04-08-2015)
- 2) MDM2 associates with Polycomb repressor complex 2 and enhances stemness-promoting Chromatin modification independent of p53. Wieken et.al., Molecular Cell 2016. (IMT, 26-01-2016)
- 3) Chromatin remodelling inactivates activity genes and regulates neural coding. Yang et al., Science 2016 (IMT, 19-07-16)
- 4) Polycomb Repressive Complex 1 generates discrete compacted domains that change during differentiation. Kundu et al. Molecular Cell 2017 (BMFZ, 16-02-17)
- 5) Targeting of Polycomb Repressive Complex 2 to RNA by short repeats of Guanines. Wang et.al., Molecular Cell 2017 (BMFZ, 20-07-17)
- 6) Conserved RNA-binding specificity of Polycomb repressive complex 2 is achieved by dispersed amino acid patches in EZH2. Long et. al., eLife (BMFZ, 08-02-18)
- 7) The interaction of PRC2 with RNA and chromatin is mutually antagonistic. Beltran et al., Genome Research 2016 (BMFZ, 29-11-18)
- 8) Xist RNA antagonizes SWI/SNF chromatin remodeler BRG1 on the inactive X-chromosome. Jégu et al., Nature Structural and Molecular Biology 2019 (BMFZ, 23-05-19)
- 9) G-tract RNA removes Polycomb repressive complex 2 from genes. Beltran et al., Nature Structure and Molecular Biology 2019 (BMFZ, 14-11-19).
- 10) RNA is essential for chromatin occupancy and function in human pluripotent stem cells. Long et al. Nature Genetics. 2020

Chromatin Methods Club:

-

**Retreats:**

Oral presentation

- 1) 5<sup>th</sup> Winter school of the collaborative research center TRR81 'Chromatin Changes in Differentiation and Malignancies' Marburg-Gießen-Bad Nauheim-Rotterdam. dMi2 and RNA (Kleinwalsertal, 22-02-2015).
- 2) 7<sup>th</sup> Winter school of the collaborative research center TRR81 'Chromatin Changes in Differentiation and Malignancies' Marburg-Gießen-Bad Nauheim-Rotterdam. Recruitment mechanisms of dMi-2 (Klienwalsertal, 01-03-2017)
- 3) 9<sup>th</sup> Winter school of the collaborative research center TRR81 'Chromatin Changes in Differentiation and Malignancies' Marburg-Gießen-Bad Nauheim-Rotterdam. Role of dMi-2 RNA interaction (Kleinwalsertal, 25-02-19)

**(Inter)national conferences**

4<sup>th</sup> TRR81 Symposium –Chromatin Changes in Differentiation and Malignancies 18 – 20 September 2017, Rotterdam, Netherlands. Poster Presentation - dMi-2 recruitment mechanism to heat shock genes (18-09-2017).

34<sup>th</sup> Ernst Klenk Symposium in Molecular Medicine – 'Epigenetics: Basic principles and clinical applications 04-06 October 2018, Cologne, Germany. Poster Presentation – Role of RNA in the recruitment of dMi-2 (06-10, 2020).

**Organisation of student workshops/mini-symposia on specific topics**

- 12<sup>th</sup> TRR81 PhD Mini-Symposium 'Advanced methods in chromatin and gene expression research, Marburg, Germany (08-04-2019)

**Internships in laboratories of the TRR81 in Rotterdam**

-

**Others**

- Supervised a Lab internship for a Bachelors student of Human Biology course, Marburg (04-08-2015) to (06-10-2015)
- Supervised a Lab internship for a Bachelor student of Human Biology course, Marburg (15-10-2018) to (15-12-2018)
- Supervised the Bachelor's Thesis for an Erasmus exchange student from Università degli Studi de Perugia, Italy. 'Salt sensitivity of dMi-2 binding to RNA'. Marburg (15-04-2019) to (22-07-2019)
- Supervised the Master Thesis for a student of Human Biology course 'Characterization of dMi-2 RNA binding' Marburg (01-09-2019) to (01-09-2020).
- Supervised practical course for students of medicine, Marburg (15-06 to 24-06, 2015).
- Supervised practical course for students of medicine, Marburg (17-06 to 22-06, 2016).
- Supervised practical course for students of medicine, Marburg (21-06 to 27-06, 2017).
- Supervised practical course for students of medicine, Marburg (11-06 to 13-06, 2018).
- Supervised practical course for students of medicine, Marburg (13-06 to 19-06, 2019).

## Curriculum vitae

### Personnel Information

Surname: Ullah  
First name: Ikram  
Date of birth: 21<sup>st</sup> January 1989  
Place of birth: Srinagar, Kashmir  
Nationality: Indian

### Education

**02-12-2014 – present. PhD in the lab of Prof. Dr Alexander Brehm**  
Institute of Molecular Biology and Tumor Research (IMT)  
School of Medicine, Philipps-Universität, Marburg  
PhD dissertation – ‘Characterization of RNA interactions of dMi-2’

**08-2012 – 06-2013 MSc Thesis Project in the lab of Dr Chetana Sachidanandan**  
Institute of Genomics and Integrative Biology (IGIB)  
Council of Scientific and Industrial Research (CSIR), New Delhi  
Master Thesis – ‘Identification of Bio-actives using Zebrafish embryonic development screens.’

**08-2011 – 06-2013 MSc in Biomedical Genetics**  
School of Biosciences and Technology (SBST)  
Vellore Institute of Technology (VIT), Vellore

**08-2007 – 08-2010 BSc in Genetics**  
MS Ramaiah College of Arts, Science and Commerce  
Bangalore  
Bachelor Thesis – ‘Field study of gene and genotype of Autosomal Mendelian traits in Humans.’

**01-2005 – 12-2006 High Secondary School**  
Iqbal Memorial Institute (IMI), Srinagar

### Fellowships

**08-2013 – 11-2014 Research Assistent Fellow**  
Dr Chetana Sachidanandan Lab  
Institute of Genomics and Integrative Biology (IGIB)  
Council of Scientific and Industrial Research (CSIR), New Delhi  
Project 1- ‘Role of lincRNA in the expression regulation of Foxd3 gene.’  
Project 2- ‘Understanding the role of chd7 gene in CHARGE Syndrome using zebrafish models.’



**Publications**

Ullah, I., Thölken, C., Zhong, Y., John, M., Roßbach, O., Lenz, J., Vasquez, O., Chung, H.R., Mackay, J.P., Brehm, A. (2021). RNA binding regulates enzymatic activity and chromatin binding of the nucleosome remodeler dMi-2/CHD4. In preparation.

Jox, T., Buxa, M.K., Bohla, D., Ullah, I., Mačinković, I., Brehm, A., Bartkuhn, M., and Renkawitz, R. (2017). Drosophila CP190- and dCTCF-mediated enhancer blocking is augmented by SUMOylation. *Epigenetics & Chromatin* *10*, 32.

Kreher, J., Kovač, K., Bouazoune, K., Mačinković, I., Ernst, A.L., Engelen, E., Pahl, R., Finkernagel, F., Murawska, M., Ullah, I., et al. (2017). EcR recruits dMi-2 and increases efficiency of dMi-2-mediated remodelling to constrain transcription of hormone-regulated genes. *Nature Communications*.

ASAD, Z., PANDEY, A., BABU, A., SUN, Y., SHEVADE, K., KAPOOR, S., ULLAH, I., RANJAN, S., SCARIA, V., BAJPAI, R., ET AL. (2016). RESCUE OF NEURAL CREST-DERIVED PHENOTYPES IN A ZEBRAFISH CHARGE MODEL BY SOX10 DOWNREGULATION. *HUMAN MOLECULAR GENETICS*.

## **List of academic teachers**

### **My academic teachers at Vellore Institute of Technology (VIT) were –**

Abilash	Prof. Dr. V.G.
Devi	Prof. Dr A.
Kalaivani	Prof Dr. T.
Kanabiran	Prof. Dr. K.
Patel	Prof. Dr T.N.
Sabina	Prof. Dr. E-P.
Saraswati	Prof. Dr. R.
Subhashree	Prof. Dr V.
Tamizhselvi	Prof. Dr. R.
Thiagrajan	Prof. Dr. P.

### **My academic teachers at School –**

Sofi	Prof. Dr. A.
Sofi	Prof. Dr I.
Hanief	M.
Jawahira	M.
Majid	A.
Manzoor	A.
Mustufa	A.
Nabi	G.
Naqshbandi	J.
Shafi	M.
Sidique	M.

## **Acknowledgement:**

First of all, I would like to thank Prof. Dr. Alexander Brehm for providing me an opportunity to pursue a PhD in his Lab. I express my gratitude to you for constantly guiding me, helping me throughout with suggestions. In addition to science, I learned several other aspects of life from you. Your humbleness is one of the several things I would like to carry on in my life. It was a pleasure to do my PhD under your supervision. Thank you so much!

Also, I would like to thank my Thesis committee members – Prof. Dr. Uta-Maria Bauer and Prof. Dr. Tilman Borggreve for valuable suggestions during the process of my PhD.

I would also like to thank Dr. Clemens Thölken for helping me with Bioinformatics analysis and Dr Stephan Awe for always being kind and helpful. I thank all the internship students whom I supervised, especially, Mara John. I wish you a successful PhD.

Importantly, I would like to thank Dr. Karim Bouazoune for constantly answering my questions. You are an extremely helpful person and I am so glad that over my PhD we became good friends. Thank you!

My family, I am blessed to have the most kind and caring parents. I appreciate that you believed in me. My two sister, who I cannot thank enough. Your constant support and encouragement kept me going. You are the world to me!

A special thank you to Prof. Dr. Roland Lill and Angelika-Haag Lill. You both are like a family to me. All these years living away from home would have been much more difficult if it were not for you. My sincere gratitude for the love and care you have shown to me.

A very special person, my teacher who inspired me to pursue research Prof. Dr. Ali Muhammad Sofi. I wish there were more teachers like you. May you rest in peace!

To all my teachers, I am grateful that you helped me through my education.

To all my friends, especially Owais Altaf and Zeeshan Mushtaq, I am so glad to have you in my life.

Also, thanks to all other members of the Brehm Lab!

### **Ehrenwörtliche Erklärung**

„Ich Ikram Ullah geboren am 21-01-1989, in Srinagar, Kashmir (India), erkläre ehrenwörtlich, dass ich die dem Fachbereich Medizin Marburg zur Promotionsprüfung eingereichte Arbeit mit dem Titel '**Characterization of RNA Interactions of dMi-2**', im Institut für Molekular Biologie und Tumorforschung (Geschäftsführender Direktor: Prof. Dr. Alexander Brehm) unter Leitung von Prof. Dr. Alexander Brehm ohne sonstige Hilfe selbst durchgeführt und bei der Abfassung der Arbeit keine anderen als die in der Dissertation aufgeführten Hilfsmittel benutzt habe. Ich habe bisher an keinem in- oder ausländischen Medizinischen Fachbereich ein Gesuch um Zulassung zur Promotion eingereicht, noch die vorliegende oder eine andere Arbeit als Dissertation vorgelegt.

Ich versichere, dass ich sämtliche wörtlichen oder sinngemäßen Übernahmen und Zitate kenntlich gemacht habe.

Mit dem Einsatz von Software zur Erkennung von Plagiaten bin ich einverstanden.

**Marburg, den 25-01-2021**

**Ikram Ullah**

**Die Hinweise zur Erkennung von Plagiaten habe ich zur Kenntnis genommen.**

**Marburg, den 25-01-2021**

**Prof. Dr. A. Brehm**