Analysis of Argonaute - Chromatin Interactions in Human Somatic Cells

DISSERTATION ZUR ERLANGUNG DES
DOKTORGRADES DER NATURWISSENSCHAFTEN (DR. RER. NAT.)
DER FAKULTÄT FÜR BIOLOGIE UND VORKLINISCHEN MEDIZIN

DER UNIVERSITÄT REGENSBURG

vorgelegt von
Nicholas Strieder, geb. Putz

aus
Deggendorf, Deutschland

im Jahr
2015

Das Promotionsgesuch wurde eingereicht am:
07.Mai.2015

Die Arbeit wurde angeleitet von:
Prof. Dr. Gunter Meister

Unterschrift:

For Lilly, Joshua and Katrin.

Summary

The cellular abundance of a protein is regulated at many different steps including transcription, splicing and translation of its encoding mRNA but also degradation of both, the mRNA and the protein. One regulator influencing some of these processes are Argonaute (Ago) proteins which are guided by an associating small RNA molecule to complementary mRNAs in the cytoplasm resulting in their translational repression and degradation. Apart from their cytoplasmic function which is conserved across many species, some Ago proteins were also found to influence protein abundance at the step of mRNA transcription in the nucleus of cells. There, Ago proteins bound to a small RNA recognize nascent transcripts of mRNAs and transposable elements by complementary base pairing leading to their transcriptional silencing. This is often linked to the recruitment of chromatin modifying factors. The nuclear function of Ago proteins, which was studied especially in yeast, plants and worms, was found to be inducible also in human cells by targeting Ago to chromatin by promoter complementary small RNA species [Meister, 2013]. Therefore, the question after endogenous triggers and the genomic target sites of this potential function emerged.

In order to identify potential target gene loci in human somatic cells and mouse embryonic stem cells, we examined Ago chromatin interaction sites by applying two complementary techniques, Chromatin Immunoprecipitation (ChIP) and DNA Adenine Methyltransferase Identification (DAMID). For a genome wide characterisation of chromatin binding sites both techniques were combined with next generation sequencing. At first, we analysed the chromatin interaction sites of Ago 1 and 2 by comparing data sets derived from ChIP experiments using three different antibodies each. However, the really low number of overlapping peaks (<11) from the three different ChIP experiments as well as the bad signal-to-noise level of the data, suggested either lack of strong chromatin interaction sites or experimental failure.

As ChIP experiments are highly dependent on the performance of the antibody under these harsh conditions, we next applied the antibody-independent DAMID approach. This method deduces chromatin association sites from DNA-methylation events caused by a recombinant fusion protein consisting of a DNA methyltransferase from *E. coli* and the target protein. The successful establishment of the method was shown by the results for a positive control which were in agree-

ment with the literature [Vogel *et al.*, 2006]. However, the validation of the predicted nuclear Ago-binding sites by ChIP experiments failed for the human cell line, which could be due to the lack of replicates needed to identify enrichment sites with a low false-positive rate or the absence of true chromatin binding sites of Ago in human somatic cells. For mESC only few significant peaks were detected which are currently validated by our collaboration partner. Excitingly, both DAMID experiments showed high enrichment of mitochondrial DNA sequences in Ago samples compared to controls. Therefore, we tested a potential mitochondrial localisation of Ago proteins, which was also suggested by hints in the current literature but was not thoroughly assessed [Bian *et al.*, 2010; Bandiera *et al.*, 2011; Sripada *et al.*, 2012; Zhang *et al.*, 2014]. Assisted by the expertise of the lab of Prof. Dr. P. Rehling in mitochondrial biology, we analysed import and localisation of Ago in mitochondria. While some of the experiments supported a potential inter membrane space localisation, no evidence for matrix-import required for specific interaction of Ago proteins with mitochondrial DNA could be found. When the results from mitochondrial localisation were combined with in depth analysis of the sequencing reads form the DAMID experiments, it appeared that a high fraction of the reads were highly amplified non-specific mitochondrial fragments reducing the complexity of DNA library. As implied in the literature [Casbon *et al.*, 2011], this low complexity has a negative influence on quantification of the lower abundant DNA fragments in the library. In addition, hints for a correlation between amplification of non-specific mitochondrial DNA and the extent of cytoplasmic localisation of the Dam-fusion protein could be found.

When the results of the DAMID experiment from human cells were compared with another related study [Benhamed *et al.*, 2012], some overlap of Ago-chromatin interaction sites was found. However, these sites ranked rather low in the bioinformatic analysis probably due to the high level of non-specific amplification of mitochondrial sequences, high background noise and the lack of replicate experiments. This was likely aggravated by a strong cytoplasmic localisation of human Ago proteins compared to their low abundance in the nucleus.

In conclusion, we can assume that the DAMID approach is also feasible for Ago proteins but needs further improvements like nuclear targeting and depletion of mitochondrial sequences. Based on our results we can only speculate about a nuclear function of Ago proteins. Several studies showed induction of Ago-chromatin association under special conditions like cellular senescence, transfec-

tion of promoter complementary siRNAs or DNA double strand break repair [Benhamed *et al.*, 2012; Kim *et al.*, 2006; Janowski *et al.*, 2006; Wei *et al.*, 2012]. However, others suggested Ago-chromatin binding under steady-state conditions to regulate alternative splicing and silencing or activation of transcription [Allo *et al.*, 2009; Ameyar-Zazoua *et al.*, 2012; Huang *et al.*, 2013a]. In order to clarify the controversially discussed nuclear functions of Ago proteins, antibody-independent techniques like DAMID are needed to critically assess the proposed mechanisms.

Contents

Symbols

α-	anti
Δ	deletion
μ	micro (10^{-6})
A	
A	Ampere
AA	amino acid
AB	antibody
Amp	ampicillin
A. thaliana	*Arabidopsis thaliana*
ATP	adenosine triphosphat
B	
b	base
Bq	Becquerel
bp	base pairs
BSA	bovine serum albumin
C	
cDNA	copy DNA
C. elegans	*Caenorabditis elegans*
ChIP	Chromatin Immunoprecipitation
Ci	Curie
C-terminal	carboxy terminal
C-Terminus	carboxy terminal end of a protein
CTD	c-terminal domain
D	
Da	Dalton
Dam	DNA adenine methyltransferase
DAMID	DNA adenine methyltransferase identification
DEPC	Diethylpyrocarbonat
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	Deoxyribonuclease
D. melanogaster	*Drosophila melanogaster*

ds	double stranded
DSB	Double strand break
DTT	dithiothreitol
E	
E. coli	*Escherichia coli*
EDC	N′-ethylcarbodiimide hydrochloride
EDTA	ethylendiamin-tetraacetate
EGFP	enhanced green fluorescent protein
C. elegans	*Caenorabditis elegans*
ENCODE	ENCyclopedia Of DNA Elements
ER	endoplasmatic reticulum
EtBr	ethidium bromide
et al.	et alii, et aliae
EtOH	ethanol
F	
FH	FLAG-HA-fuision peptide (DYKDDDDK-YPYDVPDYA)
G	
g	Gramm
GAP	GTP activating protein
GDP	guanosine diphosphat
GFP	green fluorescent protein
GTP	guanosine triphosphat
GW	Glycine-tryptophane repeat
H	
HA	gemagglutinin-epitope (YPYDVPDYA)
His	6x histidine tag
HR	homologous recombination
H_2O	water
I	
IF	immunfluorescence
IP	immunoprecipitation
K	
k	kilo

kd.	knockdown
L	
l	Litre
LB	Luria Broth, bacterial medium
M	
mAB	monoclonal antibody
min	Minute
ml	Millilitre
mRNA	messenger RNA
miRNA	microRNA
miRNP	miRNA-Ribonucleoprotein
N	
NGS	Next-generation-sequencing
nm	Nanometer (10^{-9})
nt	nucleotides
N-Terminus	amino-terminal end of a protein
O	
OD_{600}	optical density, wavelength λ = 600 nm
ORF	open reading frame
P	
^{32}P	radioactive phosphate-32 (-PO_3^{3-})
PBS	phophate-buffered saline
PAGE	Polyacrylamide Gel-electrophoresis
PCR	Polymerase chain reaction
PEG	polyethylenglykol
PNK	T4 Polynucleotide Kinase
piRNA	PIWI-interacting RNA
Poly(A)	chain of adenine nucleotides
pre-	precursor
pri-	primary
PTGS	post transcriptional gene silencing
Q	
qRT-PCR / qPCR	quantitative real-time PCR
R	

	RDRC	RNA-directed RNA polymerase complex
	RISC	RNA-induced silencing complex
	RIP	RNA-Immunoprecipitation
	RITS	RNA-induced transcriptional silencing complex
	rpm	Rounds per Minute
	RNA	ribonucleic acid
	RNAi	RNA-interference
	RNase	ribonuclease
	RNP	Ribnucleoprotein
	rRNA	ribosomal RNA
	RT	room temperature
S		
	SDS	sodium dodecyl sulphate
	siRNA	small interfering RNA
	S. pombe	*Schizosaccharomyces pombe*
	ss	single stranded
T		
	TAE	Tris/ acetate/ EDTA
	TBS	Tris buffered saline
	TE	Tris/ EDTA
	TEMED	N,N,N′,N′-tetramethylethylenediamine
	TGS	transcriptional gene silencing
	Tris	Tris(hydroxymethyl)aminomethane
	tRNA	transfer RNA
U		
	U	Units
	UV	ultraviolet light
V		
	V	Volt
W		
	WT	wild type

Genes and proteins were abbreviated according to standard nomenclature in the literature.
Abbreviations of amino acids and nucleotides used are according to international rules.

Chapter 1

Introduction

1.1 The Discovery of Argonaute Proteins

The first member of the Argonaute (Ago) family of proteins to be discovered was Argonaute 1 in *Arabidopsis thaliana* [Bohmert *et al.*, 1998]. He named the gene after the octopus *Argonauta argo* due to the squid like phenotype of the mutant plant (Figure 1.1). Ago plays a role in germination, development of leaves and inflorescence, which is apparent from the radial symmetry of the stem leaves instead of an adaxial/abaxial differentiation or the lack of branching of the stem for the homozygous mutant. The flowers of mutant plants also fail to develop a proper pistil, anthers and pollen which causes complete sterility of the plants [Bohmert *et al.*, 1998]. Sequence comparisons revealed that Argonaute proteins are conserved in many different organisms ranging from archaea to mammals [Swarts *et al.*, 2014]. The eukaryotic part of the family of Argonaute proteins will be described in Section 1.2.

A hint on the molecular function of Argonaute proteins came from a genetic screen in *Caenorhabditis elegans* for factors involved in a mechanism called RNA interference (RNAi) [Tabara *et al.*, 1999]. RNAi was discovered by Fire *et al.* [1998] when they used long dsRNAs to repress the expression of a complementary gene in *C. elegans*. As already low amounts of the dsRNA were sufficient to silence an excess of endogenous transcripts, they proposed a catalytic component to be involved in the depletion of these transcripts. In addition, they suggested targeting of chromatin or transcription as alternative explanations for the observed repression [Fire *et al.*, 1998]. The catalytic complex postulated by Fire *et al.* [1998] was soon identified to comprise an Argonaute protein bound to a single stranded

small RNA with a length ranging from 21 to 30 *bp* [Hamilton and Baulcombe, 1999; Tabara *et al.*, 2002]. In the case of *C. elegans* these small RNAs derive from the long dsRNA used in the experiment by Fire *et al.* [1998]. Today, small RNAs are divided into three main classes which comprise small interfering RNAs (siRNAs), microRNAs (miRNAs) and Piwi interacting RNAs (piRNAs) and are further explained in Section 1.3.

All the mechanisms proposed by Fire *et al.* [1998] to explain RNAi are conserved in different groups of organisms and are referred to as post transcriptional gene silencing (PTGS) and transcriptional gene silencing (TGS). In the first process Argonaute is guided by a small RNA to target an mRNA by complementary base pairing and mediates, in most of the cases, degradation of the target RNA [Huntzinger and Izaurralde, 2011]. This will be explained in section Section 1.4 with a focus on human cells. In the second mechanism an Argonaute protein associates with a small RNA in the cytoplasm and is subsequently imported into the nucleus [Guang *et al.*, 2008]. There the small RNA-Ago complex associates with nascent transcripts by complementary base pairing [Verdel *et al.*, 2004; Bühler *et al.*, 2006; Guang *et al.*, 2010] and leads to inhibition of RNA PolII [Burkhart *et al.*, 2011; Guang *et al.*, 2010] and silencing of the gene at the chromatin level [Gu *et al.*, 2012; Bühler *et al.*, 2006]. Following a short introduction to chromatin and transcription, which are both influenced by TGS (Section 1.5.1), examples of TGS mechanisms from several organisms will be described in Section 1.5. The insights from these model systems will help to interpret the divergent publications on this process in the human system, which set the basis for the initiation of this project.

Figure 1.1: ***Squid like Ago mutant of A. thaliana.*** *Picture of a 35-day-old ago1-1 seedling, taken from Bohmert et al. [1998].*

Today, the field of RNAi is still strongly expanding and has a high impact on biological and medical research. SiRNAs as tools for silencing of specific genes to

study their biological function in mammalian cells are commonly used in molecular biology laboratories [Elbashir *et al.*, 2001]. Furthermore, small RNA derived methods for cancer diagnostics and therapy are constantly improved [Farooqi *et al.*, 2014]. But even in the general public RNAi gained awareness, when C. Mello and A. Fire were awarded the Nobel Prize for physiology and medicine in 2006.

1.2 The Argonaute Family of Proteins

The eukaryotic family of Argonaute proteins is divided into four clades: the Trypanosoma Ago, the WAGO, the Ago and the Piwi family of proteins [Swarts *et al.*, 2014]. The protozoan parasite *Trypanosoma brucei* comprises two Ago-like proteins TbAgo1 and TBPWI1, of which only the first one is required for RNAi [Durand-Dubief and Bastin, 2003]. TbAgo1 is mainly targeting retrotransposon transcripts by cleavage, thus protecting the parasite from the deleterious effects of these mobile genetic elements. WAGO proteins are *C. elegans* specific Argonaute proteins, which are involved in PTGS and TGS mechanisms and bind to small RNAs with a length of 22 *nt* and a 5′-tri-phosphorylated G as the first base [Swarts *et al.*, 2014]. The human genome encodes for eight proteins of the Argonaute family. Four proteins belong to the Ago clade (Ago1-4) and four proteins are part of the Piwi clade (Hili, Hiwi, Hiwi2, Hiwi3) [Ender and Meister, 2010]. The proteins of the Ago clade are often found in the cytoplasm or the nucleus and bind to siRNAs and miRNAs. They trigger PTGS and TGS processes of various transcripts in the cell. The Piwi proteins interact with piRNAs and function mainly in the control of transposable elements in the germline [Swarts *et al.*, 2014].

1.2.1 The Domain Organisation of Argonaute Proteins

Argonaute proteins consist of two lobes, the first one containing the N-terminal and PAZ (PIWI-AGO-ZWILLE) domain which is connected via a hinge region to the second lobe comprising the MID and PIWI domain [Schirle and MacRae, 2012]. The N-terminal domain is required for cleavage activity [Hauptmann *et al.*, 2013] and presumably assists in unwinding of the small RNA duplex during loading of the RISC complex [Meister, 2013]. Interaction with the 3′-end of one strand of the small RNA duplex is mediated by the PAZ domain [Song *et al.*, 2003; Lin-

gel *et al.*, 2003; Ma *et al.*, 2004]. The bound RNA strand is called the guide strand, as it binds by complementary base paring to the target RNA and thus mediates the specific interaction between the target RNA and the Argonaute protein. In Piwi proteins the PAZ domain is specific for the 2′-O-methylated 3′-end of piRNAs. The MID domain forms a basic nucleotide binding pocket and harbours the phosphorylated 5′-end of the single stranded small RNA. In addition, also stacking interactions between a conserved tyrosine in the MID domain and the 5′-base of the small RNA stabilise the binding [Boland *et al.*, 2010; Frank *et al.*, 2010; Faehnle and Joshua-Tor, 2010]. The paring with the target RNA is assisted by the MID domain which additionally interacts with nucleotide 2-7 or 8 of the small RNA, the so called seed sequence of the small RNA [Ha and Kim, 2014]. Cleavage activity is conveyed to Ago2 by the PIWI domain which resembles RNAseH [Meister *et al.*, 2004; Liu *et al.*, 2004] and Piwi proteins [Brennecke *et al.*, 2007; Gunawardane *et al.*, 2007]. The cleavage activity is conferred to Ago2 by a catalytic tetrade consisting of the amino acids DEDH (more generally DEDX, where X is D or H) [Nakanishi *et al.*, 2012] in combination with two domains in the N-terminus. These N-terminal domains were identified by comparison of Ago2 with Ago3 which also contains the catalytic tetrade but is defective in target RNA cleavage [Hauptmann *et al.*, 2013]. The other two members of the Ago clade Ago1 and 4 have also lost cleavage or slicing activity [Meister *et al.*, 2004; Hauptmann *et al.*, 2013].

However, slicing is only found in the case of perfect complementarity between the Ago2 bound siRNA or miRNA guide strand and the target transcript. This Ago-small RNA complex is commonly referred to as the small RNA induced silencing complex (RISC) [Hammond *et al.*, 2000]. For most miRNAs imperfect base pairing of the Ago bound guide strand and the target RNA is observed. This complex is therefore called miRNA containing RISC (miRISC) or microribonucleoprotein (miRNP) [Meister, 2013].

1.3 Classes and Biogenesis of small RNAs

The main principles in biogenesis of small RNAs will be described for the three main classes: siRNAs, miRNA and piRNAs.

SiRNAs

SiRNAs are ~21 *nt* in size and mostly derive from endogenous or exogenous long dsRNAs. In fission yeast, plants and worms these long dsRNAs have various sources: First, long dsRNAs are formed by transposable elements, convergent transcription by antisense promoters, from viral genomes following infection, natural antisense transcripts or inverted repeats, which have also been identified as source of endo-siRNAs in mouse oocytes [Bologna and Voinnet, 2014; Grishok, 2013; Watanabe *et al.*, 2008; Tam *et al.*, 2008; Kawamura *et al.*, 2008; Ghildiyal *et al.*, 2008]. Second, in fission yeast, worms and plants there is a mechanism for secondary siRNA production. Recruited by a RISC complex bound to a target transcript, an RNA-dependent-RNA-polymerase (RdRP) associates with the target RNA and replicates a second strand to form a long dsRNA molecule [Bologna and Voinnet, 2014; Grishok, 2013]. As the RISC complex can later dissociate and trigger the generation of many long dsRNAs, this leads to strong enhancement of the RNAi response. Third, feeding (*C. elegans*) or transfection of cells with long dsRNA also triggers an RNAi response in invertebrates and plants [Fire *et al.*, 1998; Hammond *et al.*, 2000; Bologna and Voinnet, 2014]. In mammalian cells long dsRNAs, however, induce the interferon response which is part of the immune response to viral infections and causes inhibition of translation and apoptosis. In contrast, the transfection of short dsRNAs $<$27 *nt* results in efficient induction of RNAi in cultured mammalian cells [Elbashir *et al.*, 2001].

Apart from secondary siRNAs in *C. elegans* the long dsRNAs are processed into siRNAs by an RNAse-III-like enzyme, called Dicer or Dicer-like in plants [Hutvagner *et al.*, 2001; Ha and Kim, 2014]. Dicer proteins have a size of ~200 *kDa* and contain several functional domains which are described in the following. The ATP binding and the helicase domain bind to one end of the dsRNA and serve as interaction domain with associated proteins. These are followed by the domain of unknown function (DUF283). The PAZ domain is similar to the one of Argonaute proteins (Section 1.2) and recognises the 2 *nt* overhang at the 3′-end of dsRNAs. In mammals it additionally binds to a phosphorylated 5′-end [Ha and Kim, 2014; Grishok, 2013]. This explains the preference of Dicer for overhanging ends over blunt ends of its substrate dsRNAs [Vermeulen *et al.*, 2005]. Two RNAse-III-like domains form an intermolecular dimer and are required for endonucleolytic cleavage of the dsRNA. In case a 2 *nt* overhang at the 3′-end of the dsRNA is present, the region between the PAZ and RNAse-III-like domains

functions as a "molecular ruler" and results in Dicer producing small RNAs of a discrete size ranging from 20 to 24 *nt* depending on the isoform and species [Vermeulen *et al.*, 2005; MacRae *et al.*, 2006]. Last, Dicer contains a dsRNA binding domain RBD [Ha and Kim, 2014]. The binding to substrate RNAs as well as processing and efficiency of Dicer are regulated by associating proteins reviewed in Ha and Kim [2014]. In plants, worms and files small RNAs are protected from degradation by the methyltransferase HEN1, adding a methyl-group to the 2′-O at the 3′-end [Ha and Kim, 2014]. Subsequently, these siRNAs are loaded into Ago proteins and unwound resulting in a single strand remaining bound, which requires the heat-shock protein 90 (HSP90) chaperon and is further described in [Dueck and Meister, 2014].

MiRNAs

Another species of small RNAs loaded into Ago proteins are miRNAs. Theses are encoded in the genome and transcribed by RNA Polymerase II (RNA PolII) into long primary miRNAs (pri-miRNA) transcripts, which can base pair due to partial reverse complementarity in certain regions and form so called stem loop or hairpin structures [Lee *et al.*, 2004a] (see Figure 1.2). These are recognized and processed into precursor-miRNAs (pre-miRNA) of ~70 *nt* by the Microprocessor complex consisting of the proteins Drosha, an RNase-III-like enzyme [Lee *et al.*, 2003] and DGCR-8 [Gregory *et al.*, 2004]. Drosha is a nuclear protein of ~160 *kDa*. Similar to Dicer described above, Drosha contains two RNAse-III-like domains forming an intramolecular dimer. Therefore, it cleaves pri-miRNA to yield a hairpin carrying at one end a phosphorylated 5′-end opposite to a 2 *nt* 3′-overhang (see Figure 1.2). As the dsRBD of Drosha is not sufficient for pri-miRNA interaction, DGCR8 is required [Gregory *et al.*, 2004; Ha and Kim, 2014]. This ~90 *kDa* protein localises to the nucleus and contains two dsRBDs to associate with pri-miRNAs. The C-terminal part of DGCR8 interacts with Drosha [Ha and Kim, 2014]. Recent reports by Weitz *et al.* [2014] and Quick-Cleveland *et al.* [2014] suggest the involvement of a Fe III heme group bound to DGCR8 in the recognition of the basal and/or the apical transition from single stranded to double stranded RNA. DGCR8 serves thus as molecular ruler, resulting in the cleavage of pri-miRNA 11 *bp* from the basal junction of the stem loop by Drosha [Han *et al.*, 2006]. The newly created pre-miRNA hairpin is then transported into the cytoplasm by Exportin 5 which forms an export complex with the pre-miRNA

and RAN-GTP. After release of the pre-miRNA in the cytoplasm it is bound and further processed by Dicer (see Figure 1.2). The helicase domain of Dicer interacts with the terminal loop of the pre-miRNA while the 5′-end and the 3′-overhang are bound by the PAZ domain. In complex with the dsRBDs containing proteins TRBP or PACT Dicer processes the pre-miRNA into a ~22 *nt* long miRNA duplex [Bernstein *et al.*, 2001; Grishok *et al.*, 2001; MacRae *et al.*, 2006]. The miRNA biogenesis process is regulated by a wide range of factors influencing for example pre-miRNA processing by Dicer or its degradation which are reviewed in Treiber *et al.* [2012]. In addition to this, miRNAs can also derive from different sources so called MIRTRONS were miRNA are encoded in introns or from precursor tRNAs, from which they are released during splicing [Meister, 2013]. These pathways are either Drosha or Dicer independent [Meister, 2013]. One prominent Dicer independent example is miRNA-451. Its precursor pre-miR-451 is bound by Ago2 and entirely dependent on its slicing activity and the shortening by an exonuclease to reach its mature form. As miRNA-451 is vital for erythrocyte maturation, Ago2 knockout leads to early death of mice [Cheloufi *et al.*, 2010; Cifuentes *et al.*, 2010]. Finally, the mature miRNA duplexes are loaded into Ago proteins requiring the chaperone HSP90 [Dueck and Meister, 2014].
The loading of siRNAs and miRNAs is influenced by the length and the 5′-nucleotide in plants and worms [Grishok, 2013; Bologna and Voinnet, 2014]. The loading into *D. melanogaster* Agos is dependent on unpaired regions, bulges, in the miRNA leading to incorporation into Ago1 or the perfect, complementary siRNAs or miRNAs resulting in association with Ago2 [Ha and Kim, 2014]. In mammals all Ago proteins are equally loaded with small RNAs, however, the strand with the less stably paired 5′-end is preferentially selected as the guide strand bound to Ago [Meister, 2013].

PiRNAs

The last class of small RNAs are the Piwi interacting RNAs, called piRNAs, which are conserved in flies, fish and mammals [Meister, 2013] (Piwi proteins see Section 1.2). These 25-33 *nt* long small RNAs are expressed in the germline and generated in a Drosha and Dicer independent pathway. PiRNAs derive either from different piRNA clusters consisting of defective transposon copies or from transposable elements (TEs) including long terminal repeats (LTR), short and long interspersed elements (SINE/LINE) and internal repeats (IR) [Aravin *et al.*,

2003, 2007; Saito *et al.*, 2006; Gunawardane *et al.*, 2007].

PiRNA Biogenesis

For primary piRNA production, one or both strands of these piRNA clusters are transcribed, presumably by a bidirectional promoter, and form long single stranded RNAs which are bound by Piwi proteins and trimmed at the 3′-end by a yet unknown exonuclease [Thomson and Lin, 2009]. Some of these primary piRNAs induce secondary piRNA generation and amplification by the so called Ping-Pong cycle [Brennecke *et al.*, 2007; Saito and Siomi, 2010; Aravin *et al.*, 2007]. A primary antisense piRNA in complex with a Piwi protein binds by complementary base pairing to a sense transcript from the same piRNA cluster or an actively transcribed transposon. Due to its slicing activity the Piwi protein creates a new 5′-end on the sense transcript exactly 10 *nt* away from the binding site of the 5′-end of its own bound piRNA. The sense transcript with its new 5′-end is then bound by another Piwi protein and transformed into a new piRNA by degradation of its 3′-end by a yet unknown exonuclease. Subsequently, the Piwi protein bound to the sense piRNA is targeted to an anti-sense transposon or piRNA cluster transcript, which leads to generation of new anti-sense piRNAs [Brennecke *et al.*, 2007; Aravin *et al.*, 2007]. This mechanism is supported by the finding of pairs of sense and antisense piRNAs which show an overlap of exactly 10 *nt*. In addition, often a preference for a 5′-U of the antisense piRNA is found. While there is no preference observed for the 5′-end of the sense piRNA, which is opposite to the cleavage site of the antisense piRNA and no bias is expected, there is a strong enrichment for an A at position 10 of the sense piRNA. However, this A is complementary to the 5′-U of the antisense piRNA [Brennecke *et al.*, 2007; Aravin *et al.*, 2007; Gunawardane *et al.*, 2007]. After association of the 5′-end with the MID domain of the Piwi protein, the 3′-end of piRNAs is protected by 2′O-methylation involving Hen1 [Kirino and Mourelatos, 2007; Saito *et al.*, 2007; Horwich *et al.*, 2007].

Tudor proteins involved in PiRNA generation

Another important family of proteins involved in the piRNA biogenesis are the Tudor proteins, which are found in flies (i.e. Tudor, PAPI, spnE) and mammals (TDRD1,2,4-7,9) [Siomi *et al.*, 2011; Iwasaki *et al.*, 2015]. Tudor proteins interact

with symmetric dimethyl arginine modifications in the N-terminus of Piwi proteins and are needed for several processes [Iwasaki *et al.*, 2015]. Thus, Tudor is involved in quality control of piRNA production by interacting with Piwi proteins during the Ping-Pong cycle [Iwasaki *et al.*, 2015], whereas TDRD1 is required for correct localisation of Mili (a mouse Piwi protein) in the cytoplasm [Siomi *et al.*, 2011]. Even though many Tudor proteins have been identified and linked to the piRNA pathway, their exact functions still need to be determined.

piRNA Functions

The role of piRNAs in germline development is twofold. In the early phase piRNAs protect the genome of germline cells by silencing (PTGS and TGS) of transposable elements [Aravin *et al.*, 2007; Brennecke *et al.*, 2007]. In a later stage in development of mammalian testis, piRNAs were suggested to associate with mRNAs by imperfect base pairing, thus leading to mRNA degradation. This step seems necessary for the maturation of functional sperm cells, which contain only little mRNA and condensed nuclei [Gou *et al.*, 2014].

1.4 Overview on PTGS at the Model of Human Cells

In the cytoplasm mature miRNA, endo- or exo-siRNA duplexes, processed by many different organism-specific pathways (see Section 1.3), are incorporated into Ago proteins. In most cases they guide the RISC or miRNP by complementary base pairing to the 3′-UTR of the target RNA, however, binding to the 5′-UTR or the ORF has also been observed [Huntzinger and Izaurralde, 2011] (Figure 1.2). When targeted by the RISC or miRNP the capped and poly-adenylated mRNA is in a closed loop conformation awaiting translation by the ribosome. The loop is formed by the mRNA, the eukaryotic translation initiation factor E (eIF4E) recognising the 5′-cap of the mRNA and eIF4G which mediates the interaction with the poly-A binding proteins (PABPC) associated with the 3′-end of the RNA [Huntzinger and Izaurralde, 2011]. The recognition of a target RNA by the guide strand of the RISC or miRNP is accomplished by two distinct ways.
First, small RNAs and also some animal miRNAs are perfectly or nearly perfectly complementary to the target RNA over their entire length. Here, especially the interaction of the bases 10 and 11 of the guide strand with the target transcript enable cleavage of the latter by Ago2 [Huntzinger and Izaurralde, 2011].

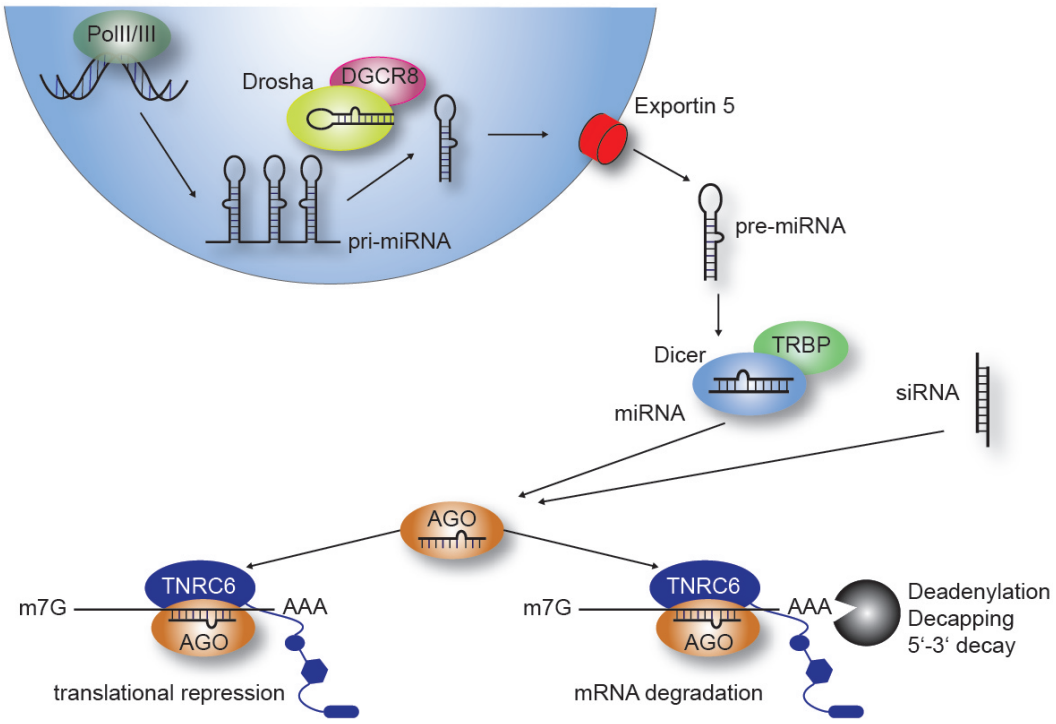


Figure 1.2: ***Overview of the PTGS mechanism in human cells.*** *Most miRNAs are transcribed by RNA PolII into pri-miRNAs. These stem loop containing transcripts are recognized by the Microprocessor complex, consisting of the RNAse-III-like enzyme Drosha and DGCR-8. Via cleavage near the base of the stem, pre-miRNA hairpins are generated by the Microprocessor complex and exported to the cytoplasm by Exportin-5. However, also other Drosha-independent processes for pre-miRNA formation exist. The pre-miRNA hairpins are subsequently bound by Dicer in complex with TRBP which removes the loop structure to yield a miRNA duplex. After miRNA or siRNA loading mediated by chaperones, miRNPs or RISC complexes consisting of an Ago protein and the guide strand of a miRNA or siRNA are formed. The Ago protein is then recruited to a target transcript by complementary base pairing of its bound small RNA. After association of TNRC6 proteins, which interact with an Ago protein and Poly-A binding proteins on the transcript, the translation of the transcript is either repressed or enzymes for transcript degradation are recruited. These comprise the deadenylase complexes Pan2-Pan3 and CAF1-CCR4-NOT, the decapping complexes DCP1/DCP2 and the 5′-3′-specific exonuclease XRN1 [Huntzinger and Izaurralde, 2011].*

The 3′-ends of the sliced RNA are subsequently degraded by the 5′-3′-exonuclease XRN1 starting at the unprotected 5′-end of the fragment [Orban and Izaurralde, 2004]. The second part of the mRNA with a free 3′-end is attacked by the exosome, a multimeric assembly of 3′-5′ exonucleases [Orban and Izaurralde, 2004]. The exosome is also involved in mRNA surveillance pathways like nonsense-mediated decay by which aberrant stop codons in an mRNA are detected result-

ing in the degradation of the latter [Parker and Song, 2004].
Second, most miRNAs recognise their target mRNA mainly by Watson-Crick base pairing of the seed sequence (2-7 or 8 *nt*, see Section 1.2), which is in case of weak interaction assisted by bases towards the 3′-end of the small RNA [Huntzinger and Izaurralde, 2011; Meister, 2013]. Then, one of three TNRC6 (A/B/C) proteins associates with the mRNA bound miRNP (Figure 1.2). These paralogues of the *D. melanogaster* GW182, contain several characteristic glycine-tryptophane (GW) repeats which bind into two tryptophan binding pockets on the surface of Ago proteins [Schirle and MacRae, 2012]. In addition, TNRC6 proteins contain two PABPC-binding domains, which enables them to interact with PABPCs located at the 3′-end of the target transcript. Finally, TNRC6 family proteins recruit the deadenylase complexes Pan2-Pan3 and CAF1-CCR4-NOT to the target mRNA [Braun *et al.*, 2011; Chekulaeva *et al.*, 2011]. Following deadenylation, the mRNA is either stored in a translationally silenced state, which is rather cell type specific [Huntzinger and Izaurralde, 2011], or decapped by DCP1/DCP2 complexes [Chen *et al.*, 2009] (Figure 1.2). The mRNA deprived from its 5′-end is subsequently degraded by the 5′-3′-specific exonuclease XRN1 [Huntzinger and Izaurralde, 2011].

1.5 TGS Pathways in Different Species

The following description of small RNA pathways in different species focuses on mechanisms of TGS in the nucleus and the required interaction partners. As the chromatin network builds the molecular framework for TGS mechanisms the main features and regulators will be introduced.

1.5.1 The Chromatin Environment of TGS

The DNA of a human cell comprises ~$3 * 10^9$ *bp*. In order to store and functionally use the DNA inside the eukaryotic nucleus, several layers of compaction, named chromatin, are needed as well as proteins which reorganise this structure to make DNA accessible for the transcription, replication and repair machinery. The basic building block of chromatin is the nucleosome which consists of ~145 *bp* of DNA wrapped around an octamer of histone proteins. The disc-shaped octamer is assembled by two copies of each histone: H2A, H2B, H3 and H4 [Al-

berts *et al.*, 2002]. Nucleosomes are connected by a short linker DNA to form the 10 *nm*-fibre, which resembles beads on a string [Alberts *et al.*, 2002]. Fibre-fibre interactions then drive the creation of higher compaction levels eventually resulting in condensed chromosomes. Apart from cell division, when condensed chromosomes are formed, chromatin is found in two states: heterochromatin or euchromatin. Even though this classical view is more and more abandoned in current research describing several mixed chromatin states [Luger *et al.*, 2012], it still helps to understand the basic principles. These different states show remarkable differences in DNA methylation status, incorporation of histone variants, post-translational modification (PTM) of histone N-terminal tails and associated proteins. The enzymes responsible of DNA modification are DNA methyltransferases (i.e. DNMT1-3 proteins) which predominantly methylate DNA at the cytosine of CpG dinucleotides [Denis *et al.*, 2011]. Whereas DNMT1 serves mainly by re-methylating hemi-methylated DNA during cell division, DNMT3A and DNMT3B are required for symmetric *de novo* methylation of CpGs at imprinted loci in the germline [Denis *et al.*, 2011]. Lysine residues of protruding histone tails are altered by methyl- and acetyltransferases (HMTs and HATs) and the respective demethylating and deacetylating enzymes (HDMs and HDACs). Other proteins involved in transition between states are nucleosome chaperones for the exchange of histone variants and nucleosome remodelling enzymes which influence the position of nucleosomes along the DNA. Last also transcription factors which bind to specific DNA sequences and prepare a binding site for RNA polymerases and insulator proteins separating heterochromatic from euchromatic regions (CTCF) shape the accessibility of the chromatin [Zhou *et al.*, 2011].

Euchromatin

Euchromatic regions are marked by histone modifications mediating a permissive state. Thus, PTM of histone tails comprising monoubiquitylation of Histone H2B (H2Bub1) accompanied by di/tri-methylation of histone H3 at lysine 4 (H3K4me2, H3K4me3), acetylation (ac) of histone H4 and the histone variant (H2A-Z) are found especially at transcriptionally active promoters [Luger *et al.*, 2012; Teves *et al.*, 2014]. Actively transcribed regions accumulate H3K36me3 and H3K79me2 at the N-terminal tails of histones and H3K4me1, H3K4me2, H3K27ac are enriched at enhancers [Zhou *et al.*, 2011; Luger *et al.*, 2012].

Heterochromatin

In contrast, heterochromatin is described as transcriptionally silent, rather condensed form of chromatin. It is found at both ends of the chromosome, the telomeres, and at the site where sister chromatides are associated, the centromeres. Both regions are rich in repetitive elements like α-satellites and are often referred to as constitutive heterochromatin. In addition, heterochromatin is used by the cell for transcriptional inactivation at the second X chromosome in females of mammals or certain genomic regions silenced in differentiated cells [Alberts *et al.*, 2002]. The heterochromatin is marked by several modifications ranging from DNA-methylation for long term genetic silencing to post-translational modification of histone H3 at H3K9me2/H3K9me3 or H3K27me3 [Zhou *et al.*, 2011]. These modifications recruit distinct chromodomain proteins like heterochromatin binding protein 1 (HP1α, β) associating with H3K9me2 or the polycomb repressive complex 1 (PRC1) which interacts with H3K27me3 [Grewal and Elgin, 2007]. HP1 is a binding platform for HMTs and HDACs which cause spreading and maintenance of heterochromatin, a mechanism conserved from yeast to humans [Grewal and Elgin, 2007]. The polycomb repressive complex 2 (PRC2), which incorporates the histone methyltransferase EZH2, is, in contrast, recruited by deacetylation of H3K27 and causes its methylation to H3K27me3. Then a PRC1 complex associates with H3K27me3 catalysing monoubiquitylation of histone H2A at lysine 119, another silencing mark, and formation of heterochromatin [Chen and Dent, 2014]. In addition, also specific histones are found in heterochromatin for example macroH2A which accumulates on the inactivated X chromosome or the centromere specific Cenp-A (CenH3) which is needed for sister chromatid separation [Luger *et al.*, 2012].
A mechanism tightly linked to chromatin is transcription. The nucleosome and higher compaction levels of chromatin form a barrier to RNA Polymerases, which they have to overcome in order to enable mRNA and protein production in a cell (Section 1.5.2).

1.5.2 Transcription

In eukaryotes there are three RNA polymerases which transcribe DNA into RNA. RNA PolI and III are engaged in the transcription of ribosomal rRNA genes. RNA PolIII is in addition required for formation of tRNA and several small RNAs.

RNA PolII transcribes all protein coding genes as well as many small and long non-coding RNAs [Alberts *et al.*, 2002]. The activity of RNA PolII is mainly regulated by the phosphorylation state of its C-terminal domain (CTD), which contains multiple heptad repeats (YSPTSPS) [Hsin and Manley, 2012]. For initiation of transcription the unphosphorylated RNA PolII associates with a complex of general transcription factors (TFII A, B, D, E, F, and H) bound to the promoter region of a gene [Hsin and Manley, 2012]. After binding to the promoter, RNA PolyII is phosphorylated at serine 5 and starts to synthesise a short stretch of RNA assisted by TFII H, which comprises DNA helicase activity and unwinds the DNA double strand for RNA PolII to enter [Hsin and Manley, 2012]. Then RNA PolII is stalled at the promoter till elongation is initiated. During elongation phosphorylation at serine 2 increases while it is successively removed from serine 5 [Hsin and Manley, 2012]. In addition to these, more modifications of the CTD have been identified which are reviewed in Hsin and Manley [2012].

For transcription RNA polymerases need to progress through DNA organised in arrays of nucleosomes. In the case of RNA PolII nucleosomes are translocated to yield way for the processing polymerase by ATP dependent remodellers [Studitsky *et al.*, 2004]. The detailed mechanisms of nucleosome remodelling required for transcription by RNA PolI to III remain poorly understood. Transcription by RNA PolII leads to local melting of the DNA double strand at the promoter creating positive supercoils in front and negative supercoils behind. Positive supercoiling destabilises nucleosomes [Teves *et al.*, 2014]. However, for efficient transcription RNA PolII is dependent on nucleosome remodellers, histone chaperones, histone modifying enzymes and elongation factors. Facilitates Chromatin Transcription (FACT) is a nucleosome chaperone which belongs to a set of proteins minimally required for RNA PolII transcription through chromatin *in vitro* [Teves *et al.*, 2014]. It is activated by the monoubiquitylation of histone H2B (H2Bub1), which is a mark for active promoters as described in Section 1.5.1. FACT is suggested to weaken histone-DNA contacts to enable RNA PolII progression, however, the exact mechanism is controversially discussed. Also, ATP-dependent remodelling complexes are implicated in transcription by RNA PolII but due to their number and redundancy their precise role could not be determined. There are four main families of chromatin remodellers comprising SWI/SNF, ISWI, INO80, and CHD. Whereas SWI/SNF proteins were suggested to evict H2A-H2B and whole nucleosomes, ISWI remodellers were found to as-

semble nucleosomes and induce their regular spacing [Teves *et al.*, 2014]. INO80 and CHD proteins serve also regular spacing of nucleosomes, while INO80 is additionally involved in deposition and removal of H2A-Z at TSS nucleosomes [Teves *et al.*, 2014]. Therefore, RNA PolII was recently suggested to play an important role for correct nucleosomal spacing in transcribed regions of the genome [Struhl and Segal, 2013].
A chromatin silencing mechanism which is dependent on transcription for recruitment of its subunits is found at the centromeres of fission yeast and is described in the next section.

1.5.3 Fission Yeast a Role Model for TGS

Nearly 13 years ago a study conducted by Volpe *et al.* [2002] linked heterochromatin formation in *S. pombe* to RNAi. Heterochromatin formation has been found at three sites in the fission yeast chromatin: at centromeres, telomeres and the mating type locus. While an RNAi based mechanism is essential for the establishment of heterochromatin at centromeres, it is redundant for heterochromatin formation at telomeres and the mating type locus [Aravind *et al.*, 2000; Volpe *et al.*, 2002; Woolcock and Bühler, 2013]. The role of RNAi in heterochromatin formation at pericentric repeats is illustrated in Figure 1.3.

Small RNA induced heterochromatin formation

Recognition of the target locus is mediated by the RNA induced transcriptional silencing (RITS) complex, which consists of an Ago protein (Ago1), Chp1 and Tas3 [Verdel *et al.*, 2004]. Fission yeast contains only one Ago protein, which is essential and is loaded with an siRNA assisted by the Ago siRNA chaperone complex (ARC). Chp1 is a chromodomain protein (see Section 1.5.1), which stabilises the binding of Ago1 to pericentric chromatin by interacting with methylated H3K9 [Holoch and Moazed, 2015]. However, Chp1 is not only binding to but is also required for H3K9 methylation at centromeres as deletion of Chp1 leads to loss of this modification from histone H3 [Sadaie *et al.*, 2004]. Tas3 is a glycine tryphtophan (GW) containing protein which binds Ago1 by its GW motif and interacts with the C-terminal half of Chp1 by its N-terminal domain and thus stabilises the trimeric protein complex [Schalch *et al.*, 2011]. RITS is loaded with an siRNA generated by Dcr1 from transcripts of non-coding repetitive elements [Reinhart

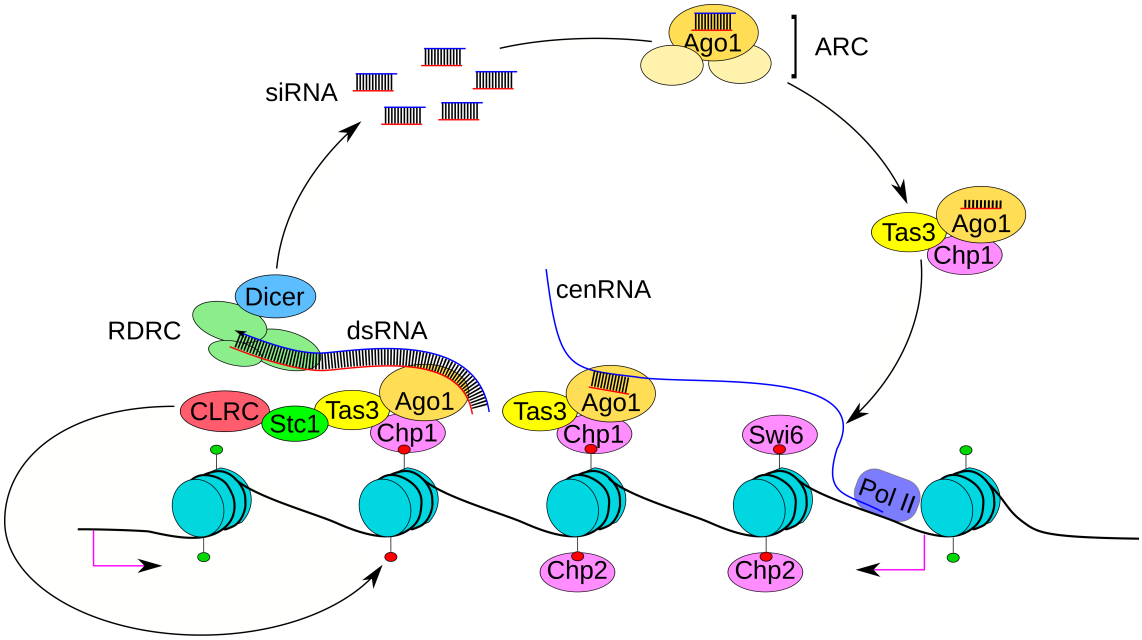


Figure 1.3: ***The role of RNAi in heterochromatin formation in fission yeast.***
At yeast centromeres an RNAi-dependent mechanism is required for heterochromatin formation. A small RNA processed by the RNAase III-enzyme Dcr1 is loaded into Ago1 by the chaperone complex ARC. Ago1 complexes with the chromodomain protein Chp1 and the GW containing Tas1 protein and forms the RNA-induced silencing complex (RITS). The RITS is recruited by complementary base-paring interaction of its bound siRNA and a nascent RNA PolII transcript to the pericentric chromatin. Mediated by Stc1 the RITS complex associates with the histone methyltransferase Clr4. Subsequent methylation at H3K9 results in binding of the chromodomain proteins Chp1/2 and Swi6. In addition, the RITS complex interacts with the RNA-dependent RNA Polymerase Complex (RDRC) consisting of Rdp1, Hrr1 and Cid12. This complex synthesises a second strand to the Ago1 bound nascent transcript. This double stranded RNA is processed by Dcr1 and loaded into the RITS complex causing a positive feedback loop, which leads to further spreading of the heterochromatin along the centromeres. The silencing mark H3K9me2/3 and the recruitment of heterochromatin protein 1 homologues Swi6 and Chp2 is required for heterochromatin formation and its stable maintenance. (Similar to [Holoch and Moazed, 2015])

and Bartel, 2002; Volpe *et al.*, 2002; Verdel *et al.*, 2004] and associates with chromatin by base pairing interactions of the Ago-bound small RNA and a nascent transcript [Verdel *et al.*, 2004; Bühler *et al.*, 2006]. Then the Clr4 methyltransferase complex (CLRC) is recruited by RITS. This step is mediated by Stc1, which has been shown, to interact with RITS and CLRC [Bayne *et al.*, 2010]. Stc1 consists of two zinc finger domains ZF1 and ZF2 implicated in Ago1 interaction and an unstructured C-terminal domain, which is sufficient for binding of the CLRC [He *et al.*, 2013]. The subunit Clr4, the sole H3K9 methyltransferase known in fission yeast [Aravind *et al.*, 2000], mediates the H3K9 methylation enabling the binding of the chromodomain proteins Chp1, Swi6 and Chp2. Swi6 and Chp2 (HP1 homologues) maintain the H3K9 methylation and heterochromatin [Bannister *et al.*, 2001; Sadaie *et al.*, 2004].
Finally, the RNA-dependent RNA polymerase complex (RDRC) is recruited to the nascent transcript, which contains the RNA-dependent RNA polymerase (Rdp1), a helicase Hrr1 and the non-canonical poly(a) polymerase Cid12 [Motamedi *et al.*, 2004]. Hrr1 is highly similar to DEAD box RNA helicases of the Smg2 family, which have been linked to the RNAi response also in other species [Motamedi *et al.*, 2004; Dalmay *et al.*, 2001]. The Cid12 protein contains a nucleotidyltransferase domain and belongs to the family of Trf4/Trf5 poly(A) polymerases [Halic and Moazed, 2010]. A second RNA strand is synthesised to the RITS bound nascent transcript by the RDRC complex (see Figure 1.3). The resulting dsRNA is then processed by Dcr1 into siRNAs which are in turn loaded into Ago1 causing further spreading of the heterochromatin [Bühler *et al.*, 2006].

Initiation of siRNA Mediated Heterochromatin Formation

The primary trigger for RITS has recently been discovered by Halic and Moazed [2010]. The ssRNAs incorporated into RITS can be supplied by two different mechanisms both dependent on slicer activity of Ago1. The first one is based on priRNAs, which are supposed, to stem from degradation products of centromeric transcripts. These are degraded by the exosome, loaded into Ago1 [Holoch and Moazed, 2015] and trimmed to the required length by the 3′-5′-exonuclease Triman [Marasovic *et al.*, 2013]. The loaded Ago1 assembles into a RITS complex leading to H3K9 methylation by recruitment of CLRC. In the second suggested pathway, priRNA loaded Ago1 binds to Tas3 and forms a primer for RDRC on centromeric transcripts [Halic and Moazed, 2010]. The dsRNAs produced are

then further processed by Dicer and incorporated into Ago1. According to Halic and Moazed [2010] Ago1 may be part of a RNA surveillance pathway, detecting low level aberrant transcripts from repetitive elements, leading to genomic silencing. However, more studies have to be performed, to elucidate the ribonucleases and trimming enzymes involved in this process.

1.5.4 Silencing by TGS in *A. thaliana*

A. thaliana comprises 10 different Ago proteins, which can be grouped into three different clades: Ago1, 5, 10; Ago2, 3, 7 and Ago4, 6, 8, 9 [Mallory and Vaucheret, 2010]. The loading of Ago proteins follows certain loading preferences for the identity of the 5′ nucleotide and for the length of the incorporated small RNA strand [Bologna and Voinnet, 2014]. However, TGS is mediated solely by the Ago4 clade of proteins, as piRNAs are absent from plants [Matzke and Mosher, 2014].

Generation of small RNAs for the class of Ago4 proteins

The small RNAs derive mainly from transposons and dispersed repeats in an RNA PolIV and RNA-dependent RNA Polymerase 2 (RDR2) dependent process [Zhang *et al.*, 2007]. RNA PolIV and RDR2 associate with the SNF2-like chromatin-remodelling factor classy 1 (CLSY1) (see Figure 1.4 and Section 1.5.2), which supports their localisation and interaction in the nuclear compartment [Smith *et al.*, 2007]. RNA PolIV and RDR2 knockout plants demonstrated, that the precursors of 90% of the 24 *nt* siRNAs involved in TGS are dependent on RNA PolIV transcription and overlap with the siRNA set, whose biogenesis relies on RDR2 [Zhang *et al.*, 2007]. Restoring of wild-type RNA PolIV by crossing of knockout plants (nrpd1a/1b and nrpd2a/2b both RNA PolIV subunits) deficient for 24 *nt* siRNAs led to full re-establishment of 24 *nt* siRNAs in the F1 generation, suggesting an siRNA-independent recruitment mechanism for RNA PolIV to its genomic targets [Zhang *et al.*, 2007]. One of those factors recruiting RNA PolIV to a subset of loci is SHH1 (Figure 1.4), which binds methylated H3K9 in combination with unmodified H3K4 [Law *et al.*, 2013]. The RDR2 derived dsRNAs are then recognised and cleaved into 24 *nt* long small RNAs by the ribonuclease Dicer-like 3 protein (DCL3). After siRNA export to the cytoplasm Ago4 is loaded with small RNAs assisted by the chaperone HSP90 (see Figure 1.4). The class

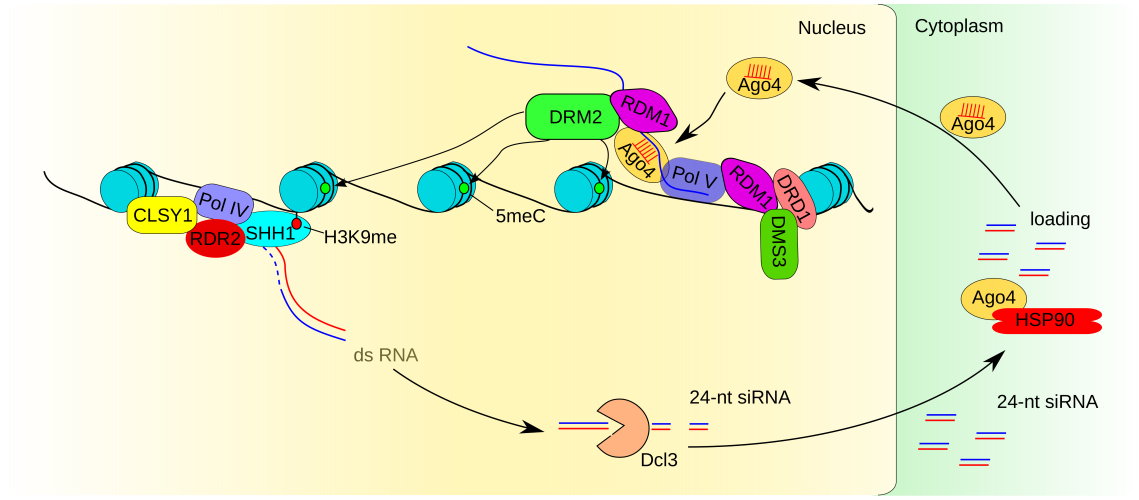


Figure 1.4: ***TGS mechanism in plants.***
RNA PolIV is recruited to chromatin by proteins like SHH1, where it starts transcription. Subsequently, it associates with the remodeller CLSY1 and RDR2, which synthesises as second RNA strand resulting in dsRNA molecule. This is cleaved into 24 nt siRNAs by DCL3, which are then exported to the cytoplasm for loading into Ago proteins. Following HSP90-dependent small RNA loading Ago4 translocates into the nucleus and is targeted by complementary base pairing of is bound siRNA and a nascent RNA PolV-dependent transcript to loci destined for silencing. Ago4 together with RDM1 recruit the DNA methyltransferase DRM2 which in turn de novo methylates DNA and thus initiates heterochromatin formation at the target loci (similar to Matzke and Mosher [2014]).

of Ago4 proteins show a strong preference for an adenine at the 5′-end of their associated siRNAs [Bologna and Voinnet, 2014].

In addition, also siRNAs with a length of 21 and 22 *nt* are produced from trans-acting small RNA producing (TAS) loci and transposable elements [Wu *et al.*, 2012; McCue *et al.*, 2014]. Firstly, transcripts of the TAS loci are targeted by specific miRNAs bound to Ago1 and sliced. These cleaved transcripts are stabilised by suppressor of gene silencing 3 (SGS3), which recruits the RNA-dependent RNA Polymerase 6 (RDR6) producing dsRNAs [Matzke and Mosher, 2014]. Secondly, RNA PolII dependent transcripts from transposable elements are also transcribed into dsRNA by an RDR6-dependent process. Both types of RDR6 dependent dsRNAs are further processed mainly by DCL4 into siRNAs serving in PTGS by Ago1 and DCL1 into 21 *nt* siRNAs bound to Ago4/6 [Matzke and Mosher, 2014; Wu *et al.*, 2012; McCue *et al.*, 2014].

Ago4 and Ago6 mediated TGS

TGS is then mediated by a process referred to as RNA-dependent DNA methylation (RdDM), which relies on RNA PolV as a binding platform for Ago4 and Ago6 [Matzke and Mosher, 2014]. The recruitment of RNA PolV to chromatin is suggested to depend in part on SUVH2 and SUVH9, which are inactive members of the SU(VAR)3–9 histone methyltransferase family (Section 1.5.1) but are still able to recognize pre-existing methylated DNA [Matzke and Mosher, 2014]. In addition, RNA PolV is also found at sites which lack pre-existing methylation and triggers *de novo* methylation [Matzke and Mosher, 2014] (compare Figure 1.4). Furthermore, RNA PolV associates with chromatin by binding to the DDR complex, which consists of the putative chromatin remodellers DRD1, DMS3 and the plant specific single-stranded DNA-binding protein RDM1 [Matzke and Mosher, 2014]. At the chromatin RNA PolV synthesises transcripts that seem to serve as a scaffold for siRNA-mediated binding of Ago4/6 (see Figure 1.4), as the transcripts are 5′-capped but deficient for poly(A) tails [Matzke and Mosher, 2014; Wierzbicki *et al.*, 2008]. Following recruitment of Ago proteins, the complex is stabilised by the association of Ago4/6 with GW repeats in the large subunit NRPE1 (NRPD1b + NRPD2a) of RNA PolV similar to human TNRC6-Ago interactions [El-Shami *et al.*, 2007] (explained in Section 1.4). Expression of a catalytically inactive Ago4 mutant in an Ago4 knockout background revealed that slicer activity is only partially required for silencing by RdDM [Qi *et al.*, 2006]. Ago4/6 in complex with the RNA PolV associated RDM1 then recruit the DNA methyltransferase DRM2, which belongs to the DNMT3 family of proteins [Xie *et al.*, 2012; Naumann *et al.*, 2011; Kuhlmann *et al.*, 2014] (see Section 1.5.1). Similar to DNMT3A in mammals DRM2 mediates *de novo* cytosine DNA methylation at symmetric CG (see Figure 1.4), but also at symmetric CHG (H is adenine, thymine or cytosine) and asymmetric CHH sites, which are confined to the binding region of the siRNA-Ago4/6 complex [Naumann *et al.*, 2011; Kuhlmann *et al.*, 2014]. The TGS by RdDM is probably reinforced by the histone methyltransferases SUVH5 and SUVH6, which bind methylated DNA and mediate heterochromatin formation by histone H3K9 methylation. In turn H3K9 methylation seems to regulate biogenesis of 24 *nt* siRNAs, which recruit RdDM and DRM2, to build a self reinforcing loop similar to the mechanism described for fission yeast [Stroud *et al.*, 2014] (compare Section 1.5.3).

TGS in the Plant Germline

The last *A. thaliana* Ago to mention from this clade is Ago9, which is involved in regulation of megaspore development from somatic ovule tissue [Olmedo-Monfil *et al.*, 2010]. On the one hand, it prevents multiple differentiation of somatic ovule cells into megaspores which further divide and give rise to the female gametophyte including the germline. On the other hand, it may silence transposons in female gametes and accessory cells. In addition, Ago9 is expressed in anthers and seed coats.

The Role of TGS in Heterochromatin Formation

RdDM seems to take place only at intergenic transposons and genes containing transposons, where it serves especially in *de novo* methylation of these loci as knockout plants for DCL3, RDR2 and Ago4 confirmed [Chan *et al.*, 2004]. Unlike *S. pombe* pericentromeric heterochromatin formation is established and maintained mostly by RNAi-independent processes [Matzke and Mosher, 2014]. Further functions for Ago-mediating RdDM were implicated in pathogen defence, stress response and genomic imprinting [Matzke and Mosher, 2014; Vu *et al.*, 2013].

1.5.5 TGS Mechanisms in Drosophila

The *D. melanogaster* genome encodes for five Argonaute proteins, Ago1 and Ago2, which are found in all cells, and Piwi, Aub (Aubergine) and Ago3, which are predominantly located in the germ line [Hutvagner and Simard, 2008]. Of these proteins only Ago2 and Piwi were implicated in TGS mechanisms. Ago2 associates with exogenous siRNAs produced by Dicer2 (Dcr2) in combination with the dsRNA-binding protein R2D2. The latter mediates efficient loading of siRNAs into Ago2 [Liu *et al.*, 2003]. The siRNA specificity of Ago2 was shown by Okamura *et al.* [2004], generating Ago1 and Ago2 knockout flies. Ago1 knockouts were defective in miRNA-mediated translational repression, late embryonic lethal and sterile. In contrast, Ago2 depletion inhibited siRNA RISC formation but had no influence on fertility. Additionally, Ago2 is loaded with endogenous-siRNAs, called endo-siRNAs, resulting in the PTGS and potential TGS of target TEs in somatic tissue [Kawamura *et al.*, 2008; Ghildiyal *et al.*, 2008]. This is further supported by Dcr2 knockdown cells and Dcr2 mutant ovaries, which lack endo-

siRNAs and show increased transcription of some TEs [Kawamura *et al.*, 2008]. Furthermore, Ago2 knockdown correlated with increased TE expression level.

Potential Ago2-mediated TGS

Several reports suggest that Ago2 has also a nuclear function in addition to the cytoplasmic control of transcripts (PTGS) from repeat elements [Deshpande *et al.*, 2005; Peng and Karpen, 2007]. In the nucleus, Ago2 is supposedly needed for pericentric heterochromatin formation during embryonic development [Deshpande *et al.*, 2005]. Applying immunofluorescence on Ago2 mutant fly embryos, defects in chromosome condensation, nuclear kinesis and assembly of spindle apparatus could be detected [Deshpande *et al.*, 2005]. In addition, silencing of transgenes inserted in percentric regions was also diminished in heterozygous mutants [Deshpande *et al.*, 2005]. Thus, both observations could be explained by a role of Ago2 in pericentric heterochromatin formation but direct evidence is missing. In a later study the RNAi machinery was suggested to be involved in nucleoli organisation [Peng and Karpen, 2007]. Assessing nucleoli number and H3K9me2 in Ago2 and Dcr2 mutant larvae Peng and Karpen [2007] found increased numbers of nucleoli and reduced levels of H3K9me2 at rDNA loci. This links the RNAi machinery to a histone mark of heterochromatin, which was shown to be dependent on the histone methyltransferase Su(var)3–9 [Peng and Karpen, 2007]. While these findings are reminiscent of TGS pathways components described for *A. thaliana*, *C. elegans* and *S. pombe*, a self-reinforcing loop is missing as flies lack an RNA-dependent RNA polymerase [Bologna and Voinnet, 2014; Lee *et al.*, 2004b; Bühler *et al.*, 2006] (see also Section 1.5.3, Section 1.5.4, Section 1.5.6). Cernilogar *et al.* [2011] detected Ago2 also in the nucleus of somatic cells. However, using ChIP and immunofluorescence experiments they found Dcr2 and Ago2 at active transcription sites potentially regulating RNA PolII promoter pausing under heat shock conditions, thus opposing a role of Ago2 in TGS. Thus, the involvement of Ago2 in a TGS mechanism remains controversial.

TGS in the Germline of Flies

The second important RNAi pathway is mediated by the Piwi family of proteins, Piwi, Aub and Ago3, which are predominantly expressed in the gonads [Hutvagner and Simard, 2008; Saito *et al.*, 2006]. In detail, Aub and Ago3 are only

expressed in the cytoplasm of germ cells of both genders [Saito and Siomi, 2010]. There, Aub and Ago3 function in PTGS and degradation of transcripts derived from repetitive elements [Li *et al.*, 2009a; Nishida *et al.*, 2007; Gunawardane *et al.*, 2007; Brennecke *et al.*, 2007]. In contrast, Piwi is found in the nucleus of ovaries (somatic and germ cells) [Brennecke *et al.*, 2007]. However, in testis it is restricted to somatic niche cells, which maintain the germ cells [Saito and Siomi, 2010]. According to the current model, Piwi bound to a piRNA enters the nucleus and is targeted by complementary base pairing to nascent transcripts from transposable element encoding loci [Huang *et al.*, 2013b; Tianfang Ge and Zamore, 2013]. There, it associates with the C-terminus of heterochromatin protein 1a (HP1a) by direct interaction, which was shown by Brower-Toland *et al.* [2007] using yeast two-hybrid assays. Applying Piwi mutants defective in nuclear localisation and slicer activity Sienski *et al.* [2012] showed that silencing of certain TEs is dependent on nuclear Piwi, but not on its slicer function. In addition, Piwi seems to recruit the histone methyltransferase Su(var)3–9, which di-methylates histone H3 (H3K9me2) [Huang *et al.*, 2013b]. This modification is in turn bound by HP1a which results in heterochromatin formation and depletion of RNA PolII from respective loci [Huang *et al.*, 2013b]. These findings are further supported by Piwi knockdown in ovaries combined with transcript, RNA PolII and H3K9me3 ChIP analyses. Here, Le Thomas *et al.* [2013] found that TEs show increased transcription rates, RNA PolII recruitment and reduced histone H3K9 methylation levels. Primary piRNAs seem to derive from heterochromatic piRNA clusters which are not targeted for silencing by Piwi (reviewed in [Tianfang Ge and Zamore, 2013]). Finally, opposing results based on HP1 ChIP experiments were presented by Moshkovich and Lei [2010], which indicate an endo-siRNA and piRNA independent mechanism for heterochromatin formation at TE encoding loci. However, due to the conclusive findings by Brower-Toland *et al.* [2007]; Sienski *et al.* [2012]; Le Thomas *et al.* [2013]; Huang *et al.* [2013b] arguing in favour of a Piwi dependent mechanism the results by Moshkovich and Lei [2010] are strongly questioned by Tianfang Ge and Zamore [2013].

1.5.6 Worms, Where RNAi was Discovered

In *C. elegans* there are 26 Argonaute proteins described [Dueck and Meister, 2014], which function in processes ranging from TGS to epigenetic memory. Associated with these Ago proteins are several classes of small RNAs, which can be grouped

due to their length (21, 22, 26 *nt*), the 5′- and 3′-modifications and the identity of the 5′-base [Grishok, 2013]. These are miRNAs, 22G 5′-tri-phosphorylated siRNAs, 21U RNAs also known as worm piRNAs and 26G endo-siRNAs. However, only four of these 26 Argonaute proteins are involved in regulation of transcription in the nucleus. These can be divided in nuclear RNAi defective 3 (NRDE3) involved in TGS in somatic cells and heritable RNAi defective (HRDE1/WAGO9) together with PRG1 and CSR1 which influence heterochromatin formation and transcription in the germline (see Figure 1.5).

TGS in Somatic Cells by NRDE3

In somatic cells of *C. elegans* TGS is catalysed by the NLS-containing Argonaute NRDE3 [Guang *et al.*, 2008]. For nuclear import to proceed, NRDE3 needs to associate with small RNAs in the cytoplasm [Guang *et al.*, 2008]. These so-called secondary siRNAs derive from exogenous RNAs or endogenous mRNAs, which are targeted by the Ago proteins RDE1 or ERI1 loaded with primary, Dicer1-dependent exo- or endo-siRNAs. Then, the Ago-targeted RNAs are bound by an RNA-dependent RNA Polymerase complex (RdRP), generating dsRNAs which yield secondary 22G 5′-tri-phosphorylated siRNAs [Guang *et al.*, 2008] (compare Figure 1.5). Following nuclear import, secondary siRNA-loaded NRDE3 associates with pre-mRNA transcripts and recruits NRDE2, a serine/arginine-rich protein [Guang *et al.*, 2010]. Subsequent binding of the protein NRDE1 to the siRNA-targeted pre-mRNA causes TGS by inhibition of RNA-PolII transcription during the elongation phase and removal of the polymerase [Guang *et al.*, 2010; Burkhart *et al.*, 2011] (see Figure 1.5). In addition to inhibition of RNA-PolII, also NRDE4-dependent recruitment of NRED1 to chromatin and methylation of H3K9 was observed at target gene loci of NRDE3 [Burkhart *et al.*, 2011]. NRDE2/3-dependent nuclear TGS resulting in H3K9me marks at the pre-mRNA-encoding genes was further confirmed on a genome wide scale by Gu *et al.* [2012]. Furthermore, the histone methyltransferase SET-25 and the HP1 homologue (HPL2), which binds to methylated H3K9, were shown to be involved in stable heterochromatin formation in somatic cells [Ashe *et al.*, 2012; Shirayama *et al.*, 2012; Juang *et al.*, 2013]. In summary, a close resemblance with the TGS mechanisms observed in yeast and plants (compare Section 1.5.3, 1.5.4) is seen, which differ mainly by the localisation and mechanism of secondary siRNA production by an RdRP.

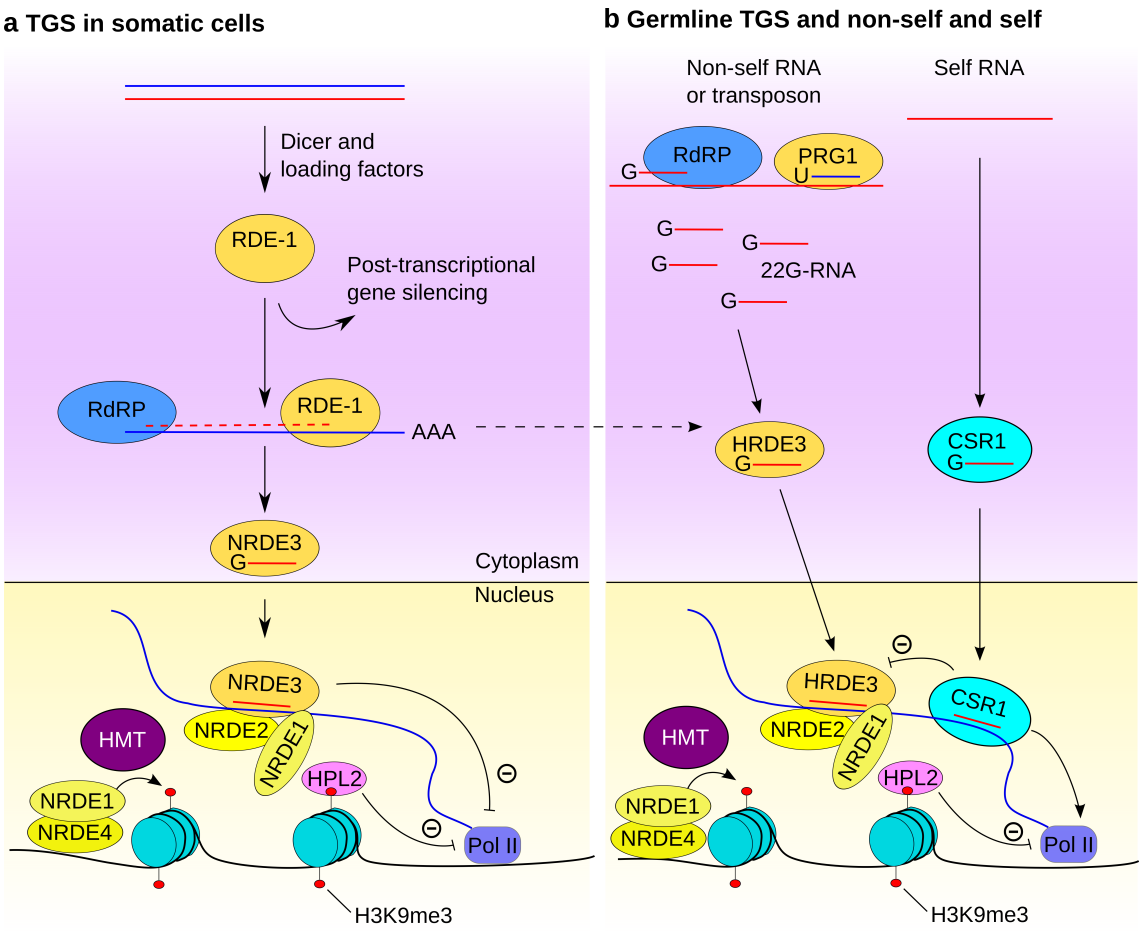


Figure 1.5: ***TGS mechanisms in C. elegans.***
a) Exogenous or endogenous double stranded RNAs are bound by Dicer and processed into primary siRNAs, which are loaded into the Ago homologue RDE1. For generation of secondary siRNAs (22G RNAs), RDE1 binds to a complementary transcript and recruits an RdRP. After synthesis of a complementary strand, secondary siRNAs are produced which are loaded into NRDE3. Following nuclear import, NRDE3 associates with NRDE2 and NRDE1 to inhibit RNA PolII transcription. Subsequently, a complex of NRDE1 and NRDE4 recruits histone methyltransferases, which initiate heterochromatin formation requiring the HP1 homologue HPL2. b) In the germline, the loading and targeting of nascent transcripts and the encoding gene is basically similar to somatic cells, here the Ago protein HRDE3 is involved. However, secondary siRNAs (22G RNAs) can additionally derive from foreign RNAs, which are recognised by the Ago homologue PRG1 bound to 21U RNAs, and are also generated by an RdRP dependent process. Furthermore, another Ago homologue CSR1 interacts with 22G RNAs, which are complementary to endogenous transcripts. After nuclear import, CSR1 protects its complementary loci from silencing by inhibiting HRDE1-mediated processes. This means that HRDE1 in collaboration with CSR1 establish a germline memory of self and non-self sequences. (Adapted with minor modifications from [Holoch and Moazed, 2015].)

TGS in the Germline

The Ago protein HRDE1 (WAGO9) is responsible for TGS of transposable elements in the germline [Gu *et al.*, 2009; Yigit *et al.*, 2006; Grishok, 2013]. Similar to NRDE3, HRDE1 is loaded with secondary siRNAs derived from endogenous or exogenous RNA targets of RDE1 or ERI1 as described above. Additional secondary siRNAs incorporated into HRDE1 are generated by a mechanism dependent on the Ago homologue PRG1, which is loaded with *C. elegans* piRNAs (21U RNAs) (see Figure 1.5). Reminiscent of *D. melanogaster* piRNA clusters (see Section 1.5.5), 21U RNAs are transcribed by RNA PolII from two clusters on chromosome IV and other sites in the germline [Grishok, 2013]. Every 21U RNA-encoding sequence contains a conserved promoter sequence 34 *bp*-upstream from the start [Ruby *et al.*, 2006; Cecere *et al.*, 2012]. They associate exclusively with the Piwi homologue PRG1 from *C. elegans*, which was shown by animals carrying a mutant PRG1 gene [Batista *et al.*, 2008; Das *et al.*, 2008]. These piRNAs are independent of Dicer processing and protected by HENN1-mediated methylation at the 3′ end [Grishok, 2013]. Following targeting of a transcript by 21U RNAs, 22G 5′-tri-phosphorylated secondary siRNAs complementary to the target at regions upstream of the 21U binding site are produced involving an RdRP complex [Ashe *et al.*, 2012; Bagijn *et al.*, 2012; Luteijn *et al.*, 2012].

After loading with a secondary siRNA in the cytoplasm, HRDE1 is imported into the nucleus and targeted by complementary base paring to nascent transcripts (see Figure 1.5). There, it interacts with NRDE2 and NRDE1 and inhibits RNA PolII transcription. In addition, NRDE4 recruits NRDE1 to the HRDE1-associated chromatin, which results in induction of H3K9me3 by the histone methyltransferases SET-25 and SET-32 [Ashe *et al.*, 2012; Buckley *et al.*, 2012]. Furthermore, the HP1 ortholog HPL2 was shown to be involved in heterochromatin formation at the targeted loci [Ashe *et al.*, 2012]. Interestingly, PRG1-dependent secondary siRNAs are only needed for initiation of heterochromatin formation at target loci, while trans-generational silencing requires HRDE1, NRDE2, HPL-2 and the methyltransferases SET25/32 [Ashe *et al.*, 2012; Shirayama *et al.*, 2012; Bagijn *et al.*, 2012; Luteijn *et al.*, 2012].

Self and Non-Self RNA Mechanism

There is a big class of endogenous secondary siRNA, 22-G RNAs, found especially in the germline of *C. elegans*. These small RNAs are generated by processes depending on RdRP complexes and can be further divided in WAGO-associated small RNAs, which target mRNAs by PTGS and TGS, and CSR1-bound 22G RNAs, which are complementary to several thousand protein coding genes of *C. elegans* [Gu *et al.*, 2009; Claycomb *et al.*, 2009]. However, there was no silencing effect observed for the CSR1-associated endo-siRNAs. On the one hand, Claycomb *et al.* [2009] found that CSR1 is targeted to chromatin and needed for chromosome segregation in mitosis and meiosis. In addition, CSR1 was implicated in 3′-UTR processing of histone pre-mRNAs, in part explaining the lethal phenotypes of CSR1 pathway mutants and chromatin condensation defects in CSR1 knockdown experiments [Avgousti *et al.*, 2012]. On the other hand, Cecere *et al.* [2014] analysed the role of CSR1 in regulation of gene expression and found that the CSR1 promoted sense transcription by RNA PolII. Its loss, however, led to reduced expression of many target genes [Claycomb *et al.*, 2009; Avgousti *et al.*, 2012].

Combining the knowledge on PRG1 from the previous section with the findings on CSR1, a model for differential target regulation could be deduced [Seth *et al.*, 2013]. In depth analysis of 21U RNA targeting mechanism revealed that only partial complementarity is required for initiation of the TGS pathway, thus allowing the recognition of a broad spectrum of non-self RNAs by PRG1 [Shirayama *et al.*, 2012]. In contrast, CSR1 counteracts heterochromatin formation and promotes expression of genes in the germline [Seth *et al.*, 2013]. From these findings, a model was deduced, which suggests that 21U RNAs bound to PRG1 are able to target any foreign transcript and supply HRDE1 with complementary siRNAs leading to TGS of the foreign RNA and its encoding sequence. CSR1 bound 22G RNAs, however, protect protein coding genes of *C. elegans* from silencing by HRDE1 [Luteijn *et al.*, 2012; Shirayama *et al.*, 2012; Seth *et al.*, 2013] (compare Figure 1.5). As the silencing and protecting/activating signal are transgenerationally transmitted, they are named small RNA-induced epigenetic silencing (RNAe) and small RNA-induced epigenetic gene activation (RNAa), respectively [Seth *et al.*, 2013]. However, still many open questions remain regarding the phenotype of animals mutant for PRG1 and CSR1 and their role in TGS.

1.5.7 TGS Pathways in Mammals

A prerequisite for TGS in human cells in the nuclear localisation of Ago proteins. This was conclusively shown in cellular fractionation experiments [Rüdel *et al.*, 2008; Robb *et al.*, 2005; Chu *et al.*, 2010] and also *in vivo*, by using GFP-tagged Ago2 in combination with cross-correlation spectroscopy [Ohrt *et al.*, 2008]. With this technique a ratio of 4:1 for cytoplasmic versus nuclear Ago2 was suggested [Ohrt *et al.*, 2008]. For cytoplasmic RISC (cRISC) a molecular weight of 3 MDa was estimated, while the size of nuclear RISC (nRISC) was determined with 158 *kDa*. In order to test the siRNA loading of cRISC and nRISC, experiments using microinjection of siRNAs in combination with wheat germ agglutinin (WGA), which inhibits nuclear transport [Finlay *et al.*, 1987], were performed [Ohrt *et al.*, 2008]. These experiments separated the cytoplasmic protein machinery from the one in the nucleus and suggested that siRNA loading of both RISCs is restricted to the cytoplasm [Ohrt *et al.*, 2008]. Applying MS analysis on purified Ago-complexes led to the identification of Importin 8 (Imp8). While initial hints for an Ago import function of Imp8 were observed, the main function of Imp8, however appears to be the cytoplasmic targeting of RISC to specific target mRNAs [Weinmann *et al.*, 2009].
In addition, also slicing activity of Ago2 in the nucleus was proposed by experiments targeting 7SK snRNA [Robb *et al.*, 2005]. Using siRNA against the nuclear 7SK snRNA, they showed efficient cleavage of the target RNA presumably by an Ago2-dependent mechanism. This is supported by the observation of a correctly sized cleavage product in nuclear and cytoplasmic fractions and *in vitro* tests with nuclear and cytoplasmic lysates, which were incubated with 5′-labelled target RNAs.

Transcriptional Gene Regulation in Somatic Cells Induced by siRNAs

One step further in elucidating a potential role of nuclear Ago in transcriptional regulation was taken by several publications starting form 2004, which applied exogenous siRNAs directed against promoter sequences (see Table 1.1).

SiRNA induced Transcriptional Gene Silencing After transfection of promoter specific siRNAs, several studies observed reduced mRNA or protein expression levels of the respective target proteins (EF1a, CDH1, PR, AR, MYP, COX2)

[Morris *et al.*, 2004; Ting *et al.*, 2005; Janowski *et al.*, 2005]. Inspired by the work in other organisms (see Sections 1.5.3, 1.5.4, 1.5.5, 1.5.6), these studies also indicated DNA-methylation [Morris *et al.*, 2004] or silencing histone modifications [Janowski *et al.*, 2005] to be involved in the human TGS mechanism. Using stable shRNA expressing HeLa cells or siRNAs, to target the promoter region of CCR5 and the tumor suppressor RASSF1A, Kim *et al.* [2006] suggested a role for a network consisting of Ago1, TRBP2 and RNA PolII in TGS in human cells. Also, the polycomb repressive complex PRC2, which comprises methyltransferase activity and is involved in heterochromatin formation [Chen and Dent, 2014], (compare Section 1.5.1), was implicated to be recruited via Ago1. Then, PCR2 should induce histone H3 modifications (H3K27me3, H3K9me2) at these loci and mediate heterochromatin formation. In addition, they also found enrichment of Ago1 at the promoters of MYT1 and CCR5 in untreated HeLa cells [Kim *et al.*, 2006].

In the breast cancer cell line T47D the promoters of huntingtin (htt), progesterone and androgen receptor (PR, AR) were analysed for TGS by targeting with specific siRNAs [Janowski *et al.*, 2006]. All gene loci responded with reduced transcription, however using the same approach in MCF7 cells, no response for PR (siRNAs: PR26, PR24), which has a lower expression level in these cells, was observed. A role of histone modifications could not be confirmed, as suggested by [Kim *et al.*, 2006]. Applying siRNA knockdown and ChIP experiments for Ago1 and Ago2, Janowski *et al.* [2006] suggest that both proteins are necessary for transcriptional silencing and associate with the respective promoters. Puzzling is, however, the fact that none of the indicated Ago proteins complements the silencing effect under knockdown of its homologue. Only recently, the study from Janowski *et al.* [2006], was confirmed by Cho *et al.* [2014], who used and improved their TGS-model in T47D cells. The findings on an inhibition of siRNA-mediated TGS upon Ago2 knockdown were extended by a dependence on SETDB1, which was shown to interact with Ago2 in IP experiments, but is recruited to the AR promoter only after Ago2 binding [Cho *et al.*, 2014]. TGS at the androgen promoter coincided with the accumulation of the heterochromatic marks H3K9me3, H3K27me3 and its modifier the polycomb repressor complex 2 (PRC2) with its methyltransferase component EZH2, all in Ago2- and SETDB1-dependent fashion. Only under combined histone modifications a loss of RNA PolII at the AR proximal promoter was observed. In addition, they found Ago2 to associate with an antisense transcript and a chromatin remodelling complex consisting of SIN3A

and HDAC1, which interacts with SETDB1. Upon targeting with siRNAs against the AR promoter, only the mRNA transcription level decreased, while the antisense transcript level remained constant. In their study Cho *et al.* [2014] found no DNA methylation activity associated with the AR promoter under silencing conditions. In agreement with the study of Cho *et al.* [2014], Gary Felsenfeld suggested SETDB1 as methyltransferase involved in Ago2 TGS at the Keystone conference on "Noncoding RNAs in Development and Cancer" in 2013.

SiRNA induced Transcriptional Activation In contrast to TGS observed by earlier studies, Li *et al.* [2006] used siRNAs targeting CDH1, p21 and VEGF promoters leading to various effects from none to strong induction of transcription dependent on the cell line, which was termed transcriptional activation [Li *et al.*, 2006; Janowski *et al.*, 2007]. Further support for this observation was contributed by Janowski *et al.* [2007] for the promoter of PR in MCF7 cells (compare Table 1.1). This is presumably due to the different epigenetic modifications found at the promoter regions of the cells lines used. Several other loci examined, failed, to generate an activating response [Li *et al.*, 2006; Janowski *et al.*, 2006], suggesting that the gene activation is not a general and conserved phenomenon. At the activated CDH1 promoter, a reduction of silencing histone modification (H3K9me2/3) was measured, following targeting [Li *et al.*, 2006]. DNA methylation seemed, however, not to be involved and Ago2 was identified as effector protein [Li *et al.*, 2006].

siRNA/miRNA	Method	Cell line	Target promoter	Detection method	Regulatory effects	Ago protein	Reference	Additional factors
siRNA-EF1a	transfection	HEK-293	EF1a, EF1a-GFP	target mRNA levels, NucRunOn	gene silencing	.	[Morris *et al.*, 2004]	DNA methylation
siRNA-CDH1	transfection	HCT-116 ,MCF7	CDH1	target mRNA and protein levels	gene silencing	.	[Ting *et al.*, 2005]	H3K9me2
siRNA-PR9/RP24/ AR/MVP/COX2	transfection	T47D	progestron and androgen receptor (PR, AR), major vault protein (MVP), cyclooxygenase-2	target mRNA and protein levels	gene silencing	.	[Janowski *et al.*, 2005]	
siRNA/shRNA-CCR5/RASSF1A	transfection	HEK-29, HeLa	CCR5, RASSF1A	target mRNA and protein levels, ChIP	gene silencing	Ago1	[Kim *et al.*, 2006]	EZH2(PCR2), H3K27me3, TRBP2
.	.	HeLa	MYT1, CCR5	ChIP	gene silencing	Ago1	[Kim *et al.*, 2006]	EZH2(PCR2), H3K27me3, TRBP2
siRNA-CDH1	transfection	PC-3, HeLa	CDH1	target mRNA and protein levels	gene activation	Ago2	[Li *et al.*, 2006]	(-) H3K9me2/3
siRNA-p21	transfection	PC-3, MCF7, HeLa	p21	target mRNA and protein levels	gene activation	Ago2	[Li *et al.*, 2006]	
siRNA-p21	transfection	HEK-293, J82, LNCaP, T24	p21	target mRNA and protein levels	mild gene activation	Ago2	[Li *et al.*, 2006]	
siRNA-CDH1/p21/VEGF	transfection	HeLa	VEGF	target mRNA levels	gene activation	Ago2	[Li *et al.*, 2006]	
siRNA-PR24/PR26/AR/Htt	transfection	T47D	Progersteron and Androgen receptor, Huntingtin	target mRNA and protein levels, NucRunOn, ChIP	gene silencing	Ago1/Ago2	[Janowski *et al.*, 2006]	
siRNA-PR24/26	transfection	MCF7	Progersteron receptor	target mRNA and protein levels	no effect	.	[Janowski *et al.*, 2006]	
siRNA-PR11/26	transfection	MCF7	Progesteron receptor	target mRNA and protein levels	gene activation	.	[Janowski *et al.*, 2007]	(+) H3K4me2
siRNA-PR/AR	transfection	MCF7, T47D	Progesteron receptor	target mRNA and protein levels, RIP	gene activation/silencing	Pan-Ago	[Schwartz *et al.*, 2008]	antisense transcript bound by Ago, HP1γ, hnRNPK
siRNA-PR/AR	transfection		Progersteron receptor	target mRNA and protein levels, RIP	gene activation/silencing	Ago2	[Chu *et al.*, 2010]	antisense transcript
siRNA-AR50, PR26	transfection	T47D	Progersteron receptor, Androgen receptor	target mRNA and protein levels, RIP	gene silencing	Ago2	[Cho *et al.*, 2014]	SETDB1, (+) H3K9me3, (+) H3K27me3, PRC2 (EZH2), antisense transcript, SIN3A, HDAC1
miR-373	transfection	PC3	CDH1, CSDC2	target mRNA levels	gene activation	.	[Place *et al.*, 2008]	.
miR-373	transfection	HCT-116	CDH1, CSDC2	target mRNA levels	gene activation (CSDC2)	.	[Place *et al.*, 2008]	.
miR-373	transfection	LNCaP	CDH1, CSDC2	target mRNA levels	no effect	.	[Place *et al.*, 2008]	.
miR-320	transfection	HEK-293	POLR3D	target mRNA levels, ChIP	gene silencing	Ago1	[Kim *et al.*, 2008]	EZH2(PCR2), H3K27me3
miR-320	endogenous	HEK-293	POLR3D	target mRNA levels antagomiR	gene silencing	.	[Kim *et al.*, 2008]	.
miR-423-5p	transfection	T47D	Progesteron receptor, IGSF1	target mRNA levels, RIP	gene silencing	Ago2	[Younger and Corey, 2011]	H3K9me2, antisense transcript
miR-372, miR-373, miR-520c-3p	transfection	T47D	Progesteron receptor	target mRNA levels	gene silencing	.	[Younger and Corey, 2011]	
miR-423-5p	endogenous	MCF7, T47D	Progesteron receptor	target mRNA levels + antagomiR	no effect	.	[Younger and Corey, 2011]	
miR-423-5p	transfection	MCF7	IGSF1	target mRNA levels	gene silencing	.	[Younger and Corey, 2011]	

continued on the next page

siRNA/miRNA	Method	Cell line	Target promoter	Detection method	Regulatory effects	Ago protein	Reference	Additional factors
miR671	inducible	HEK-293	CDR1	target mRNA levels	gene silencing	Ago2 (slicer)	[Hansen *et al.*, 2011]	circular antisense transcript
miR671	endogenous	HEK-293	CDR1	target mRNA levels + antagomiR	gene silencing	Ago2 (slicer)	[Hansen *et al.*, 2011]	circular antisense transcript
miRNAs	endogenous	WI-38 (senescent)	E2F target genes (PCNA, CDCA8, CDC2, CCNA2)	RIP-Seq, Small RNA-IP-Seq, microscopy, target mRNA levels	gene silencing	Ago2	[Benhamed *et al.*, 2012]	(+) H3K27me3, (+) H3K9me2
miR-let7f, miR185	transfection	WI-38 (presenescent)	CDC2, CDCA8 (miR-let-7f), ORC6L (miR-185)	RIP, Small RNA-IP, ChIP, target mRNA levels	gene silencing	Ago2	[Benhamed *et al.*, 2012]	(+) H3K27me3, (+) H3K9me2
miRNAs	endogeonus	PC-3		ChIP-Seq, target mRNA levels (microarray Ago1 kd), microscopy	gene activation	Ago1	[Huang *et al.*, 2013a]	(+) H3Kme4
miRNAs	endogeonus	PC-3	PIK3CA, PRKCH, CDC6, and RRM1	ChIP	gene activation	Ago1	[Huang *et al.*, 2013a]	
miRNAs	endogeonus	PC-3/ RWPE	SMC1A, CDC20, SMAD3, BUB1	ChIP, target mRNA levels (less then 2 fold change)	gene activation	Ago1	[Huang *et al.*, 2013a]	

Table 1.1: *Overview of studies regarding small RNA-mediated transcriptional gene regulation.*

A Unifying Mechanism for Transcriptional Silencing and Activation In several follow up papers by the Corey lab [Janowski *et al.*, 2007; Schwartz *et al.*, 2008; Chu *et al.*, 2010] working with MCF7 and T47D cell lines, the above described findings were refined. Thus, some siRNAs (PR11, PR26) targeting the promoter of PR could be found, which activated transcription from the PR gene in MCF7, while others (PR9) were shown to silence PR in T47D [Schwartz *et al.*, 2008]. Gene activation was linked to an increase in histone H3K4me2/3 and reduction of histone H3K9 and H3K14 acetylation [Janowski *et al.*, 2007]. In contrast to this finding on differential histone modification, all of these three activating histone modifications were implied to be associated with gene activation or bivalent promoters by Karmodiya *et al.* [2012]. The chromatin interaction of the RISC at the PR promoter, was explained by long antisense transcripts, spanning the region in question [Schwartz *et al.*, 2008]. Using Ago ChIP, RIP and biotin labelled siRNA pull-down, Ago association with chromatin at the promoter and direct interaction with the antisense transcript by complementary base pairing of the bound small RNA could be shown [Schwartz *et al.*, 2008]. The accumulation of RNA PolII at the siRNA activated promoter and reduction of RNA PolII levels at silenced loci support their findings. Finally, also HP1γ, was detected in ChIP experiments at the PR promoter. HP1γ, a member of the heterochromatin binding proteins, interacts with histone H3K9me and is implicated in heterochromatin formation [Chen and Dent, 2014] (see Section 1.5.1). Treating the cells with an activating siRNA led to reduction of HP1γ association and thus relieve of silencing, while the interaction was unchanged for the silencing siRNAs, specific for the PR promoter. Analysing Ago1 to 4 in parallel, Chu *et al.* [2010] showed nuclear localisation for all Ago proteins. However, they found only evidence for Ago2 being involved in transcriptional silencing and activating pathways in MCF7 and T47D cells [Janowski *et al.*, 2007; Schwartz *et al.*, 2008; Chu *et al.*, 2010]. Additional proteins involved in the modification of histones and their interaction with Ago2 were not yet described. This lack of mechanistic details renders the observations on transcriptional activation and silencing by promoter targeted siRNAs strongly controversial in the field.

Transcriptional Gene Regulation by miRNAs in Somatic Cells

Using bioinformatic prediction several miRNAs complementary to gene promoter sequences were identified [Place *et al.*, 2008; Kim *et al.*, 2008; Younger and

Corey, 2011] (compare Table 1.1). Gene activation was observed by targeting the CDH1 and CSDC2 promoters in different cell lines with transfected miR-373 [Place *et al.*, 2008]. In contrast, Ago1-dependent gene silencing was observed at the promoter of POLR3D, when HEK-293 cells were transfected with miR-320 [Kim *et al.*, 2008]. Younger and Corey [2011] tried to link gene silencing at the PR gene [Janowski *et al.*, 2006; Schwartz *et al.*, 2008] to endogenous sources by predicting miRNA complementary to the PR promoter. While they found several miRNAs resulting in gene silencing (compare Table 1.1), they in depth studied miR-423-5p mediated TGS [Younger and Corey, 2011]. For TGS to occur, they found the seed sequence (2-8 *nt*) of miR-423-5p as well as interaction of Ago2 with a promoter associated antisense transcript to be required. In addition, the histone silencing mark H3K9me2 was induced by miR-423-5p-dependent TGS. Finally, they analysed endogenous miR-423-5p function in T47D and MCF7 cells by transfection with antagomiRs. These are 2′-O-methyl modified ssRNAs fully complementary to miR-423-5p, which bind to their target miRNA and cause its degradation. However, in both cell lines no increase of PR transcription was observed after treatment with antagomiRs. Therefore, Younger and Corey [2011] conclude that, while transfected miR423-5p can induce TGS, the endogenous miRNA is not mediating this function in T47D or MCF7 cells.

Recently, a report by Huang *et al.* [2013a] described a role of Ago1 but not Ago2 in transcriptional gene activation by applying ChIP-Seq in prostate adenocarcinoma PC-3 cells. Interestingly, this interaction was Dicer and thus presumably small RNA dependent. By correlating the Ago1- with H3K4me3-ChIP-Seq results, they could identify Ago1 bound to TSS of active genes. In order to prove their findings, the TSS of the active genes PIK3CA, PRKCH, CDC6, and RRM1 were tested by ChIP and shown to be enriched for Ago1 and RNA PolII. From the comparison of their ChIP-Seq data with expression array data from cells treated with siRNA against Ago1 they drew the conclusion, that Ago1 promotes transcription of a subset of genes. Given the lack of replicates in the ChIP-Seq experiments the genome wide predictions by this study have to be interpreted rather carefully.

A quite different example of miRNA mediated TGS was presented by Hansen *et al.* [2011], who found miR-671 to silence transcription from the CDR1 gene. During further analysis they showed that miR-671, which is antisense to a sequence proximal to the transcription start site of CDR1, binds to a circular antisense transcript deriving from the CDR1 gene locus. This binding is mediated by

Ago2 and leads to cleavage of the circular antisense transcript coinciding with reduced transcription of CDR1 mRNA [Hansen *et al.*, 2011]. In summary the study shows a stabilising effect of a circular antisense RNA on transcription of the gene locus encoded mRNA. However, further studies are need to gain more mechanistic insight of the proposed pathway.
Recently, Benhamed *et al.* [2012] found Ago proteins from senescent primary lung fibroblast cells (WI-38) to be associated with E2F-target gene promoters (320 bound promoters) in ChIP experiments (pan-Ago) compared to pre-senescent control cells (77 bound promoters), which was supported by nuclear Ago localisation in microscopic experiments. E2F is a transcription factor involved in cell cycle regulation which interacts with the tumor suppressor Retinoblastoma protein (RB1). Additionally, they isolated small RNAs from RNA-IP experiments with a pan-Ago-specific antibody and an H3K9me2-directed antibody. Using ChIP experiments at several E2F target promoters they could establish a coinciding association of Ago2 with the repressive histone marks H3K27me3 and H3K9me2 at these loci. Supported are these findings by a reduced expression from the targeted loci, which is relieved upon Ago2 knockdown. Furthermore, they validated their findings at the promoters of CDC2 and CDCA8 targeted by let-7f and ORC6L targeted by miRNA-186, which were identified by their small RNA-IP-Seq experiment. A silencing and heterochromatin inducing function for let-7f on E2F target promoters was shown in luciferase reporter assays in combination with Ago2 and the repressor RB1 and in ChIP and nuclear run on assays *in vivo*. Still missing in the description of the TGS pathway of Benhamed *et al.* [2012] are factors like SUV39 and a link to RNA PolII. Also reproductions of these findings by independent groups are required.

Endogenous siRNAs with a Potential Function in TGS

Another view to nuclear Argonaute function comes from the finding of endo-siRNAs in human somatic cells like HeLa and HEK293. Yang and Kazazian [2006] set out, to confirm the post-transcriptional silencing of TE in somatic cells by using a L1 reporter construct. This was generated by fusing the 5′-UTR of L1 to GFP and followed by integration in 293 FlipIn-cells. They detected endo-siRNAs by northern blotting, which derive from long dsRNAs (~400 bp) generated by pairing of the L1 transcript with transcripts generated from an antisense promoter (ASP) located in the 5′-UTR of L1. This is also supported by experiments using

siRNAs generated from a dsRNA of a full length LINE element ($L1_{RP}$) by recombinant Dicer, which were co-transfected with reporter constructs consisting of the 5′-UTR of $L1_{RP}$ or L1.3 and firefly luciferase. These $L1_{RP}$-derived siRNAs could target not only the 5′-UTR of $L1_{RP}$, but also of the related LINE element (L1.3) due to 98% sequence homology in this region [Soifer *et al.*, 2005]. Deletion constructs lacking the ASP led to accumulation of transcripts, due to the loss of L1 siRNAs [Yang and Kazazian, 2006].
In addition, both groups applied retrotransposition experiments with reporter constructs carrying a promoter and either GFP or a resistance marker in antisense orientation fused to the LINE element. Both reporters contained the γ-globin intron in the orientation of the LINE element. Therefore, the marker could only be detected after transcription, splicing, retro-transcription and insertion. Here, the destruction of the ASP and thus endo-siRNA production gave rise to retrotransposition events [Yang and Kazazian, 2006]. In case of treatment with siRNAs derived from the 5′-UTR, the retro-transposition could be reduced [Soifer *et al.*, 2005]. In ChIP experiments, they examined the loading of wild type L1 constructs and the ones lacking a functional ASP with RNA PolII [Yang and Kazazian, 2006]. As the RNA PolII association was comparable for both constructs, they concluded post-transcriptional targeting of L1 by RNAi. Given that long dsRNA induces an interferon response in mammalian cells, they supposed that the dsRNA derived from the L1 5′-UTR is diced in the nucleus, followed by post-transcriptional degradation by an RNAi mechanism. The nuclear localisation of Dicer is also supported by other studies [Doyle *et al.*, 2013; Gagnon *et al.*, 2014]. In contrast to exogenous siRNAs targeting promoters (see above, Section 1.5.7), the potentially nuclear endo-siRNAs derived from 5′-UTR of LINE elements lead to post-transcriptional degradation and do not modify the transcription process at the target locus [Yang and Kazazian, 2006].

TGS in Mammalian Germ Cells

Mouse Piwi Proteins The knockout of Piwi proteins leads to opposing phenotypes for male and female mice. Whereas the loss of any of the Piwi proteins leads to male sterility accompanied by stop of meiosis in spermatocytes and reduction of germ cells (Miwi2), females are fertile and normal [Carmell *et al.*, 2007; Deng and Lin, 2002; Kuramochi-Miyagawa *et al.*, 2004]. Mili and Miwi proteins are loaded with piRNAs derived from primary piRNA clusters [Aravin *et al.*, 2007;

Vourekas *et al.*, 2012] (see Section 1.3). Given that Miwi2 shuttles between the cytoplasm and the nucleus, a model is proposed in which the piRNA amplification cycle of Mili and Miwi2, supplies nuclear Miwi2 with antisense sequences needed for targeting of TE genes for DNA-methylation [Aravin *et al.*, 2008], while Mili is involved in PTGS of TE transcripts in the cytoplasm [Iwasaki *et al.*, 2015]. By analysing Mili and Miwi2 knockout animals using high-throughput sequencing, *in situ* hybridisation and bisulfite sequencing an increase in TE (LINE, IAP) expression accompanied by a reduction in DNA methylation was observed [Aravin *et al.*, 2007; Carmell *et al.*, 2007; Kuramochi-Miyagawa *et al.*, 2008]. Finally, Mili (E12.5) and Miwi2 (E15.5) expression coincides with *de novo* methylation of TEs by the DNA methyltransferases Dnmt3a, Dnmt3b and the cofactor Dnmt3L at embryonic day E16.5 to E18.5 [Kuramochi-Miyagawa *et al.*, 2008]. A downstream function of Dnmt3L in the Piwi TGS pathway is further supported by knockout experiments. Absence of Dnmt3L resulted in a complete loss of DNA methylation, while piRNA levels remained unchanged for some classes of TEs and increased for others [Aravin *et al.*, 2008]. These findings imply that mammalian Piwis are involved in genomic silencing of TEs by DNA methylation, reminiscent of the TGS mechanism in plants (Section 1.5.4). However, many factors involved in the Piwi-mediated TGS mechanism are still unknown and need to be characterized.

In addition to silencing of TEs, Mili and Miwi2 were also suggested being involved in *de novo* DNA methylation of the imprinted locus Rasgrf1, hinting to a role of Piwi proteins in epigenetic regulation of genes [Watanabe *et al.*, 2011]. Unlike Mili, the Piwi protein Miwi loaded with pachytene piRNAs seems to have a function in PTGS of mRNAs in the cytoplasm of developing sperm cells [Thomson and Lin, 2009; Reuter *et al.*, 2011; Grivna *et al.*, 2006].

Human Piwi Proteins The function of human Piwi proteins is mostly analysed in cancer and somatic cells, whereas their role in the germline most likely resembles the one described above for their mouse homologues. This is suggested by analysis of Piwi proteins form seminomas, which are germ cell tumours of the testis [Qiao *et al.*, 2002]. There are several publications on human Piwis in cancer cells as reviewed by Thomson and Lin [2009]. Expression of human Piwis in somatic cells and cancer cell lines was tested by qPCR experiments [Keam *et al.*, 2014]. Here, only Hiwi2 showed an expression level of 10% compared to testis

[Qiao *et al.*, 2002; Keam *et al.*, 2014]. The protein expression level of Hiwi2 in breast cancer cells was also similar to the one observed in somatic cells [Keam *et al.*, 2014]. However, the protein expression of Hili and Hiwi was not assessed. In contrast to Miwi2, Hiwi2, which was detected using a Miwi2 antibody recognizing also Hiwi2, showed cytoplasmic localisation. In addition, a strong difference between the piRNAs in Miwi2 libraries from testis, human testis libraries and Hiwi2 libraries from breast cancer cells was detected. If the Hiwi2-bound piRNAs from these cancer cells, which derive predominantly from tRNAs, are functional, however needs further experimental support [Keam *et al.*, 2014]. In addition, also other components of a Piwi pathway would need to be expressed in conjunction with Hiwi2, which have not yet been indicated.

1.5.8 Additional Nuclear Functions of Ago Proteins

Regulation of Splicing

Another mechanism, which was linked to nuclear Ago function is alternative splicing. Using siRNAs against specific intron or exon sequences, Allo *et al.* [2009] showed increased inclusion of an alternative exon into the mature mRNA of fibronectin (FN1). SiRNA knockdown of either Ago1 or Ago2 in combination with the siRNA treatment promoting alternative exon inclusion inhibited the effect. This is astonishing as Ago1 and Ago2 can bind to the intron targeting siRNA and could therefore complement the knockdown of their homologue. The identification of an antisense transcript overlapping the coding sequence, raised the possibility that both siRNA strands could be effectively targeting the locus and supported a possible TGS like mechanism. Along this line also the processivity of RNA PolII was reduced upon treatment with the intron targeting siRNA, which explained the increase of exon inclusion. The alternative splicing effect could also be linked to accumulation of H3K9me2, K3K27me3, HP1α and DNA methylation at the targeted intron [Allo *et al.*, 2009].
These findings are in part supported by phorbol-12-myristate-13-acetate (PMA) induced alternative splicing [Ameyar-Zazoua *et al.*, 2012]. Mass spectrometric analysis of Ago associated proteins revealed among others also factors involved in splicing: snRNP U2 and U5 subunits which are part of the core spliceosome [Alberts *et al.*, 2002]; several SR proteins which are involved in alternative splicing, mRNA export and translation [Long and Caceres, 2009]; PTBP1 and PTBP2

which bind intronic polypyrimidine tracts and are involved in splicing [Kafasla *et al.*, 2012]; Sam68, which was implicated in alternative splicing of CD44, inducing exon V5 inclusion upon phosphorylation [Matter *et al.*, 2002]. In knockdown experiments using siRNAs against Ago1, Ago2 and Dicer a slight reduction of alternative splicing was shown. Additionally, physical interaction of Ago1 and Ago2 with the gene locus of the alternative spliced transcript was demonstrated by ChIP experiments. In addition, Ameyar-Zazoua *et al.* [2012] could show a dependence on HP1γ for Ago1 and Ago2 in induction of alternative splicing. The mechanism behind, is explained by an increase in H3K9me3 methylation marks which reduce the elongation rate of RNA-PolII and thus allow for integration of alternative exons similar to the hypothesis by Allo *et al.* [2009]. Here, the increase in H3K9me3 does lead to a slow down of RNA-PolII while the CD44 transcript level persists. However, these findings stand in contrast to the mechanism described for *C. elegans* where NRDE3-dependent inhibition of RNA PolII is followed by RNA PolII depletion form the genomic locus and silencing [Guang *et al.*, 2010]. In conclusion, many potential factors involved in RNAi dependent splicing have been suggested. However, given the role of Sam68 in CD44 regulation [Matter *et al.*, 2002], which has to be separated from the proposed Ago function [Ameyar-Zazoua *et al.*, 2012] and the current lack of Ago proteins in spliceosome preparations [Meister, 2013], more experimental work has to be performed to conclusively explain the role of Ago proteins in splicing.

Ago Proteins and DNA Double Strand Break Repair

Recently, evidence for involvement of Ago proteins and Dicer in DNA double strand break (DSB) repair was found in *A. thaliana* and humans [Wei *et al.*, 2012]. Working with GFP reporter cell lines inducing only specific DSB breaks, Wei *et al.* [2012] were able to use high-throughput sequencing to characterize potentially DSB-induced small RNAs. Within the libraries, small RNAs surrounding the DSB sites were indeed identified. In addition, DSB repair efficiency was severely reduced after Ago2 or Dicer knockdown [Wei *et al.*, 2012]. These small DSB-induced RNAs (diRNAs) were shown to guide Ago2 to the specifically induced DSB sites on the DNA using ChIP experiments [Gao *et al.*, 2014]. Furthermore, direct interaction of Ago2 with the ssDNA binding protein RAD51, a well known component of the homologous recombination repair mechanism [Yamanaka and Siomi, 2014], was demonstrated [Gao *et al.*, 2014]. Ago2 mutants deficient in small RNA

binding or slicer activity still bound to RAD51 but were unable to recruit RAD51 to DSB sites or restore homologous recombination repair [Gao *et al.*, 2014]. In contrast to this, an earlier study implied Drosha and Dicer involved in DNA double strand break repair [Francia *et al.*, 2012]. Even though the study by Gao *et al.* [2014] nicely explained the link of Ago2 with RAD51 and its importance for homologous recombination, many open question regarding the production of precursors, the processing and loading of diRNAs remain [Yamanaka and Siomi, 2014].

1.6 Objectives of this Work

RNA interference (RNAi) is a potent regulatory mechanism, were a small RNA bound to an Argonaute protein targets a complementary RNA sequence resulting in many diverse processes such as translational repression, mRNA decay, DNA-methylation or heterochromatic gene silencing (Dueck and Meister [2014]). Encouraged by RNAi mediated heterochromatic silencing mechanisms in yeast and plants, reports accumulated were siRNAs were successfully used to trigger transcriptional gene silencing in mammalian cells (see Table 1.1). In addition, several studies implied markers of heterochromatin at the targeted gene promoters, reminiscent of pathways described in *S. pombe*, *A. thaliana* and *C. elegans* (compare Section 1.5.3, 1.5.4, 1.5.6). Furthermore, Ago proteins are found in the nucleus of many different somatic cell types.
Thus, the aim of this work was, to gain insight into a possible endogenous driven Ago-protein-chromatin interaction. First, the nuclear localisation and chromatin interaction of Argonaute proteins was analysed using biochemical assays. Second, a ChIP Sequencing approach was undertaken to identify potential chromatin interaction sites of Argonaute proteins. Third, an antibody-independent approach was established, called DNA Adenine Methyltransferase Identification (DAM-ID) [Steensel and Henikoff, 2000], where the potential binding sites were marked by a fusion protein of Dam and Argonaute at GATC stretches. Forth, bioinformatic analyses were planned to be applied in order to map potential Ago binding sites to the human genome. With these approaches we aimed at a comprehensive overview on chromatin loci potentially regulated by small RNA-guided mechanisms in human somatic cells.

Chapter 2

Material

2.1 Antibodies

Name	Epitope	Technique	Species	Brand	Publication
07-599	hAgo1 (1-46)	ChIP	rabbit pAB	Millipore	
1525	hAgo1 (peptide)	ChIP	rabbit pAB	own	
1C9	hAgo1	ChIP	rat mAB	own	
1C9	hAgo1	WB (1:10)	rat mAB	own	
4B8	hAgo1	ChIP	rat mAB	own	Beitzinger *et al.* [2007]
07-590	hAgo2 (7-48)	ChIP	rabbit pAB	Millipore	
1526	hAgo2 (peptide)	ChIP	rabbit pAB	own	
1526	hAgo2 (peptide)	WB(1:200)	rabbit pAB	own	
2A8	HAgo2 (47-879)	ChIP	mouse mAB		Nelson *et al.* [2007]
5D4	hAgo2	ChIP	rat mAB	own	Beitzinger *et al.* [2007]
11A9	HAgo2 (1-15)	ChIP	rat mAB	own	Rüdel *et al.* [2008]
11A9	HAgo2 (1-15)	WB (1:10)	rat mAB	own	Rüdel *et al.* [2008]
	hCox6A	WB	rabbit pAB	Rehling lab	
ABIN968286	hERP72	WB (1:500)	mouse mAB	antikörper-online.de	
mms-101p	HA-tag 16B12	WB (1:1000)	mouse mAB	Convance	
ab1791	hHistone H3	WB	rabbit pAB	Abcam	
sc-20681	hLamin A/C	WB (1:1000)	rabbit pAB	Santa Cruz	
	hMFN2	WB	rabbit pAB	Rehling lab	
	hMITRAC12	WB	rabbit pAB	Rehling lab	
	hMITRAC15	WB	rabbit pAB	Rehling lab	
	hMRPS15	WB	rabbit pAB	Rehling lab	
	Rabbit IgG	ChIP	rabbit pAB	Santa Cruz	
	Rat IgG	ChIP	rat pAB	Santa Cruz	
ab24758	hRNA Pol II phospho S2	ChIP	mouse mAB	Abcam	
	hSEC61β	WB	rabbit pAB	Rehling lab	

continued on the next page

Name	Epitope	Technique	Species	Brand	Publication
	hTim23	WB	rabbit pAB	Rehling lab	
	hTim44	WB	rabbit pAB	Rehling lab	
	hTom70	WB	rabbit pAB	Rehling lab	
T9026	hTubulin α (DM1A)	WB (1:10000)	mouse mAB	Sigma-Aldrich	
sc-931	hUBF (1-220)	ChIP	rabbit pAB	Santa Cruz	
AB3792	V5-tag	WB	rabbit pAB	Millipore	
15164	His-Probe-HRP	WB	Ni^{2+}-POD	Thermo	
115-035-068	anti-mouse IgG	WB (1:10000)	goat POD	Jachskon Research	
112-035-068	anti-rat IgG	WB (1:5000)	goat POD	Jackson Research	
A2074	anti-rabbit IgG	WB (1:10000)	mouse POD	Sigma-Aldrich	
IRDye 800CW mouse	anti-mouse IgG	WB (1:15000)	goat	LI-COR-Biosciences	
IRDye 800CW rat	anti-rat IgG	WB (1:15000)	goat	LI-COR-Biosciences	
IRDye 800CW rabbit	anti-rabbit IgG	WB (1:20000)	goat	LI-COR-Biosciences	

Table 2.1: *Antibodies used in this work*

2.2 Enzymes

Benzonase	Merck
DNase RQ1	Promerga
DPN I	NEB
DPN II	NEB
T4-DNA-Ligase	NEB
Lysozyme	AppliChem
Phusion ® High-Fidelity DNA Polymerase	New England Biolabs
λ Protein Phosphatase	New England Biolabs
Proteinase K	Thermo Scientific
Restriction endonucleases	Thermo Scientific and New England Biolabs
RNAse A	Thermo Scientific
Taq-DNA-Polymerase	Thermo Scientific
T4-DNA-Ligase	Thermo Scientific
T4 Polynucleotide Kinase (PNK)	Thermo Scientific
Tobacco Etch Virus nuclear inclusion a endopeptidase (TEV)	lab stock

2.3 Chemicals, Enzymes, Peptides and Oligonucleotides

All chemicals were, unless stated otherwise, purchased from one of the following companys: Amersham Bioscience (Buckinghamshire, UK), AppliChem GmbH (Darmstadt, Germany), Biorad (Hercules, USA), GE Healthcare GmbH (Chalfont St Giles, UK), Macherey-Nagel GmbH & Co.KG (Düren, Germany), Merck (Darmstadt, Germany), Qiagen (Hilden, Germany), Roche (Basel, Switzerland), Roth (Karlsruhe, Germany) and Sigma-Aldrich (Munich, Germany). Radiochemicals were provided by Hartmann Analytic GmbH (Braunschweig, Germany). Commonly used enzymes were purchased from New England Biolabs (Ipswich, USA) and Thermo Fisher Scientific Inc. (Waltham, USA). DNA oligonucleotides were synthesized by Metabion (Martinsried, Germany). siRNAs were produced by MPI for Biochemistry (Martinsried, Germany) or Biomers (Ulm, Germany). Plastic consumables and labware were supplied by Sarstedt (Nümbrecht, Germany) and BD Biosciences (San Jose, USA).

2.4 Buffers and Solutions

general buffers		
PBS	NaCl	137 mM
	KCl	2.7 mM
	Na_2HPO_4	10 mM
	KH_2HPO_4	1.8 mM
	adjust with HCl or NaOH to pH 7.5	
TAE	Tris	40 mM
	Glacial Acetic Acid	20 mM
	Na_2-EDTA pH 8.0	2 mM
TBE	Tris	89 mM
	Boric Acid	89 mM
	Na_2-EDTA pH 8.0	2 mM
TBS	NaCl	150 mM
	Tris pH 8,0	10 mM
TE	Tris/HCl pH 7.5	10 mM
	Na_2-EDTA pH 8.0	1 mM
DNA-loading dye	Tris/HCl, pH 7.8	10 mM
	Na_2-EDTA pH 8.0	1 mM
	Bromphenolblau	2.5 mg/ml
	Glycerin	300 mg/ml
Buffer for chemically competent bacteria		
Tfb I	KOAc	30 mM
	$MgCl_2$	50 mM
	KCl	100 mM
	Glycerin	15 %
	H_2O	ad 250 ml
adjust pH 5.8 with Acertic Acid (0.2 M)		
sterile filtrate solution		
Tfb II	MOPS	10 mM
	$CaCl_2$	75 mM
	KCl	10 mM
	Glycerin	15 %
	H_2O	ad 50 ml
adjust pH 7.0 with NaOH (diluted)		
sterile filtrate solution		
SDS-PAGE, Western Blot and Coomassie		
Lämmli-Loading buffer (4×)	Tris/ HCl pH 6.8	240 mM

	SDS	8%
	Glycerin	50 %
	Bromphenolblue	800 $\mu g/ml$
	β-Mercaptoethanol	10 %
SDS running buffer	Tris	25 mM
	Glycine	50 mMÂ
	SDS	1 %
Towbin blotting buffer	Glycine	192 mM
	Tris	25 mM
	adjust with HCl to pH 8.6	
	Methanol	20%
TBS-T	NaCl	150 mM
	Tris pH 8.0	10 mM
	Tween (100 %)	0.1 %
PonceauS staining solution	Ponceau-Red	0.2 %
	Acetic Acid	1 %
Coomassie straining solution	Acetic Acid	10 %
	Ethanol	20 %
	Coomassie R250	0.25 %
Coomassie destaining solution	Acetic Acid	10 %
	Ethanol	30 %
Immunprecipitation		
IP wash buffer	Tris/ HCl pH 7.5	50 mM
	NaCl	300 mM
	$MgCl_2$	5 mM
	NP-40	0.05%
	NaF	1 mM
Lysis buffer	Tris/ HCl pH 7.5	25 mM
	KCl	150 mM
	Na_2-EDTA pH 8.0	2 mM
	NP-40	0.5 %
	NaF	1 mM
	DTT	0.5 mM
	AEBSF	1 mM
Proteinase K buffer	NaCl	300 mM
	Tris/ HCl pH 7.5	200 mM
	Na_2-EDTA (0,5 M)	25 mM
	SDS	2%
Northern Blot		
MOPS-NaOH (50x)	MOPS	1M
	adjust with NaOH to pH 7	
	Keep in the dark at 4 °C	

RNA sample buffer (2x)	Bromphenolblau	0.25 %
	Xylencyanol	0.25 %
	Formamid	
SSC (20x), pH 7	NaCl	3 *M*Â
	Trinatriumcitrat-Dihydrat	0.3 *mM*
	Adjust with HCl to pH 7	
5 x SSC/ 1 % SDS	SSC (20 ×)	5 ×
	SDS	1 %
1 x SSC/ 1 % SDS	SSC (20 ×)	1 ×
	SDS (20 %)	1 %
50 x Denhardts solution	Albumin fraction V	1 %
	Polyvinylpyrrolidon K30	1 %
	Ficoll 400	1 %
	Store at 20 °C	
Hybridisation buffer	SSC (20x)	12.5 *ml*
	Na_2HPO_4 pH 7.2 (1*M*)	0.996 *ml*
	SDS (10%)	34.86 *µl*
	Denhardtâs solution (50 x)	0.996 *ml*
	H_2O	ad 1 litre
Media		
LB-Medium	NaCl	10 *g*
	Yeast-Extract	5 *g*
	Bacto-Trypton	10 *g*
	adjust with NaOH to pH 7.2 - 7.5	
	i.e.(Ampicillin (50 *mg/ml*)	100 *µg/ml*
LB-Platten	LB medium	
	Agar	1,5 %

2.4.1 Oligonucleotides

Target	Name	Sequence 5′-3′	Purpose
Tubulin	Tubulin-Pr-qPCR-F	CCCTCTCCTTTCTCCCTCTC	ChIP-qPCR
Tubulin	Tubulin-Pr-qPCR-R	CAGGCTGGAAGGTTGGAAT	ChIP-qPCR
UBF	UBF-qPCR-F	GTGTGTGGCTGCGATGGT	ChIP-qPCR
UBF	UBF-qPCR-R	CGACTCGGAGCGAAAGATA	ChIP-qPCR
Chr10-248	chr10-248-ChiPqPCR-F	CTCAAGGCAATATTTCTGTTCTCA	ChIP-qPCR
Chr10-248	chr10-248-ChiPqPCR-R	CACTTGGTTGGAAAGGTGC	ChIP-qPCR
Chr1-191	chr1-191-ChiPqPCR-F	GAGTGTGCCACCACATTCAG	ChIP-qPCR
Chr1-191	chr1-191-ChiPqPCR-R	GAGCCAGGAGTTTGAAACCA	ChIP-qPCR
chr1-191-Co-5000	chr1-191-Co-5000-ChiPqPCR-F	GCATTTTGCTACACTGACCACT	ChIP-qPCR
chr1-191-Co-5000	chr1-191-Co-5000-ChiPqPCR-R	CCACAAGTTTGGTGCTGTTC	ChIP-qPCR
Chr12-169	chr12-169-ChiPqPCR-F	TTTTATGTGATGGCACACAGG	ChIP-qPCR
chr12-169-ChiPqPCR-R	chr12-169-ChiPqPCR-R	TTGTGCAGTTCATGCCCTAA	ChIP-qPCR
chr1-2490-ChiPqPCR-F	chr1-2490-ChiPqPCR-F	GATCCTCCCACCACATTCC	ChIP-qPCR
chr1-2490-ChiPqPCR-R	chr1-2490-ChiPqPCR-R	GGGAGACCCCATCTCTACAAA	ChIP-qPCR

continued on the next page

Target	Name	Sequence 5′-3′	Purpose
chr13-1-ChiPqPCR-F	chr13-1-ChiPqPCR-F	AGTGTTCTGGCGAGCAGTTT	ChIP-qPCR
chr13-1-ChiPqPCR-R	chr13-1-ChiPqPCR-R	AACACACAATAGCTAAGACCCAAA	ChIP-qPCR
chr1-436-ChiPqPCR-F	chr1-436-ChiPqPCR-F	GCATTGGTAAGGGCGACTAA	ChIP-qPCR
chr1-436-ChiPqPCR-R	chr1-436-ChiPqPCR-R	CAGAAGAGTTGCGCTCACTG	ChIP-qPCR
chr1-628-ChiPqPCR-F	chr1-628-ChiPqPCR-F	GCCTTGGAAATGGCAAATAG	ChIP-qPCR
chr1-628-ChiPqPCR-R	chr1-628-ChiPqPCR-R	TGGACAGAGTTATGATGAAGAAAGA	ChIP-qPCR
chr16-73-ChiPqPCR-F	chr16-73-ChiPqPCR-F	ACTCCACTTCCATGGGACAC	ChIP-qPCR
chr16-73-ChiPqPCR-R	chr16-73-ChiPqPCR-R	TTTTCCTTGCAGCTCCATCT	ChIP-qPCR
chr17-360-ChiPqPCR-F	chr17-360-ChiPqPCR-F	CTGCACCTGCATTACCAGAA	ChIP-qPCR
chr17-360-ChiPqPCR-R	chr17-360-ChiPqPCR-R	AGCTGTAAGCCACCACAACC	ChIP-qPCR
chr18-157-ChiPqPCR-F	chr18-157-ChiPqPCR-F	TTCCATTCATCGAATTCTTCAC	ChIP-qPCR
chr18-157-ChiPqPCR-R	chr18-157-ChiPqPCR-R	TGAGCAACAAAGTCAACAAACA	ChIP-qPCR
chr18-47-ChiPqPCR-F	chr18-47-ChiPqPCR-F	TCTTCTCTCCCGTAGCTCCTT	ChIP-qPCR
chr18-47-ChiPqPCR-R	chr18-47-ChiPqPCR-R	GGGATTTACAAAACCCAACAAA	ChIP-qPCR
chr19-32-ChiPqPCR-F	chr19-32-ChiPqPCR-F	ATTCTTTCAAAACCCACAACAATTA	ChIP-qPCR
Chr19-32-ChiPqPCR-R	Chr19-32-ChiPqPCR-R	TGGGGATACCACAAATCCTG	ChIP-qPCR
chr19-41-ChiPqPCR-F	chr19-41-ChiPqPCR-F	TGAGACAGGGTCTCGCTTTT	ChIP-qPCR
chr19-41-ChiPqPCR-R	chr19-41-ChiPqPCR-R	AATCAAGGCTGCAGTGAACC	ChIP-qPCR
chr20-189-ChiPqPCR-F	chr20-189-ChiPqPCR-F	CCTGCTCCATGTGTCTCTGA	ChIP-qPCR
chr20-189-ChiPqPCR-R	chr20-189-ChiPqPCR-R	GCAACCTCCTCTTCCCAGTA	ChIP-qPCR
chr20-189-Co-5000-ChiPqPCR-F	chr20-189-Co-5000-ChiPqPCR-F	TGCAGGTGAAAGTTTTGGTTT	ChIP-qPCR
chr20-189-Co-5000-ChiPqPCR-R	chr20-189-Co-5000-ChiPqPCR-R	TGGGCAGAAATCATGTGAAA	ChIP-qPCR
chr2-244-ChiPqPCR-F	chr2-244-ChiPqPCR-F	GGAACTCAGCTCAACATCAGGT	ChIP-qPCR
chr2-244-ChiPqPCR-R	chr2-244-ChiPqPCR-R	CATAGTGCCATAGATGGTTCTCAA	ChIP-qPCR
chr2-65-ChiPqPCR-F	chr2-65-ChiPqPCR-F	CACCCATGTTTTTAAACAAAACTTAAA	ChIP-qPCR
chr2-65-ChiPqPCR-R	chr2-65-ChiPqPCR-R	TCAAGCATTGTTAAGGGAAGAAA	ChIP-qPCR
chr2-82-ChiPqPCR-F	chr2-82-ChiPqPCR-F	CCACCACGTCTGGCTAATTT	ChIP-qPCR
chr2-82-ChiPqPCR-R	chr2-82-ChiPqPCR-R	GCAGATGGATCACTTGAGGTC	ChIP-qPCR
chr2-8-ChiPqPCR-F	chr2-8-ChiPqPCR-F	GTATGGCCCACCATAATTGC	ChIP-qPCR
chr2-8-ChiPqPCR-R	chr2-8-ChiPqPCR-R	TTTTATGGGCTTTGGTGAGG	ChIP-qPCR
chr3-119-ChiPqPCR-F	chr3-119-ChiPqPCR-F	TGCTTTCCCTTGAGCTAACC	ChIP-qPCR
chr3-119-ChiPqPCR-R	chr3-119-ChiPqPCR-R	ATTCTGGGACCACATGGAAA	ChIP-qPCR
chr3-47-ChiPqPCR-F	chr3-47-ChiPqPCR-F	ACCCCACTCCCCAAATCTAC	ChIP-qPCR
chr3-47-ChiPqPCR-R	chr3-47-ChiPqPCR-R	TTCAGGCGCATTATACGTGA	ChIP-qPCR
chr6-557-ChiPqPCR-F	chr6-557-ChiPqPCR-F	CTTCCTGATTCACGGTTGGT	ChIP-qPCR
chr6-557-ChiPqPCR-R	chr6-557-ChiPqPCR-R	TTTAACTTTGTGGGGCCTTG	ChIP-qPCR
chr8-271-ChIPqPCR-F	chr8-271-ChIPqPCR-F	GCAGAGCGTGGCTTAGTTTC	ChIP-qPCR
chr8-271-ChIPqPCR-R	chr8-271-ChIPqPCR-R	GATCCGGGATTCTGCAGTTA	ChIP-qPCR
chr8-84-ChiPqPCR-F	chr8-84-ChiPqPCR-F	GCCTGGCCACGTAATTGTAA	ChIP-qPCR
chr8-84-ChiPqPCR-R	chr8-84-ChiPqPCR-R	TGGAGGAACATCCAGTGACA	ChIP-qPCR
chrX-33-ChiPqPCR-F	chrX-33-ChiPqPCR-F	CCTCACCAAATAAACACAGGTCT	ChIP-qPCR
chrX-33-ChiPqPCR-R	chrX-33-ChiPqPCR-R	TCTGTTTTGTGGATATTGGGGTA	ChIP-qPCR
Control3-47-2000-ChiPqPCR-F	Control3-47-2000-ChiPqPCR-F	CGATCTGTTGCCTTAGGTTCA	ChIP-qPCR
Control3-47-2000-ChiPqPCR-R	Control3-47-2000-ChiPqPCR-R	CCCCTGTGCCTTCTGAGTAGT	ChIP-qPCR
Control3-47-6000-ChiPqPCR-F	Control3-47-6000-ChiPqPCR-F	CATCTCCCCACAACAGGTCT	ChIP-qPCR
Control3-47-6000-ChiPqPCR-R	Control3-47-6000-ChiPqPCR-R	GAGAAGAGAGGGGGACCATC	ChIP-qPCR
chr1-191-bChiPqPCR-F	chr1-191-bChiPqPCR-F	GGCTCAAGCAATCTGCCTAC	ChIP-qPCR
chr1-191-bChiPqPCR-R	chr1-191-bChiPqPCR-R	ATACGGCATCTGTGCCAAG	ChIP-qPCR
chr8-84-5000ChiPqPCR-F	chr8-84-5000ChiPqPCR-F	CCTTGTCAGAGGATTGCTGAA	ChIP-qPCR
chr8-84-5000ChiPqPCR-R	chr8-84-5000ChiPqPCR-R	CAAGCACACACCCTAGAGCA	ChIP-qPCR
chrX-33-5000ChiPqPCR-F	chrX-33-5000ChiPqPCR-F	TCACAGGCTGCATTAACAGC	ChIP-qPCR
chrX-33-5000ChiPqPCR-R	chrX-33-5000ChiPqPCR-R	TTGTCTGCGCAATCTTTTTG	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	401_F	TTCATCTTGGCAAATGCTGT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	51_F	GCTACTCAGGCACACATAGGC	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	851_F	TCTTCTCTCCCGTAGCTCCTT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	402_F	CACTCAGGAGAACAATTTCACA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	52_F	TCCTTTTCTCACTTAGACTTTCCA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	351_F	AACAAGTTACCCTAAGACCGACTG	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	202_F	TCTATCATGAGGGCTGACTTTTT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	1_F	GAGGTTGCCAACACAAGCTA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	53_F	TAGCTCAGTGGGGGAGAAGA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	101_F	CATCATGCTGTATGGTGGAAA	ChIP-qPCR

continued on the next page

Target	Name	Sequence 5′-3′	Purpose
DAM-ID-Seq-Ago2-Targets	2_F	GACATTAAGAAGGTTCATGTCATTTT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	402-Co_F	CAAAGTCAATGGATTTGCTTTATTT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	53-Co_F	TCAGTCCCCACCTGTGAAAT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	401_R	TGTTTATCTGAAGGAGAGCAACAA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	51_R	TGCAAATGAAGAGCAAGTTTGT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	851_R	GGGATTTACAAAACCCAACAAA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	402_R	AAGCCAAATAATCATTTCTTTTAATC	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	52_R	TTCAAAGTGCCTAGGGATTCA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	351_R	TCAGATACCAAGTGGAATAGGTG	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	202_R	TTTTGCAGTTTCGGTGTTGA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	1_R	TGTGCCCTAGGATTCCTCAC	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	53_R	GCCAGTTGCACAGCATACAT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	101_R	GAGTGGTGAAAGTGGGCATC	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	2_R	TGTTTGATGCTGGACATTGTT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	402-Co_R	TGCCTGGCTAAATCCTTTCT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	53-Co_R	GTGACCTCCTTGCTGGAGAC	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	351-U1-F	GCTTCCCAGAAAACATACGG	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	351-U1-R	tttcctgtctgcaggcttgt	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	351-U2-F	CCATTGGGCTGGTAGGATAA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	351-U2-R	GCACTCAGAACATCCAAGCA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	351-U3-F	TCACAGATTTCAAGCCTCAAGTT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	351-U3-R	CTTACGGGATTGGGTGCTT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	351-U4-F	AAAATATTTGTGGCAGGCAGA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	351-U4-R	GACCCTCAATTGTTGTGTTGG	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	351-IN2-F	catgtgctcaacaggtacgc	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	351-In2-R	ctttctgtgatgatggaaatgtt	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	351-D1-F	AGCTAGCCTGGCCTAGAAGG	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	351-D1-R	CCTGATAACCCCAGACAGGA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	351-D2-F	TTGACTAATCACTGCTCCCAAA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	351-D2-R	GAAAACGTCCCTGGGAAAAT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	351-D3-F	GATAATCAGGGGCAACAGGA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	351-D3-R	TCTTCTCAACCTGGTGTGTATGA	ChIP-qPCR
DAM-ID-Seq-DAM-Targets	DAM-P1-F	CAGCTAAGCCTCCATTCCTG	ChIP-qPCR
DAM-ID-Seq-DAM-Targets	DAM-P1-R	CTCCTTGTCTGGCTGGGATA	ChIP-qPCR
DAM-ID-Seq-DAM-Targets	DAM-P2-F	ACAGCACAGGCAAGGTTTTT	ChIP-qPCR
DAM-ID-Seq-DAM-Targets	DAM-P2-R	AGCAGCCACCTAGATTTCCA	ChIP-qPCR
DAM-ID-Seq-DAM-Targets	DAM-Co1-F	TCCAAACCAGTAAATGAGCAAA	ChIP-qPCR
DAM-ID-Seq-DAM-Targets	DAM-Co1-R	CCTTCAAAGGGAGGTTGTGA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	401-U1-F	ATGCCCAGCCGTATGTTTTA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	401-U1-R	CCACTCAAATGAGGGGTGTT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	401-IN2-F	acagctggctagtgcgtacc	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	401-IN2-R	gcctggagaccaagagacag	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	401-IN3-F	gctttgggacctggtgtaga	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	401-IN3-R	AAGGGAACCATCCAAGGACT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	401-D1-F	CCACCACACCTGGCTAATTT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	401-D1-R	AGGTGGGTGAATCGTCTGAG	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	401-D2-F	ATTTCTGCATGACTGCACCA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	401-D2-R	GGCACCCAATTTTCCTTTTT	ChIP-qPCR
DAM-ID-PCR-Cloning	Ago2-clBaC3-1-F	TGGACACAGCTGGGATTACA	ChIP-qPCR
DAM-ID-PCR-Cloning	Ago2-clBaC3-1-R	GAGTTCGAGACCAGCCTGAC	ChIP-qPCR
DAM-ID-PCR-Cloning	Ago1-clBaC1-1-F	AGCTTGGCGTAATCATGGTC	ChIP-qPCR
DAM-ID-PCR-Cloning	Ago1-clBaC1-1-R	TTTATGCTTCCGGCTCGTAT	ChIP-qPCR
DAM-ID-PCR-Cloning	ChrM-12754F	ACTGTTCATCGGCTGAGAGG	ChIP-qPCR
DAM-ID-PCR-Cloning	ChrM:12852R	ATTGCTTGAATGGCTGCTGT	ChIP-qPCR
DAM-ID-PCR-Cloning	Ago1-Mito2F	GGCTCACATCACCCCATAAA	ChIP-qPCR
DAM-ID-PCR-Cloning	Ago1-Mito2R	AACTCACTGGAATGGGGATG	ChIP-qPCR
DAM-ID-PCR-Cloning	Ago1-Mito3F	AAATCTTACCCCGCCTGTTT	ChIP-qPCR
DAM-ID-PCR-Cloning	Ago1-Mito3R	CTACCTTTGCACGGTTAGGG	ChIP-qPCR
DAM-ID-PCR-Cloning	Ago2-8108-F	AATTCCCGGACGTCTAAACC	ChIP-qPCR
DAM-ID-PCR-Cloning	Ago2-8205-R	TGAAACTGTGGTTTGCTCCA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	ChrM-13247-f	TTCTCCACTTCAAGTCAACTAGGAC	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	ChrM-13247-r	CTTTGAAGAAGGCGTGGGTA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	chrM-13762-f	ATCCCCCTTCCAAACAACA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	ChrM-13762-r	TGGTTAGGTAGTTGAGGTCTAGGG	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	ChrM-14253-f	CCAATAGGATCCTCCCGAAT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	chrM-14253-r	GGGTGGTGGTTGTGGTAAAC	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	chrM-8515-f	CCTGAGAACCAAAATGAACGA	ChIP-qPCR

continued on the next page

Target	Name	Sequence 5′-3′	Purpose
DAM-ID-Seq-Ago2-Targets	chrM-8515-r	CAGTACTGCGGCGGGTAG	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	chrM-8959-f	CGAAACCATCAGCCTACTCA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	chrM-8959-r	AGGTGGCCTGCAGTAATGTT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	chrM-9515-f	CACTCCAGCCTAGCCCCTAC	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	chrM-9515-r	GAGTGGGACTTCTAGGGGATTT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	chrM-7539-f	GAAAAACCATTTCATAACTTTGTCA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	chrM-7539-r	GCTGCATGTGCCATTAAGAT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	chr1:24611176-f	CCTGCACAATTGTTGTCTCG	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	chr124617731-r	CCATGGGCTGGTTAGTCAAT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	chrM-6300-f	gtgatgatgtgaggccatgt	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	chrM-11139-r	gcaaccctacacggaggtaa	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	Ago1-c1-226144392	AGCTGCAGACCCACAACTTT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	Ago1-c1-226144395	TTCTATTTGGACAATCCACAGATTT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	Ago2-chrM-6905	CACACTCCACGGAAGCAATA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	Ago2-chrM-6908	GCCACCTACGGTGAAAAGAA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	Ago2-chrM-1229	GCTAAGACCCAAACTGGGATT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	Ago2-chrM-1232	AGTGTTCTGGCGAGCAGTTT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	Ago2-c2-46623686	TCATAGAGAAAAACTGGGTGTGAG	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	Ago2-c2-46623689	GATTTACAGAAAAATTGCAAAAATCA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	Ago2-chrM-3660	ATGGCCAACCTCCTACTCCT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	Ago2-chrM-3663	GAATTTTTCGTTCGGTAAGCA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	Ago2-chrM-15592	GCCCTCGGCTTACTTCTCTT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	Ago2-chrM-15595	GTCGCCTAGGAGGTCTGGT	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	Ago2-chrM-953	CCCCTCCCCAATAAAGCTAA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	Ago2-chrM-956	TCAGATATGTTAAAGCCACTTTCG	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	Ago2-chrM-1001	CCAGTTGACACAAAATAGACTACGA	ChIP-qPCR
DAM-ID-Seq-Ago2-Targets	Ago2-chrM-1102	GGGCTAAGCATAGTGGGGTA	ChIP-qPCR
Ago2	Ago2-seq2	ACCGTGTCTGCAATGTGAC	Sequencing
Ago3	Ago2-seq3	GGCCATTGCGTGCTTC	Sequencing
Ago4	Ago2-seq4	TGCGCGTGCAGCAG	Sequencing
Ago5	Ago2-seq5	CGAGTTCGACTTCTACCTG	Sequencing
VP5-DamV5-Fw		ACTGGCGCGCCCAATTCAGATCTCGGCCGCA	Cloning pIRES-Neo
VP5-DamV5Ago2-Rev		GATGCGGCCGCTCTCGATCGTAGAATCGAGAC	Cloning pIRES-Neo
VP5-DamV5-Stop-Rev		ACAGAATTCTCACTCGATCGTAGAATCGAGAC	Cloning pIRES-Neo
DHFR	DHFR-F-SalI	gtagtcgac ATGGTTGGTTCGCTAAACTG	Cloning pGEM-Teasy
DHFR	DHFR-R-NdeI	tagcatatgTTAATCATTCTTCTCATATACTTCAAAT	Cloning pGEM-Teasy
Ago2	Ago2_1-F-SacII	att ccgcgg ATGTACTCGGGAGCCGGC	Cloning pGEM-Teasy
Ago3	Ago2_222-R-SpeI	tgtactagtcGATTTCCCTGTTAACTCTCCTC	Cloning pGEM-Teasy
AdRT	AdRT	ctaatacgactcactatagggcagcgtggtcgcggccgagga	DAMID
AdRb	AdRb	tcctcggccg	DAMID
AdPCR	AdPCR	ggtcgcggccgaggatc	DAMID
Ago2-attB1-F	Ago2-attB1-F	GGGGacaagtttgtacaaa aaa gca ggc t tc-ATGTACTCGGGAGCCGGCCCCGCAC	GatewayCloning
Ago2-attB1-R	Ago2-attB1-R	gggaccactttgtacaagaaagctgggtTCAAGCAAA-GTACATGGTGCGCAG	GatewayCloning
Ago1-attB1-F	Ago1-attB1-F	GGGGacaagtttgtacaaa aaa gca ggc t tc A-TGGAAGCGGGACCCTCGGGAG	GatewayCloning
Ago1-attB1-R	Ago1-attB1-R	gggaccactttgtacaagaaagctgggtTCAAGCGAAG-TACATGGTG	GatewayCloning
Ago3-attB1-F	Ago3-attB1-F	GGGGacaagtttgtacaaa aaa gca ggc t tcAT-GGAAATCGGCTCCGCAGGAC	GatewayCloning
Ago3-attB1-R	Ago3-attB1-R	gggaccactttgtacaagaaagctgggt TTAAGCGAA-GTACATTGTGCGTAAG	GatewayCloning
Ago4-attB1-F	Ago4-attB1-F	GGGGacaagtttgtacaaa aaa gca ggc t tc AT-GGAGGCGCTGGGACCCGGACCTC	GatewayCloning
Ago4-attB1-R	Ago4-attB1-R	gggaccactttgtacaagaaagctgggt TCAGGCAAAA-TACATCGTGTGCTG	GatewayCloning
Illumina Adapter	PE PCR Primer 1.0	AATGATACGGCGACCACCGAGATCTACACT-CTTTCCCTACACGACGCTCTTCCGATCT	qPCR
Illumina Adapter	PE PCR Primer 2.0	CAAGCAGAAGACGGCATACGAGATCGGTCT-CGGCATTCCTGCTGAACCGCTCTTCCGATCT	qPCR
Illumina Adapter	PE Read 1 SeqNP	GATCTACACTCTTTCCCTACACGAC	qPCR
Illumina Adapter	PE Read 2 SeqNP	GTCTCGGCATTCCTGCTG	qPCR
RPPH1-F1	RPPH1-F1	CAGCGAAGTGAGTTCAATGG	qPCR
RPPH-R2	RPPH-R2	aatgggcggaggagagtagt	qPCR
GAG-F1	GAG-F1	ggagctagaacgattcgcagtta	qPCR
GAG-R1	GAG-R1	ggttgtagctgtcccagtatttgtc	qPCR
GAPDH RV	GAPDH-R	CGCCCCACTTGATTTTGG	RT-qPCR

continued on the next page

Target	Name	Sequence 5′-3′	Purpose
GAPDH FW	GAPDH-F	AATGGAAATCCCATCACCATCT	RT-qPCR
Ago1	Ago1 qPCR-fwd	tttcctggtttcccactagg	RT-qPCR
Ago2	Ago1 qPCR-rev	aaccaccaccttctccagtc	RT-qPCR
Ago3	Ago2 qPCR-fwd	cgtggtcatgaggcacttgcca	RT-qPCR
Ago4	Ago2 qPCR-rev	gatggaagccaaaccacacttc	RT-qPCR
MultiBac	MCS1	GGTACCGCATGCTATGCATCAGCTGC	Sequencing
Multibac	LP14rev	TCATTTTATGTTTCAGGTTCAGGG	Sequencing
MultiBac/pFL	LP6for	AAATGATAACCATCTCGCAAATAAA	Sequencing
HILI Seq 4	HILI Seq 4 fwd	ATCTCACCCTCACAAAATGGTATT	Sequencing
HILI Seq 1	HILI Seq 1 fwd	GTGGTATCAGCAGAGAAGTGGAC	Sequencing
HILI Seq 2	HILI Seq 2 fwd	CTAGCATCCGAAGGACAGATG	Sequencing
HILI Seq 3	HILI Seq 3 fwd	CAAAAGGATGTACATAAGATTGAAGGAC	Sequencing
HILI Seq 506	HILI Seq 506 rev	TTGTCCACTTCTCTGCTGAT	Sequencing
HIWI Seq1	HIWI Seq1	AAAAACAGGTTCTTCAGGCATT	Sequencing
HIWI Seq2	HIWI Seq2	AACAATAAGACATACAGAGTGGATGA	Sequencing
HIWI Seq3	HIWI Seq3	TGAAGCAGCCAATTCATTGAT	Sequencing
HIWI Seq4	HIWI Seq4	CCAGCTGAAAACACTGGTGA	Sequencing
HIWI Seq (455)	HIWI Seq (455)	TGAAAAAGAAGAGCTGAACG	Sequencing
HIWI2 (488)	HIWI2-seq1	GGTGCCATCCTTTTTCTGTCAC	Sequencing
HIWI2 (1002)	HIWI2-seq2	ATGGCACCGAGATCACCTATG	Sequencing
HIWI2 (1522)	HIWI2-seq3	GAACTGCTTGAGAAGAGTTGCAG	Sequencing
HIWI2 (2088)	HIWI2-seq4	TGCTGGTGTAGGGGATGGTCAG	Sequencing
VP5	VP5-Seq-F	CGAAATTAATACGACTCACTATAG	Sequencing
VP5	VP5-Seq-R	CCCAACAGCTGGCCCTCGCAGA	Sequencing
Ago3	Ago3-seq1	TTCTCCGCTCCAGAÂ Â AGGATA	Sequencing
Ago3	Ago3-seq2	GTAATATTGTGGCAGGGCAAC	Sequencing
Ago3	Ago3-seq3	TATTATCÂ Â GTCATCCTGCÂ Â CGG	Sequencing
Ago3	Ago3-seq4	AGAAÂ Â GCCTGCATCAÂ Â GTTTGGA	Sequencing
Ago4	Ago4-Seq1	GGTCTGGTTTÂ Â GGTTTTCATCÂ Â A	Sequencing
Ago4	Ago4-Seq2	AAGCÂ Â TACAGCAAGAÂ Â TCTGCTCC	Sequencing
Ago4	Ago4-Seq3	TGAAACGTGTÂ Â TGGAGATACCÂ Â C	Sequencing
Ago4	Ago4-Seq4	CAGÂ Â GTAGCTTGGC 2051Â Â CAGAACT	Sequencing
universal R-Gerhard		caggaaacagctatgac	Sequencing
Ago1	Ago1 Seq1	CCTGGATGTGGCCATGAG	Sequencing
Ago1	Ago1 Seq2	CAGACAGACAGGAGGAGATCAG	Sequencing
Ago1	Ago1 Seq3	CTGGGAGCAGATGTTACACAC	Sequencing
pLGW-DAM_SEQ		CAAGTAAATCAACTGCAACTACTG	Sequencing
DAMseq	DAMseq	GCTGACTAGGGAGTGGTATCAG	Sequencing

Table 2.4: *Oligonucleotides used in this work*

2.4.2 Plasmids

Name	Insert	Backbone	Source
pIRESneo-EcoDAM-V5-Ago2	HSP-EcoDam-V5-Ago2	pIRESneo	this work
pIRESneo-EcoDAM-V5	HSP-EcoDam-V5	pIRESneo	this work
pFastBac HT A Ago1	Ago1	pFastBac HT A	lab stock
pFastBac HT A Ago2	Ago2	pFastBac HT A	[Pfaff *et al.*, 2013]
pFastBac HT A HILI	HILI	pFastBac HT A	lab stock
pFastBac HT A HIWI	HIWI	pFastBac HT A	lab stock
pFastBac HT A HIWI2	HIWI2	pFastBac HT A	lab stock
pGEM-T-Easy		pGEM-T-Easy	Promega
pLGW-EcoDam-V5-ccdB	EcoDam-V5-ccdB	pLGW	[Vogel *et al.*, 2007]
pLGW-V5-EcoDam	V5-EcoDam	pLGW	[Vogel *et al.*, 2007]
pLGW-Cbx1-V5-EcoDam	Cbx1-V5-Dam	pLGW	[Vogel *et al.*, 2007]

continued on the next page

Name	Insert	Backbone	Source
pLGW-EcoDam-V5-Ago1	EcoDam-V5-Ago1	pLGW	this work
pLGW-EcoDam-V5-Ago2	EcoDam-V5-Ago2	pLGW	this work
pLGW-EcoDam-V5-Ago2-Y529E	EcoDam-V5-Ago2-Y529E	pLGW	this work
pLGW-EcoDam-V5-Ago3	EcoDam-V5-Ago3	pLGW	this work
pLGW-EcoDam-V5-Ago4	EcoDam-V5-Ago4	pLGW	this work
pIRESneo-FlagHA-Ago1	Flag-HA-Ago1	pIRESneo	[Meister *et al.*, 2004]
pIRESneo-FlagHA-Ago2	Flag-HA-Ago2	pIRESneo	[Meister *et al.*, 2004]
PIRESneo-FlagHA-Ago2-Y529E	Flag-HA-Ago2-Y529E	pIRESneo	[Rüdel *et al.*, 2011]
psPAX2	HIV-GAG, GAG-Pol	psPAX2	gift Nevels, Regensburg/Trono Lab
pMD2G	VSV-G	pMD2G	gift Nevels, Regensburg/Trono Lab
pET-28a-Ago1	Ago1	pET-32	lab stock
pET-28a-Ago2	Ago2	pET-32	lab stock
pGem-Teasy-Ago2-ΔC-DHFR	Ago2-ΔC-DHFR	pGem-Teasy	this work

Table 2.5: *Plasmids used in this work*

Chapter 3

Methods

3.1 Working with Bacteria

The amount of bacterial cells is frequently referred to in OD_{600} where 1 OD_{600} equals about 10^7 cells. *E. coli* and mammalian cell specific shuttle vectors were cloned using the bacteria strain *E. coli* XL1-blue. Bacteria were grown at 37 °C in LB media containing the required selection.
Genotype: *recA1 endA1 gyrA96 thi*−1 *hsdR17 supE44 relA1 lac* [F° *proAB lacI*^{q} *ZδM15* Tn*10* (Tet ^{r})].

3.1.1 Preparation of Chemical Competent Cells

E. coli cells were grown overnight in SOB medium (50 *ml*) with respective antibiotics at 37 °C. On the following morning a culture (200 *ml*) was inoculated at 0.2 OD_{600} and grown to 0.5 OD_{600}. After harvesting the cells at 4500 rpm for 10 minutes at 4 °C, the cell pellet was resuspended in Tfb I (15 *ml* per 50 *ml* of liquid culture) and incubated on ice for 20 minutes. The cells were again collected by centrifugation and dispersed in Tfb II (4*ml* per 200*ml* of bacterial culture). Finally, the cells are incubated on ice for another 10 to 20 minutes, aliquoted (50 μl per transfection) and snap frozen in liquid nitrogen. The chemical competent cells are stored at -80 °C.

3.1.2 Transformation of *E. coli* by Heat Shock

Chemical competent *E. coli* were thawn on ice for 10 minutes, mixed with plasmid DNA and incubated for 30 minutes at 4 °C. The DNA was transferred into

the bacteria by a heat shock (45 seconds) applied in a 42 °C waterbath. Immediately afterwards, the cells were cooled down on ice for 3 minutes. For recovery and expression of the antibiotic resistance gene, bacteria were grown at 37 °C for 1 hour and plated on a Petri dish containing LB mixed with the respective antibiotic.

3.2 Working with Insect Cells

For recombinant protein expression Sf21 insect cells were used. This cell line originated from the IPLBSF cell line, which derives from the pupal ovarian tissue of the fall army worm, *Spodoptera frugiperda* [Vaughn *et al.*, 1977]. The advantage of insect cells lays in their ability of posttranslational modification of proteins, which is absent in /coli/. Furthermore,proteins are fully processed and targeted to their subcellular locations [Invitrogen, 2000]. Sf21 cells were grown in Gibco SF900 medium (Thermo Scientific) either adherent in a 6 well plate or in suspension cultures of $0.3 - 2.0 * 10^6 cells/ml$ at 27 °C shaking at 100 rpm.

3.2.1 Sub-culturing Insect Cells

When the culture reached a density of $2 * 10^6 cells/ml$, it was diluted in growth insect growth medium. The culture flask was changed every other week as the cells attached to the surface [Fitzgerald *et al.*, 2006].

3.2.2 Freezing of Insect Cells

Cells from a 100 *ml* Sf21 suspension culture were collected by centrifugation at 136 $\times g$ for 10 minutes and resuspended in freezing medium (Sf900 Medium supplemented with 10%FBS and 10% DMSO) at $60 * 10^6 cells/ml$ and aliquoted to freezing tubes [Fitzgerald *et al.*, 2006]. The cells were frozen by storing them for 1 hour at -20 °C in an isopropanol filled freezing container, followed by incubation overnight at -80 °C. Eventually the cells were stored in liquid N_2.

3.2.3 Thawing of Insect Cells

Pre-warmed medium (60 *ml*) was transferred into a new culture flask. Frozen cells were collected form the N_2 freezing container and quickly thawn in a 37 °C

waterbath. After dispersion of the nearly thawn cells in pre-warmed medium, the culture was grown for three days to ensure proper division of cells before they were used for virus preparation [Fitzgerald *et al.*, 2006].

3.2.4 Transfection of Insect cells for Virus Production

The Baculovirus protein expression system usually uses the *Autographa californica nuclear polyhedrosis virus* (AcNPV) as vector, which belongs to the family of Baculovirdiae [Berger *et al.*, 2004]. This virus contains a double stranded circular DNA genome and employs two different infectious viral particles. During the late phase of infection (18 to 36 hours after infection), a membrane surrounded viral particle buds from the cell to infect the surrounding tissue. The second form is produced from 24 to 96 hours post infection and consists of virions embedded in a crystalline polyhedra matrix [Invitrogen, 2000], which is released by cell lysis. For expression of recombinant proteins the strong promoters of polyhedra matrix genes are exploited. This leads to high levels of protein production during the very last phase of infection [Okano *et al.*, 2006]. In addition to this also the budded form of the virus (budded virus, BV) is used. After harvesting the BV from the medium, between 18 and 36 hours, this highly infectious virus is employed for efficient infection of Sf21 cells for high level protein production.

3.2.5 Reconstitution of the Baculovirus

Insect cells were seeded in a 6-well plate $0.5 * 10^6$ cells per well and transfected with freshly prepared bacmid (see Section 3.4.3). For the transfection reaction the lipofectant Fugene ® HD (10 μl, Promega) diluted in insect medium (100 μl) was mixed with bacmid DNA (20 μl) in medium (200 μl), incubated at 25 °C for 15 minutes and dispersed equally on two wells of the 6-well plate. The cells were grown for 48 hours to produce viral particles. This could by monitored by fluorescence microscopy due to the YFP integrated into the MultiBac [Fitzgerald *et al.*, 2006].

3.2.6 Harvesting of V_0 Virus Stock

During the late phase of the infection (24 - 48 hours) membrane coated viral particles were harvested by collecting the culture medium. The infection was

controlled by fluorescence microscopy detecting the expression of YFP, which is encoded in the genetically modified viral genome. After another 48 hours of incubation in fresh medium, the cells were harvested (2 wells), lysed in PBS (500 μl) combined with 20 pulses of sonication (duty cycle: 50%, output: 1.5, microtip). To examine the expression of the recombinant protein SDS-PAGE or WB (see Section 3.5.3, 3.5.5) was used.

3.2.7 Creating a High Titre V_1 Virus Stock

An optimal viral infection is achieved, when the targeted cells arrest their cell cycle shortly after exposure to the virus [Fitzgerald *et al.*, 2006]. However, as the virus contains many repetitive sequences, promoters, terminators and recombination sites, special care must be taken to keep the multiplicity of infection (MOI) during virus production at 1. This is one infectious virus per cell. Otherwise, several viral genomes accumulate in one cell and recombination occurs, which results in defective viruses and loss of protein production [Fitzgerald *et al.*, 2006]. Thus, different dilutions of the V_0virus stock produced, were taken to infect suspension cultures of insect cells (50 mlat $1 * 10^6 cells/ml$). In case several of these infection conditions resulted in a growth arrest, the one with the lowest viral concentration was chosen to prevent over infection of the cells. Then the viral particles from this condition were harvested 24 hours after the cell cycle arrest, to gain a highly concentrated virus stock in the supernatant of the used culture. Subsequently, the best concentration for infection by the V_1 virus was determined using above mentioned titration method. Using these cultures, it was also possible to monitor the production rate of the recombinant protein during the next 48 hours by SDS-PAGE or UV-spectrophotometric measurements of YFP.

3.2.8 Expression of Recombinant Proteins

For protein production large cultures of SF21 cells (two times 600 ml) were inoculated and infected with the required amount of V_1 virus stock (see above). The cells were harvested at the time point when most of the recombinant protein was expressed (60 hours after growth arrest), which is during the very late phase of the infection with these modified baculoviruses [Fitzgerald *et al.*, 2006].

3.3 Methods for Culturing Mammalian Cells

Human cells used in this study are HeLa, HEK293, H1299, HeLaS3, MRC-5. All cells were grown in DMEM supplemented with FBS (10 %) and P\S (1 %), which will be referred to as medium, in a CO_2 chamber at 37 °C with evaporation driven humidity level. Mouse cells (mESCs) were cultured in DMEM (Invitrogen), containing 15% of a special selected batch of fetal bovine serum (FBS; Life Technologies) tested for optimal growth of mESCs, 1000 U/ml LIF (Millipore), 0.1 mM 2-β-mercaptoethanol (Life Technologies), 0.05 mg/mL of streptomycin, and 50 U/mL of penicillin (Sigma) on a gelatin coated support in the absence of feeder cells. The mouse cells were grown by C.Constance at ETH Zürich.

3.3.1 Sub-culturing of Adherent Cells

To dissociate the cells from the culture dish (15 cm), these were washed with PBS (18-20 μl) to deplete Ca^{2+} ions from the bonds of the cell to the surface of the plate. Then Trypsin (3-4 ml) was used to digest the cells attachments at 37 °C, followed by addition of medium (6-8 ml) to stop the reaction. Then the cells were diluted according to experiment and growth behaviour of the cell line and spread on new culture plates.

3.3.2 Freezing of Mammalian Cells

Cells were collected by centrifugation at 200 $\times g$ for 10 minutes and gently resuspended in freezing medium (DMEM Medium supplemented with 10% FBS and 20% DMSO) and aliquoted to freezing tubes. The cells were frozen by storing them for 1 hour at -20 °C in an isopropanol filled freezing container, followed by incubation overnight at -80 °C. Eventually the cells were stored in liquid N_2.

3.3.3 Thawing Mammalian Cells

Pre-warmed medium was transferred into a 15 mlFalcon tube. Frozen cells were collected form the N_2 freezing container and quickly thawn in a 37 °C waterbath. After dispersion of the nearly thawn cells in pre-warmed medium, the tube was centrifuged at 200 $\times g$ for 10 minutes. The cell pellet was carefully dissipated and spread on new 10 (15) cmculture plate pre-filled with 10 (25) $ml$of medium.

3.3.4 Collagen Coating of Culture Dishes

The adherence of some cell lines i.e. HEK293 to the tissue culture plate can be increased by coating the plates with collagen. A mixture of collagen (final concentration 50 $\mu g/ml$) in sterile filtrated acetic acid (20 mM) was applied to clean culture dishes (5 ml per 10 cm plate) and incubated on a shaker at RT for two hours. Then unbound collagen was removed by washing step with PBS and two washes with sterile water. Then the plates were dried under the hood for 1 hour, stored at 4 °C or used immediately.

3.3.5 Plasmid Transfection of Cells

To analyse expression level, localisation or function of a tagged exogenous proteins, mammalian expression vector in combination with the easy transfectable HEK293 cells were used. HEK293 cells were seeded at 40 % confluence (10 cm plate) one day ahead of transfection. One hour before treatment the growth medium was renewed. Then the cells were transfected by calcium phosphate co-precipitation [Green and J. [2012]]. The expression plasmid (10 μg) and $CaCl_2$ (125 mM) were combined in water (ad 500 μl). While vortexing this solution, 2× HPS (500 μl 140 mM NaCl, 1.5 mM Na_2,HPO_4, 50 mM Hepes-NaOH pH 7.0) was added drop wise, resulting in formation of tiny DNA containing crystals. Afterwards, the suspension was incubated at RT for 15 minutes and was slowly and carefully applied to the cells. The cells were moved to a growth cabinet overnight and incubated for 24 to 72 hours.

3.3.6 Transfection of Mammalian Cells with siRNA

Transfection of cells with siRNAs leads to a transient knock down of the target protein, which is a vital tool for determining the role and function of a protein. Before siRNA transfection single stranded siRNAs, the sense and antisense strand (40 μM) were mixed, diluted two fold in siRNA annealing buffer () and annealed by 95 °C for 1 minute followed by 25 °C for 1 hour. Annealed siRNA (20 μM) were stored at -80 °C. For transfection the respective cells ($1 * 10^5$) were dispersed in a 6 well plate one day before transfection. SiRNA (5 - 50 nM final concentration) were mixed with Optimem Medium (250 μl), RNAimax (3 μl) was dissolved in Optimem Medium (250 μl) in a separate tube. Then both solutions were mixed

and incubated for 15 minutes at RT. When the nucleic acid filled lipofectant vesicles had formed, the suspension was carefully applied to the cells.

3.3.7 Production of Lentiviruses

To produce lentiviruses HEK293T cells were used as a host for production as suggested by the group of Bas v. Steensel [Vogel *et al.*, 2007]. HEK293T cells were seeded at 40 % confluence (10 *cm* plate) one day ahead of transfection. One hour before treatment the growth medium was renewed. Then the cells were transfected by calcium phosphate co-precipitation [Green and J., 2012]. In order to reconstitute viral particles, the envelop plasmid pMD-2G (3 μg), the packaging plasmid psPAX2 (8 μg), the flanking LTR and insert containing transfer vector pLGW-X (10 μg) and $CaCl_2$ (125 mM) were combined in water (ad 500 μl) [Vogel *et al.*, 2007]. While vortexing this solution, 2$\times$ HPS (500 μl 140 mM NaCl, 1.5 mM Na_2,HPO_4, 50 mM Hepes-NaOH pH 7.0) was added drop wise, resulting in formation of tiny DNA containing crystals. Afterwards, the suspension was incubated at RT for 5 minutes and was slowly and carefully applied to the cells. The cells were moved to a growth cabinet overnight. Following replacement of the medium (6 *ml*), the cells were left to grow for another 24 hours. On the next three days the virus is harvested. The virus containing cell supernatant was filtered through a 0.45 μm filter into a fresh Falcon tube and replaced by fresh medium (6 *ml*). The filtered virus containing solution was stored at 4 °C. After the final virus harvest the filtered supernatants are combined, aliquoted in 1 *ml* fractions and stored at -70 °C.

3.3.8 Infection of Tissue Culture Cells with Lentiviral Particles

For stable and high efficiency integration of a target gene into the cell genome, lentiviral transduction was used. Cells ($4 * 10^5$ per well) were seeded on a 6 well plate and grown overnight. On the following day lentivirus stocks were quickly thawn and mixed 1:2 with pre-warmed medium (0.9 *ml*). After aspiration of the growth medium from the cells, the virus dilution (1.5 *ml*) was added. The cells were incubated for 24 hours in the presence of the virus. Then the medium was replaced and the cells were grown for another 48 hours. After this time the cells were harvested and genomic DNA was isolated.

3.3.9 Creation of Stably Transfected Mouse Cell Lines

The male E14 WT mESC line (129/Ola background) [Hooper *et al.*, 1987] was used to create E14-DAM-ID independent mESCs clones by stable transfection of plasmids pIRESneo-EcoDAM-V5 (HSP-EcoDam-V5) and pIRESneo-EcoDAM-V5-Ago2 (HSP-EcoDam-V5-hAGO2) and selection on G418-containing medium. Screening has been done by QPCR on hAGO2 mRNA level. Two independent clones for each plasmid were selected and amplified before DNA extraction. This work was done by C.Ciaudo (ETH Zürich).

3.4 Genetic and Molecular Techniques

3.4.1 Purification of Plasmid DNA Using Affinity Columns

For column purification of plasmid DNA the Plasmid Kit of Macherey-Nagel was used. For large amounts of plasmid DNA the Midi Kit of Macherey-Nagel was used.

3.4.2 Production of recombined Bacmid

Chemical competent DH10MultiBac-YFP [Berger *et al.*, 2004] bacteria were transfected with a pFastBac HT A vector (Invitrogen), carrying the target gene in the MCS, preceded by a 6× His-tag. During transfection the pFastBac vector recombined with the Baculo virus Bacmid via Tn7 recombination sites. After the heat shock, the bacteria were grown in LB medium without selection marker at 37 °C overnight. On the next morning the transfected cells were spread on LB (10 $\mu g/ml$ Gentamycin, 50 $\mu g/ml$ Kanamycin,10 $\mu g/ml$ Tetracyclin,0.5 mM IPTG, 100 $\mu g/ml$ X-Gal) plates to select for white colonies after 24 hours at 37 °C. The positive clones were re-streaked on another LB (Gentamycin, Kanamycin, Tetracyclin, IPTG, X-Gal) plate to confirm the blue white screening result. Positive clones were used for bacmid isolation and reconstitution of Baculovirus in SF21 insect cells [Fitzgerald *et al.*, 2006].

3.4.3 Purification of Bacmid DNA

An overnight culture (2 ml) of successfully transfected DH10MultiBac cells was grown in the respective selection medium. After collection of the cells, these were

resuspended in buffer P1 (300 μl) containing RNAase A, followed by alkalic lysis using P2 (200 μl) and neutralized after exactly 5 minutes with P3 (200 μl). The white, denatured protein mass was pelleted by 11,000 $\times g$ for 10 minutes at 4 °C. The Bacmid DNA in the supernatant was precipitated using isopropanol (600 μl) at 11,000 $\times g$ for 30 minutes and 4 °C. The pellet was washed in ethanol (1 ml, 75 %), followed by a second wash with ethanol (50 μl, 75%). This volume was left on the DNA till it was transferred to the clean bench for transfection into insect cells.

3.4.4 Purification of DNA form Cultured Cells

The extraction of mammalian DNA from cultured cells was performed using the Blood and Tissue Kit form Qiagen with the following adaptations. The RNase digest was included in step 2 from the isolation procedure. The elution was performed twice with 200 μl buffer AE into a 2 ml tube. Followed by concentration of the DNA by Ethanol Precipitation.

3.4.5 Purification of RNA form Cultured Cells

For the purification of RNA from cultured mammalian cells the Trizol Reagent of Life Technologies was used, according to manufacturers recommendations.

3.4.6 Measurement of Concentration and Purity of Nucleic Acid Solutions

The concentration of DNA/RNA solutions was measured with the Nanodrop UV-spectrophotometer. Therefore the extinction E at 260 and 280 nm was estimated. For double stranded DNA the calculation was based on the correlation $E_{260} = 1$ equals 50 $\mu g/ml$. For RNA the calculation was based on $E_{260} = 1$ equals 40 $\mu g/ml$. The purity of a nucleic acid can be assessed by the quotient E_{260}/E_{280}, which should resemble values 1.8 for clean DNA and 2.0 for clean RNA.

3.4.7 Restriction Digest of DNA

All restriction digests were performed according to the following basic protocol:

basic reaction mix:	DNA	0.5-1 μg
(20 μl)	10x restriction buffer	2 μl
	restriction enzyme	0.5 μl
	H_2O	ad 20 μl

The specific buffer and temperature conditions and, furthermore, the inactivation method were chosen according to the enzyme properties described by the manufacturer (Thermo Scientific/NEB). The incubation time was 1-2 hours. For larger quantities of DNA fragments the digests was scaled accordingly.

3.4.8 Ligation of Plasmids with DNA Fragments

Ligation reactions prepared according to the following basic protocol, were incubated for 1 hour at room temperature:

basic reaction mix:	Vector-DNA	50 ng
(10 μl)	Insert-DNA	3-5 times Vector-DNA
	10x Ligase buffer	1 μl
	T4-DNA-Ligase	1 μl
	H_2O	ad 10 μl

3.4.9 Gel Electrophoresis of DNA

In order to separate DNA due to its size, agarose gel electrophoresis was used. Negatively charged DNA fragments migrate in the electric field at a rate mainly influenced by the density of the polymer network and the fragment size. The gels were made by boiling 0.75 to 4 % agarose in TBE. When the gel had cooled down to about 55 °C ethidiumbromide (0.2 $\mu g/ml$) or ServaStain (1 μl) was added and the gel was poured into a mold (custom-made). The solid gel was subsequently mounted in the electrophoresis-tank (custom-made), where the DNA samples mixed with 10x loading-dye were loaded. The separation was performed in TBE (1x) at a voltage of 80-100 mV.

3.4.10 Recovery of DNA from Agarose Gels

DNA was isolated from agarose gels using the Gel-Extractionkit of Macherey-Nagel, according to manufacturers recommendations.

3.4.11 Purification of DNA Fragments after PCR or Restriction Digests

DNA from enzymatic reactions (PCR, Restriction Digest) was purified employing the PCR-Purificationkit of Macherey-Nagel, according to manufacturers recommendations.

3.4.12 Amplification of DNA by the Polymerase Chain Reaction (PCR)

The polymerase chain reaction invented by Mullis is a method to quickly amplify specific DNA-fragments [Mullis *et al.*, 1986]. For amplification of a specific gene, primers aligning with the borders of the target DNA were chosen and incubated with a thermostable DNA-polymerase. The basic principle of the PCR is a repeating temperature cycle comprising a DNA denaturing step, a primer annealing step and a DNA-extension step, which is the activity period of the DNA-polymerase. For amplification of genes for subsequent cloning and expression a proof reading (3′-5′ excision repair) DNA-polymerase was used (Phusion, NEB).

3.4.13 Amplification of Genes

For preparative amplification of genes from genomic or plasmid DNA, oligonucleotides were used carrying specific restriction sites in addition to the gene aligning parts. These sites were chosen with respect to the multiple cloning site of the vectors used for subsequent work.

basic reaction mix:	5x HF buffer	10 μl
(50 μl)	dNTP-Mix (10 mM)	1 μl
	primer forward (10 $pmol/\mu l$)	2.5 μl
	primer reverse (10 $pmol/\mu l$)	2.5 μl
	Phusion DNA Polymerase (1 $U/\mu l$)	1 μl
	DNA template	10 ng
	H_2O	ad 50 μl

temperature cycle:	denature (DNA)	98 °C	2 *min*
	denature (DNA)	98 °C	20 *sec*
	annealing (primer)	Tm °C	20 *sec*
	extension	72 °C	1 *sec*/*kb*
	final extension	72 °C	5-10 *min*

The melting temperature (Tm) of the primer was computed by Primer3Plus. The extension time was chosen according to the estimated length of the DNA fragment. The amount of template DNA needed for the reaction is dependent on its concentration.

3.5 Biochemical Methods for Protein Analysis

3.5.1 Preparation of Whole Cell Lysates

Exponentially growing cells were washed twice with PBS, if needed scraped from the culture dish, and collected by centrifugation at 200 $\times g$ for 10 minutes at 4 °C. The cell pellet was resuspended in 5 Vol. of ice cold lysis buffer (25 mM Tris-HCl pH 7.5, 300 mM KCl, 2 mM EDTA, 0.3% Nonidet P-40, 1 mM AEBSF) and incubated for 20 to 30 minutes on ice. For break down of genomic DNA, which otherwise will result in a smear on the SDS-PAGE gel, the lysate was sonicated (microtip, 20 pulses, output 1.5, duty cycle 50 %). Cell debris was removed by centrifugation at 17000-20000 $\times g$ for 30 minutes at 4 °C. For analysis with SDS-PAGE a volume equivalent to 20000-40000 cells was mixed with loading dye heated to 95 °C for 5 minutes and applied to one lane of a gel.

3.5.2 Immunoprecipitation (IP)

For IPs, Protein G Sepharose (4FF GE Healthcare) beads were washed three times with PBS (supplemented with 0.1% Triton X-100) and collected by centrifugation at 1000 $\times g$ for two minutes at 4 °C. Hybridoma supernatant or antibody sera were first cleaned from debris or precipitates by high speed centrifugation for 15 minutes at 4 °C, before these were incubated with the beads on a rotating wheel at 4 °C overnight. Before the bead bound antibodies are applied to a homogenate, the beads are washed two times with PBS (supplemented with 0.1% Triton X-100) and two times the respective buffer of the homogenate. Following an incubation

on a rotating wheel for three hours, the beads are washed 5 times in the respective buffer, once in PBS and eluted by 2× SDS-Loading Dye at 65 °C at 850 rpm for 15 minutes.

3.5.3 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

A method to separate proteins according to their size is SDS-polyacrylamide Gel electrophoresis [Laemmli, 1970]. In order to only assess the primary structure, the secondary and tertiary structure of the proteins are dissolved by the detergent sodium dodecyl sulfate (SDS) and the reducing agent β-mercaptoethanol. Additionally, the negatively charged SDS associates with the proteins and leads them to the anode, when an electric field is applied. The density of the polyacrylamide network reduces the migration speed of every protein depending on its size. First the gel matrix of the resolving gel is mixed and poured between two clean glass plates, overlaid with isopropanol and left to polymerize at RT. Then the organic solvent is washed away to add the stacking gel mixture and a comb. For electrophoresis the samples are loaded on the stacking gel where they are compressed by the glycine of the running buffer as soon as the electric field is launched. When the proteins pass into the resolving gel, they have all reached the same starting point and separate according to their size. For standard analysis resolving gels with an acrylamide concentration of 10 % were used (Composition see Green and J. [2012]). The gel electrophoresis was launched with 15-35 mA per gel and was followed by Western blotting.

3.5.4 UREA-Polyacrylamide Gel Electrophoresis (UREA-PAGE)

For separation of ^{35}S labelled proteins UREA-PAGE gels were used. The method works as the SDS-PAGE above only modified by the addition of urea in the running buffer, the resolving and stacking gel. First a 10% resolving gel (6.2 ml Acrylamide:Bis-Acrylamid [60% :0.8%], 7.8 ml Tris-HCl [1M, pH 8.8], 41.5 μlNaCl [4 M], 6.925 g urea, 193.8 μl SDS [10%], 69.3 μl APS [10%], 13.8 μl TEMED) was pored between glass plates and overlaid by isopropanol. Second the stacking gel (0.56 ml Acrylamide:Bis-Acrylamid, 0.85 ml Tris-HCl [1M, pH 6.8], 1.25 g urea, 3.5 ml H_2O, 75 μl SDS, 79 μl APS [10%], 6.9 μlTEMED) was layered on top of the resolving gel.

3.5.5 Western Blotting and Immunological Detection of Proteins (WB)

By means of Western blot analysis proteins separated by SDS-PAGE can be detected specifically. For further analysis the proteins were transferred form the acrylamide gel to a nitrocellulose or PDVF membrane. The protein of interest is then recognised by a specific primary antibody, which is itself bound by a secondary antibody. The secondary antibody is covalently linked to fluorophore (CW700-CW800; Licor) or horse radish peroxidase (HRP). Then the proteins can be detected on a Fluorescence Scanner (Odyssey,Licor) or an X-ray film, by incubating HRP with its substrate, which results in chemiluminescence. For transfer of the proteins from the gel onto the nitrocellulose membrane, a Whatman paper GB005 soaked in transfer buffer was placed on the graphite electrode (anode) of the blotting device. On top of it the water activated nitrocellulose membrane and the acrylamide gel, both equilibrated in transfer buffer, were superimposed. Finally, another soaked Whatmann paper was laid down on the pile, which is followed by another graphite electrode (cathode) of the lid. The transfer was performed with an electric current of 40 mA (160 mA) per gel for 1 hour 10 minutes (45 minutes). The protein carrying nitrocellulose membrane was afterwards stained with Ponceau S for 10 minutes and scanned after 10 minutes of destaining. This served as control to compare the protein amounts loaded in the different lanes. Subsequently, the membrane was blocked with TBS-T milk (5 % milk dissolved in TBS-T) for 45-60 minutes. The blocked membrane was incubated with primary antibody diluted in TBS-T milk (7 ml) at 4 °C overnight on a shaker. Next morning the membrane was washed three times for 5 minutes with TBS-T (10 ml) and then incubated for another hour with a HRP linked secondary antibody diluted in TBS-T milk at room temperature. Again the blot was washed three times for 5 minutes with TBS-T (10 ml). Finally,it was transferred to TBS. To detect the protein by the HRP signal the membrane is placed on a glass plate and incubated with a chemiluminescence blotting substrate by Roche. Thus,reagent A (10 μl) were mixed with reagent B (990 μl) and spread on the membrane in the darkroom. On top a plastic foil was placed. Then an X-ray film was exposed to the chemiluminescence. The positions of the specific protein appear as dark spots on the X-ray film.

3.5.6 Cellular Fractionation

A method to fractionate cells in cytoplasm and nuclei, was originally published by Dignam *et.al* for preparing a nuclear cell extract with high transcription initiation activity [Dignam, 1983]. Taking care while preparing the extracts, a relatively clean fractions of cytoplasm and nuclei can be achieved. Cells were harvested by centrifugation at 1000 $\times g$ for 10 minutes at 4 °C washed in ice-cold PBS and resuspended in five packed cell pellet volumes of buffer A (10 mM Hepes-KOH pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT). After incubation for 10 minutes on ice, the cells are collected by centrifugation as above and lysed in two volumes of buffer A by 10 strokes of a Kontes Dounce homogenizer (B type Pestle). The cell lysis was examined microscopically by tryphan blue staining. The crude nuclei were pelleted at 1000 $\times g$ for 10 minutes at 4 °C and the supernatant was carefully transferred to a new tube and mixed with 0.11 volumes of buffer B (0.3 M Hepes-KOH pH 7.9, 1.4 M KCl, 30 mM $MgCl_2$). To remove residual debris and precipitates, the crude cytoplasmic fraction was centrifuged at 50000 $\times g$ for 60 minutes at 4 °C in a Beckman JA-25.50 rotor. The nuclei containing pellet was washed two times in buffer A and two times in buffer C-NP40 (3 ml per $10^9 cells$, 0.3 % NP-40, 20 mM Hepes-KOH pH 7.9, 10 % glycerol, 150 mM NaCl, 1.5 mM $MgCl_2$, 1 mM AEBSF, 0.5 mM DTT) and was further purified from cytoplasmic material by a high speed centrifugation at 25000 $\times g$ for 20 minutes at 4 °C. The crude nuclei were lysed in buffer C-Triton X-100 (3 ml per $10^9 cells$, 0.5 % Triton X-100, 20 mM Hepes-KOH pH 7.9, 10 % glycerol, 150 mM NaCl, 1.5 mM $MgCl_2$, 1 mM AEBSF, 0.5 mM DTT) for 15 minutes at RT. Then buffer C-NP40 (3 ml, adjusted to 10 mM $MgCl_2$, 10 mM $CaCl_2$) containing DNase (80 μl RQ1 RNase-free, Promega, USA) was added and DNA was digested for 20 minutes at 4 °C on a rotating wheel. After addition of 20 mM EDTA to stop the DNase, the samples were centrifuged in a microfuge at 4 °C for 1 hour at maximum speed. Then a standard IP with the Ago2 (4E1, Hybridoma SN) was performed see below. The IP was washed four times with buffer C-NP40 supplemented with 0.1 % Triton X-100.

3.5.7 Chromatin Association Assay

An assay which was original designed to preserve nuclear matrix structures [Fey E G, 1986], was used by Cernilogar *et al.* [2011] to fractionate cells and probe for

protein chromatin association. HeLa cells ($1.8 - 3.6 * 10^8 cells$) were washed in ice-cold PBS on the plate, scraped form the culture dishes and collected by centrifugation at 200 $\times g$ for 10 minutes at 4 °C. After a further washing step in PBS, the cells were lysed in CSK 1 buffer (7.2 ml 10 mM Pipes pH 6.8, 100 mM NaCl, 1 mM EDTA, 300 mM Sucrose, 1 mM $MgCl_2$, 0.5% Triton X-100, 1 mM NaF, 0.5 mM DTT, 1 mM AEBSF) for 15 minutes on ice. A centrifugation step at 500 $\times g$ for 3 minutes at 4 °C separated the lysate into a SN1 (supernatant not bound to chromatin) and P1 (pellet chromatin bound) fraction. After two washing steps in CSKI, the pellet was resuspended in CSK 2 (7 ml 10 mM Pipes pH 6.8, 50 mM NaCl, 300 mM Sucrose, 6 mM $MgCl_2$, 1 mM NaF, 0.5 mM DTT, 1 mM AEBSF). In the next step chromatin was solubilised by addition of $CaCl_2$ (to a final concentration of 10 mM), $MgCl_2$(to a final concentration of 10 mM) and DNase (140 μl RQ1 RNase-free, Promega) at 20 °C for 30 minutes. Then $(NH_4)_2SO_4$ (250 mM final concentration) was applied for 10 minutes at 25 °C to stop DNase. The nuclease treated samples was centrifuged at 1200 $\times g$ for 6 minutes at 4 °C. The chromatin bound supernatant was designated SN2 and the matrix and hnRNP containing Pellet P2. The P2 fraction was finally solubilised in RIPA buffer (150 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5% Na-Deoxycholate, 0.1% (w/v) SDS, 1% (v/v) NP-40, 1 mM AEPBF, 1 mM NaF) for 30 minutes on ice followed by sonication. A fraction of P1 and of untreated cells were also resuspended in RIPA buffer and sonicated. The samples (20 μg) were analysed by SDS-PAGE and WB.

3.5.8 Mitochondrial Isolation

HEK293 cells (4 culture dishes) were washed from the plate with PBS and collected at 4700 $\times g$ for 15 minutes at 4â¯ °C. The pellet was resuspended in TH buffer + BSA (1 mlper plate; 10 mM KCl, 10 mM Hepes-KOH pH 6.9, 300 mM trehalose, 0.1 % BSA) and 4 mlwere transferred into a 15 mlpotter. To isolate mitochondria, the potter was submerged and carefully moved up and down in the lysate for 20 times under constant rotation of the teflon rod (680 rpm). The suspension was separated into unbroken cells and debris in the pellet and a mitochondria containing supernatant by centrifugation at 400 $\times g$. The pellet was resuspended in TH buffer + BSA and extracted as before. The pooled supernatant form both extractions (400 $\times g$ centrifugation step) was depleted from debris and nuclei by 800 $\times g$ at 4 °C for 7 minutes. The mitochondria were then pelleted by 10000 $\times g$ at 4 °C for 10 minutes. The mitochondrial fraction was then re-

suspended in TH buffer and pooled. Then mitochondria were collected by centrifugation at 10000 $\times g$ at 4 °C for 5 minutes and dispersed in a small volume of TH buffer (i.e. 200 μl). The concentration of mitochondria in the solution is measured by Bradford. For fractionation assays mitochondria can be stored at -80 °C.

3.5.9 Carbonate Extraction of Mitochondria

Frozen mitochondria or freshly isolated mitochondria were resuspended in PBS, aliquoted (6 $\times$; 50 μgper tube) and centrifuged at 10000 $\times g$ for 5 minutes at 4 °C. The supernatant was removed and the were pairwise resuspended in: a) SEM (45 μl,Sucrose, EDTA, MOPS-KOH pH 7.2), KCl (45 μl, 1 M),Triton-X100 (10 μl); b) Na_2CO_3, $NaHCO_3$, pH 10.8 (100 μl, 0.1 M); c) Na_2CO_3, pH 11.5 (100 μl, 0.1 M). The samples were dispersed in the respective buffer and incubated on ice for 20 minutes. One sample of each pair was transferred into polyallomer tubes and applied to 45000 rpm (124517 $\times g$) at 4 °C for 45 minutes. Subsequently, the supernatant was carefully saved in a new tube and the high speed pellet was resuspended in SEM (100 μl). All samples were precipitated with TCA (15 % final concentration) and Na-Deoxycholate (5 μlper 100$\mu l$of sample, 1.25%) by vortexing and incubation for 30 minutes on ice. The samples were centrifuged at 10000 $\times g$ for 30 minutes at 4 °C and the supernatant was discarded. The pellet was washed with ice cold acetone (150μl) and collected by centrifugation at 10000 $\times g$ for 20 minutes at 4 °C with the tubes turned by 180 °. The pellets were dried at 37 °C for 15 minutes, dissolved in loading dye (60 μl) and heated for 20 minutes at 55 °C. For analysis by SDS-PAGE and WB 30 μl were loaded per lane.

3.5.10 Swelling of Mitochondria

The mitochondria swelling assay distinguish the localisation of mitochondrial proteins into outer membrane, intermembrane space (IMS) and matrix. Fresh mitochondria were isolated form HEK293 cells (4 culture dishes) as described above Section 3.5.8. After quantification, two tubes containing 100 μg mitochondria and one tube with 80 μg mitochondria was prepared and centrifuged at 10000 $\times g$ for 5 minutes at 4 °C. The mitochondrial pellet in the first tube (100 μg) was resuspended carefully in SEM (100 μl) and split to three new tubes. The pellet of the second tube (100 μg) was dispersed in EM (50 μl) by pipetting 40 $\times$ up and down

(yellow tip), followed by another addition of EM (50 μl, 20$\times$ up and down), before it was split to three fresh tubes. All samples were incubated on ice for 25 minutes. Then different amounts of Proteinase K (1 mg/ml; 0 μl, 5μl, 10 μl) were added to each of the three tubes. The sample (80 μg), which was serving as sonication control, got resuspended in EM (80 μl, 20 $\times$ up and down) and aliquoted into two tubes. One of those was the no Proteinase K control, the other one was treated with Proteinase K (1 mg/ml; 10 μl). Then both tubes were sonicated three times 20 pulses. Subsequently, all samples were incubated on ice for 10 minutes, to digest all accessible proteins. The reaction was stopped by addition of PMSF (3 μl; 100 mM in ethanol) on ice for 10 minutes. The Mitochondria were pelleted at 10000 $\times g$ for 10 minutes at 4 °C. The pellet was resuspended in LD (50 μl) and boiled at 95 °C for 5 minutes. The samples (25 μlper lane) were analysed using SDS-PAGE and WB (Section 3.5.3, 3.5.5).

3.5.11 Percoll Gradient of Mitochondria

Percoll Gradient for HEK293 cells The Percoll gradient was performed by C. Dennerlein. HEK293 cells (minimum 6 $\times$ 15 cm dishes) were grown and harvested in ice cold PBS + EDTA (1 mM). The cells were pelleted by centrifugation at 400 $\times g$ for 5 minutes at 4 °C. The cells were resuspended in homogenization buffer (HB,220mM mannitol, 70mM sucrose, 10 mM HEPES-KOH (pH 7.4), 1 mM EDTA, 1mM PMSF, 2 mg/ml BSA), incubated for 10 minutes on ice and homogenized with potter S (20 strokes, 800 rpm). Using centrifugation (400 $\times g$ 4 °C 10 minutes), the cell debris was pelleted and the supernatant was saved. The pellet was extracted again in HB and homogenized (20 strokes, 1000 rpm). After centrifugation, see above, both supernatants were pooled. Following another low speed centrifugation step at 400 $\times g$ for 5 minutes at 4 °C the mitochondria were pelleted at 11000 $\times g$ for 10 minutes at 4 °C. The supernatant was removed, the mitochondria containing pellet was washed in 1$\times$ Isolation buffer (IB, 2$\times$ IB: 440 mM Mannitol, 140 mM Sucrose, 20 mM HEPES-KOH (pH 7.4), 2 mM EDTA) and centrifuged at 11000 $\times g$ for 5 minutes at 4 °C. Finally, the mitochondria in the pellet were resuspended in IB buffer (1$\times$, 520 μl), a small sample was taken for WB analysis and the rest (500 μl) was loaded on top of a Percoll gradient. The Percoll gradient was reversely layered (2 ml 12% Percoll, 2 ml 26% Percoll, 500 μl 40% Percoll) in ultra-clear ultracentrifugation tubes for MLS-50 rotor. Then the loaded gradient was centrifuged in the Optima Max Ultracentrifuge (23.000 rpm,

45 minutes, 4 °C Deceleration 9). After the centrifugation all protein layers at the interphases of the gradient were transferred to Eppendorf tubes (2 *ml*) (Top layer "#1", 12% -26% interphase "#2", 26% -40% interphase "#3"). To remove the Percoll from the different fractions, these were diluted 1:5 with IB and centrifuged in a tabletop centrifuge (4700 $\times g$ 15 minutes, 4 °C). After careful removal of the supernatant above the rather loose pellet, residual Percoll was depleted by additional washing steps in 1× IB, till the pellet was tightly bound to the tube after centrifugation. At last the pellet was resuspended in the appropriate amount of 1× IB and protein concentration was estimated by a Bradford assay. For analysis the different fractions were loaded on a WB (10-20 μgper lane).

Percoll solutions (40%: 4 *ml* Percoll solution, 5 *ml* 2× Buffer IB, 1 *ml* MP-H_2O 26%: 3.25 *ml* 40% Percoll solution, 1.75 *ml* 1 × IB Buffer; 12%: 1.65 *ml* 40% Percoll solution, 3.35*ml* 1 × IB Buffer)

Percoll Gradient for MRC5 cells MRC5 cells (23 × 15 cm dishes) were harvested in 4 *ml* ice cold PBS. After spin at 200 $\times g$ for 10 minutes at 4 °C and several washes in PBS, the cells were lysed in five cell volumes of 1 × SEM buffer (20 *mM* MOPS pH 7.2, 250 *mM* Sucrose, 1 *mM* EDTA, 1 *mM* AEBSF) by douncing. The pellet, containing nuclei and unbroken cells, was removed by centrifugation at 700 $\times g$ for 10 minutes at 4 °Cã. The supernatant containing the cytoplasm was transferred to a new tube. Following another extraction of the pellet, the nuclei were pelleted and washed 2 times with 1 × SEM. The mitochondria were separated from the cytoplasm by centrifugation at 13000 $\times g$ for 10 minutes. After a washing step the mitochondria were resuspended in 10% Percoll buffer (2 *ml*) and loaded on top of a discontinuous Percoll gradient (24% [7.8 *ml*] and 40% Percoll [2.5 *ml*], all Percoll solutions were prepared in 1 × SEM). The gradient was centrifuged for 1 hour at 40000 $\times g$ in Ti70.1 rotor at 4 °C in a Beckman ultracentrifugation with no break. After the run, twelve 1 *ml* fractions were taken. The first four fractions were discarded. The residual eight fractions were diluted ten fold in 1 × SEM and centrifuged at 16000 $\times g$ for 15 minutes at 4 °C with deceleration set to slow due to the lose mitochondrial pellet. Following another two washes to completely remove any residual Percoll the samples were resuspended in loading dye (50 μl) and incubated for 15 minutes at 70 °C.

3.5.12 ^{35}S Labelling of Proteins *in vitro*

For ^{35}S-methionine labelling of proteins the TnT® Quick Coupled Transcription/-Translation System for T7 promoter was used. A plasmid encoding the protein of interest under the T7 promoter (500 ng) was mixed with TnT® T7 Mastermix (50 μl) and ^{35}S-methionine (5 μl; 10mCi/ml). After brief vortexing the reticulocyte lysate was incubated at 30 °C for 1.5 hours at 450 rpm. The reaction was stopped by addition of methionine (5 μl; 1 mM) for one minute at 30 °C followed by sucrose (7.5 μl). To examine the result of the coupled transcription and translation reaction a Sample (1 μl) was analysed on an SDS-mini gel. The rest of the labelling reaction was stored at -80 °C for later use.

3.5.13 *In vitro* Import of Proteins into Mitochondria

Fresh mitochondria were isolated form HEK293 cells (4 culture dishes) as described above Section 3.5.8. After quantification mitochondria were aliquoted into five tubes (150 μg) and centrifuged at 10000 $\times g$ for 5 minutes at 4 °C. The pellets were carefully resuspended in import buffer (150 μl; 250 mM sucrose, 5 mM MgOAc, 20 mM Hepes-KOH pH 7.4; add fresh: 2.25 mM ATP [stock 200 mM], 2.25 mM NADH [stock 200 mM], 15 mM Na-succinate [stock 1 M], 15 mM malat [stock 1 M]. One tube was treated with AVO (3 μl; 1 μM valinomycin, 20 μM oligomycin, 8 μM antimycin in ethanol, stored at -20 °C), to destroy the membrane potential of the mitochondria and, thus, inhibit membrane potential dependent protein import. All other tubes were supplemented with ethanol (3 μl). Following an incubation at 37 °C for 3 minutes, the import was started by addition of the reticulocyte lysate (4 μl). The import reaction was stopped after 30 minutes by addition of AVO (3 μl). The samples were split into two tubes and one of each pair was incubated with Proteinase K (2.5 μl, 1 mg/mlin SEM) on ice for 10 minutes. The Proteinase K was inactivated by covalent binding to the active centre of PMSF (2 μl, 100 mM) and left on ice for another 10 minutes. The mitochondria were collected by centrifugation at 10000 $\times g$ for 10 minutes at 4 °C. In order to wash the mitochondria, SEM (80 μl) was carefully added and the tubes were turned by 180 °, which leads to centrifugation (settings as above) through the washing solution. The pellet was resuspended in LD (30 μl), boiled at 95 °C for 5 minutes and analysed by a UREA-Page gel and WB.

3.5.14 ^{35}S Labelling of Mitochondrial Proteins *in vivo*

Given that the mitochondria contain 70 S ribosomes, it is possible to stop the cytoplasmic protein production by selectively inhibiting the eukaryotic 80 S ribosome by emetine. Thus, only mitochondrial proteins are labelled with ^{35}S methionine, when this is supplemented to the growth medium. This is a straight forward method to analyse the impact of target proteins on mitochondrial protein synthesis [Mick *et al.*, 2012]. Cells were grown in 25 cm^2flasks for the experiment. Knockdown of respective proteins was performed by transfection with specific siRNAs (33 nM). For transfection in a 25 cm^2 flask siRNA (8.96 μl 20 μM) were mixed with Lipofectamine RNAimax (4.48 μl, Thermo Scientific) and Optimem medium (1120 μl). After incubation for 20 minutes, HEK293 cells were added ($5.6 * 10^5$). Finally, medium (up to 5600 μl) was added and the cells were split to two 25 cm^2 flasks. After siRNA knockdown for 72 hours medium was replaced by FBS/methionine free media (2 ml) and the cells were incubated for 10 minutes at 37 °C. This step was repeated another time with media containing FBS but no methionine (1.5 ml), before the addition of medium with FBS, Emetine (100 $\mu g/ml$final concentration) and no methionine. After incubation for 10 minutes at 37 °C, 0.5 mlmedium were removed and the mitochondrial translation was pulsed with ^{35}S methionine (20 μl, $10 mCi/ml$) at 37 °C for one hour. The labelling was stopped by removal of the ^{35}S containing medium. The cells were harvested in PBS and collected by centrifugation at 2000 $\times g$ for 2 minutes at 4 °C. Following a washing step in ice cold PBS, the cells were resuspended in cold PBS (200 μl) and protein concentration was measured using Bradford. Finally, the cells (50 μg) were lysed in LD (25 μl; containing 0.25 μlBenzonase (25 kU)) for 20 minutes at 37 °C and analysed by an UREA-PAGE gel (see Section 3.5.4), autoradiography (PhosphorImager) and WB (Section 3.5.5).

3.5.15 Purification of Argonaute Proteins

SF21 cells were grown (600 ml, $1 * 10^6$ cells/ml) and infected with Baculo virus. The cells were harvested 48 hours after infection at 136 $\times g$ for 10 minutes. The cell pellet was lysed in 30 ml lysis buffer (50 mM Hepes-KOH pH 7.5, 500 mM NaCl, 0.1 % NP-40) supplemented with Benzonase (9 μl, 25 kU) and pipetted up and down for at least 25 times. The cells were broken by vortexing at maximum speed for 40 seconds, followed by cooling on ice for 20 seconds, this was

repeated another time. Then the cells were incubated on ice for 5 minutes and vortexed again as before. The lysate was centrifuged at 4000 rpm for 5 minutes at 4 °C. The Argonaute containing supernatant was saved and the pellet was extracted again with lysis buffer (20 ml) as described above. The supernatants from both extractions were mixed and applied to high speed centrifugation 50000 $\times g$ for 1 hour. The cleared lysate was filtered through a 0.45 μm low protein binding filter, supplemented with glycerol (10% final concentration) and imidazol (10 mM final concentration; pH 7.5) and loaded on a Ni-NTA column (5 ml) coupled to a AKTA-purifier system. After loading the sample, the bound protein was washed 40 column volumes of 0.5 M salt wash buffer (50 mM Hepes-KOH pH 7.5, 500 mM NaCl,10 % glycerin, 20 mM imidazol pH 7.5). This was followed by a 2 M salt wash (50 mM Hepes-KOH pH 7.5, 2 M NaCl,10 % glycerin, 20 mM imidazol pH 7.5) for 5 column volumes, to remove unspecifically attached nuclei acids. The column was re-equilibrated to normal salt conditions by washing with 10 column volume 0.5M salt wash buffer. Finally, the protein was eluted with 5 column volumes of 50% elution buffer (50 mM Hepes-KOH pH 7.5, 300 mM NaCl,10 % glycerin, 500 mM imidazol pH 7.5). If the HIS-tag had to be removed, the eluate was dialysed overnight against 0.5 M wash buffer and the HIS-tag was cleaved of at an internal TEV recognition site using TEV protease. The tag and the His-tagged TEV protease were removed by another run on a Ni-NTA column (5 ml). The flowthrough containing the purified protein was saved. Either the first eluate or the flowthrough were concentrated in a Vivaspin 20 (50 kDa MCWO) membrane based centrifugal concentrator at 3000 $\times g$ and 4 °C. The concentrated protein was further purified using gelfiltration on a HiLoad 16/600 Superdex 200 in gelfiltration buffer (100 mM Hepes-KOH pH 8.3, 10% glycerin, 500 mMNaCl, 5 mM $MgCl_2$). During the run, the 260 nm reading, detecting nucleic acids, was higher than the absorption at 280 nm which is specific for the amino acid tryptophan.

3.5.16 Protein Coupling to CNBr Beads

For purification of an Ago2 antibody, the purified protein was coupled to a CNBr column. To prepare the column, CNBr FF4 Beads (1 g) were soaked in HCl (1 mM) for 20 minutes, applied to a disposable plastic column for gravity flow. The column was washed another 10 times with HCl (5 ml, 1 mM) for 2 minutes rotating. The beads were collected by 100 $\times g$ for 20 seconds at 4 °C. The column was

prepared for coupling by two washes with coupling buffer (100 mM Hepes-KOH pH 8.3, 10% glycerin, 500 mMNaCl, 5 mM $MgCl_2$). Then Ago2 (1 mg), diluted in coupling buffer (9 ml), was applied to the beads and covalently bound to the beads during incubation overnight on a rotating wheel. On the next day the beads were blocked with ethanolamin (5 ml, 1 M, pH 8.0), rotating for 2 hours at 4 °C. Then the column was washed (5 ml) - three times each - with acidic buffer (0.1 M NaOAc pH 3.5, 0.5 M NaCl) and basic buffer (0.1 M Hepes pH 8.5, 0.5 M NaCl). The alternating washes were repeated for another two times before the column was left in basic buffer for storage overnight at 4 °C.

3.5.17 Antibody Purification against Target Protein

The Ago2 antibody serum was centrifuged at 20000 $\times g$ for one hour at 4 °C. Subsequently the supernatant was filtered through silanized glasswhool to remove lipids. The Ago2-CN-Br column was equilibrated by 10 column volumes of binding buffer (50 mM Hepes pH 8.0, 500 mM NaCl). Then the cleared antibody serum was applied to the column, which was filled up to 10 ml with binding buffer, and left for binding on rotating wheel at 4 °Cfor 2 hours. Afterwards the column was washed with 10 $\times$ 10 column volumes of binding buffer. The Antibody was eluted by three column volumes of elution buffer (96% citrate [0.2 M], 4% Na_2PO_4 [0.2 M] pH 2.4) into 2 mlEppendorf tubes containing Tris-HCl pH 9.0 (125 μl). Following elution the pH of the antibody solution was titrated by Tris-HCl pH 9.0 to roughly pH 7.5. The antibody was concentrated in a vivaspin 6 (50 $MWCO$) concentrator and stored in glycerol (50 % final concentration) at -80 °C.

3.6 Biochemical Methods for Nucleic Acid Analysis

3.6.1 Size Separation of RNAs via UREA-TBE Gels

Standard UREA-TBE gels and Sequencing Gels of different concentrations were poured between clean glass plates using the SequaGel (UreaGel System), according to manufacturers recommendations. Before use, the UREA gels were heated by applying 400 V ($\sim$ 51 mA) for 20 to 30 minutes. After boiling the RNA samples (20 μg) in RNA LD for 5 minutes at 95 °C, the samples were immediately loaded into the carefully flushed wells of the UREA gel. A radioactively labelled standard was used. The gel was run at 320 to 400 V for one hour or until the dye front

had a distance of 6 to 7 *cm*from the bottom of the well. To control for proper separation the gel was stained with EtBr for 15 minutes in TBE and analysed using a UV screen.

3.6.2 Northern Blot

Northern blots for detection of small RNAs were essentially performed as described by Pall and Hamilton [2008] with minor modifications. The RNA from the stained UREA gel (Section 3.6.1) was blotted onto a nylon membrane (N-Highbond) with three Whatman paper each on top and at the bottom of the sandwich in a semi-dry blotting chamber using Milliq-$H_2O$for blotting (20 V, 30 minutes). After blotting the transferred RNA was crosslinked with EDC (61.25 μl Methylimidazol, 2.25 ml H_2O, 75 μl HCl [1 M], 188,3 mg EDC, at 6 $ml H_2O$; pH 8.0) by soaking a Whatman paper in EDC and placing the membrane on top (RNA faces upwards). The Whatman and membrane are then wrapped into a cling film and incubated for 1 hour at 50 °C in hybridisation oven. For specific analysis of an RNA, a complementary DNA ($\sim$ 20 bp) probe is radioactively labelled. For this the DNA probe is incubated with PNK and ^{32}P-γ-ATP:

PNK labelling reaction:	PNK 10× buffer A	2 μl
(20 μl)	DNA oligo (20 μM)	1 μl
	PNK (10 $U/\mu l$)	1 μl
	^{32}P-γ-ATP (10 mCi/ml)	2 μl
	H_2O	14 μl

Following incubation for one hour at 37 °C, the reaction is stopped by EDTA (30 μl, 30mM) the labelled DNA is purified using gel filtration on MicroSpin G25 columns. Before the membrane was exposed to the DNA probe, it was blocked for one hour at 50 °Crotating in hybridisation solution (12.5 ml 20× SSC, 0.996 ml Na_2HPO_4 1 M 34.86 μl10% SDS, 0.996 ml Denhardts [in case of unspecific background; 0.498 mlSalmon Sperm DNA 10 mg/ml], ad 50 ml H_2O). Then the DNA probe ($\sim$ 60 μl) was added to the hybridisation solution and incubated at 50 °C overnight, rotating in a hybridisation oven. On the next morning the probe containing hybridisation solution is saved and stored at -20 °C for further use. The membrane is washed for 10 minutes in wash solution I (10 ml;5× SSC, 1% SDS) at 50 °C rotating. This is wash is repeated, before the membrane is incubated with wash solution II (10 ml;1× SSC, 1% SDS) as above. To expose the membrane to

a blanked phosphor (BaFBr:Eu) screen, the membrane is wrapped in a cling film and placed in film cassette. After exposure for several hours or overnight, the stored radioactive signal (oxidized phosphor) is released form the phosphor by stimulation with visible laser light (635 nm) see Biorad [2005]. The photons emitted by the reduction reaction of the phosphor are captured by a photomultiplier and processed into a digital file.

3.6.3 Proteinase K digest of co-precipitated RNAs

For the analysis of co-precipitated RNAs, samples were treated with Proteinase K (Roth) in 250 μl of Proteinase K buffer (0.3 mg/ml Proteinase K, 300 mM NaCl 25 mM EDTA, 2% SDS, 200 mM Tris-HCl, pH 7.5) and incubated for 15 minutes at 65 °C. RNA was extracted with one Vol acidic phenol (pH 4.0) by vortexing for 10 seconds. After incubation of ice for 10 minutes and centrifugation at 12000 $\times g$ for 20 minutes, the aqueous supernatant was re-extracted with acidic phenol. The resulting supernatant was precipitated by addition of NaCl (300 mM final concentration), glycogen (1 μl, 20 mg/ml) and 2.5 vol of pure ethanol, followed by incubation overnight at -20 °C. On the next morning the samples were centrifuged at 20000 $\times g$ for 1 hour at 4 °Cand washed two times with ethanol (1 ml,80%, -20 °C). After complete removal of the ethanol the RNA was dried for 3 minutes at RT and resuspended in H_2O(20 μl).

3.6.4 Chromatin Immunoprecipitation (ChIP)

ChIP is an antibody based technique to select chromatin regions bound by the protein of interest. Subsequently, DNA is purified form the enriched chromatin and quantified or analysed regarding their sequence composition. The ChIP experiments were essentially preformed as described by Luo [1998]; Braunstein [1993]. To crosslink proteins with DNA, two culture dishes of cells ($1.5 * 10^7$ cells per 15 cm plate) were treated with formaldehyde (1 %) and incubated for 10 minutes at RT. The reaction was stopped by addition of glycine (125 mM) for another 5 minutes at RT. Subsequently, the cells were washed three times with ice cold PBS and collected by centrifugation at 200 $\times g$ for 10 minutes at 4 °C. The cells were lysed in ChIP-Lysis buffer (1 ml, per cell pellet from two plates; 1% SDS, 10 mM EDTA and 50 mM Tris, pH 8.1, 1 mM AEBSF) for 10 minutes on ice. Sonication was performed in the Diagenode Bioruptor UCD-200 (settings: High, 30

seconds ON, 30 seconds OFF for 3 times 5 minutes) in 15 ml tubes (Falcon) with an ultrasound reflective metallic bar inside the tube and an approximately 0.5 cm thick layer of ice in the waterbath. Every 5 minutes the ice was replaced to ensure proper cooling of the samples during sonication. Cellular debris was removed by centrifugation at 20000 $\times g$, for 15 minutes at 4 °C. The cleared lysate was diluted 10 fold in (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.1, 167 mM NaCl) and 5-10 μg of antibody was added per 2 ml at 4 °Covernight on a rotating wheel. A sample (1/10 Vol of IP) was taken as input control. On the next morning Protein A Agarose/Salmon Sperm DNA (Millipore 17-295, 50 μl) were added per 2 ml tube and left at 4 °C for another hour. Subsequently, the beads were washed for 5 minutes with 1 $ml$of the buffers: low salt (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HC1 at pH 8.1, 150 mM NaCl), high salt (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HC1 at pH 8.1, 500 mM NaCl),LiCl buffer (0.25 M LiC1, 1% Nonidet P-40, 1% Na-Deoxycholate, 1 mM EDTA, 10 mM Tris-HC1 pH 8.1) and two times TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA). The bound DNA was eluted by incubating the beads two times for 15 minutes in elution buffer (0.1 mMNaHCO $_3$, 1% SDS, 200 μl) at 65 °Cin a shaker 850 rpm. The formaldehyde crosslinks were reversed by addition of NaCl (ad 200 mM) and heating to 65 °Covernight. The eluted DNA was treated with Proteinase K (1 $\mu l$of 20 mg/ml) after supplementing the solution with EDTA (10 μl of 0.5 M) and Tris-HCl(20 $\mu l$of 1 M, pH 6.5) for one hour at 45 °C. After PCI extraction, the DNA was recovered by ethanol precipitation (1/10 Vol. Na-Acetate pH 5.2, 1 μl Glycogen, 2 Vol. Ethanol 100 %) at -20 °Covernight. The precipitated DNA was collected by centrifugation at 20000 $\times g$ for 45 minutes at 4 °C, washed with Ethanol (75%), centrifuged for another 10 minutes and air-dried for 10 minutes at RT. The purified DNA was resuspended in TE or water (for qPCR 200 μl for input and IP samples) by incubation for 15 minutes at 37 °C and 850 rpm.

Enrichment of Nuclei was performed for the ChIP-Sequencing experiment to deplete the cells of cytoplasmic Argonaute [Janowski *et al.*, 2006]. After the formaldehyde crosslink, the samples were washed two times with ice-cold PBS. Two culture dishes (15 cm) were harvested, pooled in one falcon tube and collected by centrifugation at 500 $\times g$ for 10 minutes at 4 °C. To remove the cytoplasm, the crosslinked cell pellet was washed in hypotonic buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 2 mM $MgCl_2$, 0.5% Nonidet P-40, 1mM AEBSF) by pipetting ten times slowly up and down. The crude nuclei were centrifuged at 200 $\times g$ for

10 minutes at 4 °C. After a second washing step the pellet was frozen in N_2 or immediately used for ChIP assays.
For the ChIP-Seq experiment the eluted DNA was treated with Proteinase K, extracted with PCI and chloroform. Subsequently, contaminating RNA was removed by RNase A digestion for 1 hour at 37 °C followed by extraction with PCI and two times with chloroform. This was needed to remove trace amounts of phenol, which could negatively affect library generation. The library preparation and NGS was performed by Fasteris (Plan-les-Ouates, Switzerland). Libraries were constructed in paired-end mode including barcodes sequences for multiplexing. Sequencing was performed in two lanes of one flow cell. Four samples GPK1-4 and GPK5-8 were pooled and measured on a HiSeq2000 for 100 sequencing cycles plus another seven cycles for identification of the index. The pictures were taken in 32 tiles, which were then processed by the software packages HiSeq Control 1.1.37.8, RTA 1.7.48 and Casava 1.7.

3.6.5 DNA-Adenine-Methyltransferase Identification (DAMID)

The DAMID technique, an antibody independent approach to map chromatin interactions of proteins, was developed by Bas van Steensel and used as a complementary method to ChIP [Steensel and Henikoff, 2000]. The protein of interest is fused to DNA Adenine Methyltransferase from *E. coli* and expressed in the respective cell line. Here the protocol of Vogel [Vogel *et al.*, 2007] was used. For transduction of the fusion protein encoding genomic sequence a lentiviral system was used (Section 3.3.8). After infection and incubation for 72 hours, DNA was isolated form the cells using the DNeasy Blood and Tissue Kit (Qiagen) with following modifications. An RNase A digest was included in the first solubilisation step and the elution was performed two times with buffer AE (200 μl) in one 2 ml Eppendorf tube. The genomic DNA was subsequently precipitated with NaOAc (0.1 Vol, 3 M pH 5.2) and Ethanol (2 Vol, 100%) and stored at -20 °C for 30 minutes. After centrifugation at 20000 $\times g$ for 30 minutes, the supernatant was removed and the pellet was washed with Ethanol (70%) and air-dried for 10 minutes. The genomic DNA was resuspended in TE (10 mM Tris-HCl pH 7.5, 0.1 mM EDTA) at 1 $\mu g/\mu l$ and incubated at 55 °C for 30 minutes to completely dissolve the DNA. To select for methylated GATC sites, the purified DNA (2.5 μg) was digested with Dpn I (10 U, NEB) in Buffer 4 (NEB) in a total volume of 10 μlat 37 °C overnight. The restriction enzyme was heat inactivated by 80 °C for 20 minutes. The GATC site was restored by ligation of the dsAdR adaptor

to the digested DNA. The dsAdR is an equimolar mixture of the oligonucleotides AdRt (5′-CTAATACGACTCACTATAGGGCAGCGTGGTCGCGGCCGAGGA-3′) and AdRb (5′-TCCTCGGCCG-3′)

basic ligation mix:	DpnI digested DNA	10 μl
(20 μl)	10 $\times$ ligation buffer	2 μl
	Apdaptor dsAdR (50 μM)	0.8 μl
	T $_4$ Ligase (5 $U/\mu l$)	1 μl
	H_2O	ad 20 μl

The ligation reaction was performed at 16 °C for four hours and finally inactivated by 65 °C for 10 minutes. During the following DpnII digest, only unmethylated GATCs cut by the restriction enzyme, leading to selection of double methylated, adaptor carrying fragments.

basic DpnII digest mix:	Ligated DNA	20 μl
(50 μl)	10 $\times$ DpnII buffer (NEB)	5 μl
	DpnII (NEB 10 $U/\mu l$)	1 μl
	H_2O	ad 50 μl

The DNA was incubated with DpnII at 37 °C for one hour. Subsequently, the digested DNA was used as input for the methyl PCR, which specifically amplifies nucleic acid fragments methylated on both sides.

basic PCR digest mix:	DpnII digested DNA	10 μl
(50 μl)	5 $\times$ HF buffer (NEB)	10 μl
	Primer Adr PCR (100 μM)	0.625 μl
	dNTP mix (10 mM)	1 μl
	DpnII (NEB 10 $U/\mu l$)	1 μl
	H_2O	ad 50 μl

The PCR was performed under following cycling conditions using the Primer AdR_PCR (5′-GGTCGCGGCCGAGGATC-3′):

Cycle	Denature	Anneal	Extend
1			72 °C for 10 minutes
2	98 °C for 1 minute	65 °C for 5 minutes	72 °C for 15 minutes
3-6	98 °C for 1 minute	65 °C for 1 minutes	72 °C for 10 minutes
7-24	98 °C for 1 minute	65 °C for 1 minutes	72 °C for 2 minutes

After analysis of the Samples (4 μl) on a Agarose Gel (1 %), the Samples were purified with the Qiaquick PCR purification kit and eluted with EB buffer (15 μl).

3.6.6 DAMID-Next Generation Sequencing (NGS)

For NGS the size of the fragments has to be in the range of 200-500 bp. Therefore, the up to 3000 bp long fragments form DAMID had to be shortened. The samples were diluted (4 $ng/\mu l$ in 50 μl) and sonicated (18 cycles, low setting, 30 sec on, 30 sec off). Then the samples were applied to the NEBNext® Library Preparation Kit. For purification and size selection AMPure® XP Beads were used. For size selection the sample was incubated with 0.33 Vol of AMPure® XP Beads to remove long fragments, which bound to the beads under these conditions. The supernatant was then applied to 0.87 Vol of AMPure® XP Beads, which favours the binding of fragments between 500 and 200 bp. All lower size fragments were still in the supernatant, which was removed at this step. The beads were then washed and eluted and the purified DNA was once more applied to a bead clean up with 0.9 Vol of AMPure® XP Beads. After the final enrichment PCR (15 cycles) the samples were purified with AMPure® XP Beads and eluted in 30 μl. Then the samples were passed to the KFB (NGS Sequencing Facility) for quality control (Bioanalyzer), quantification (KAPA-PCR) and NGS on an Illumina HiSeq 2000 machine.

3.6.7 Quantification by PCR (qPCR)

For relative quantification of DNA or cDNA fragments the real-time quantitative PCR mix SsoFast EvaGreen of Biorad was applied according to the manufacturers recommendations. Here a fluorescent dye is incorporated into freshly synthesized double stranded DNA and measured. The increase of fluorescence signal is measured over the number of PCR cycles performed. For ChIP experiments input samples were diluted 10 fold and 8 μl of IP and diluted input DNA was used. For cDNA analysis 2 μg of total RNA were treated with DNaseI and reverse transcribed. The resulting cDNA was diluted 10 to 20 fold and 8 μl were used per qPCR reaction. Primers were used in a final concentration of 500 nM.

3.6.8 Quantification of Viral Genome Insertions for Transduction Based DAMID experiments

During transduction the viral genome is stably integrated into the host genome. To quantify the insertion rate, a standard plasmid containing the coding sequence for Ribonuclease P RNA Component H1 (RPPH1) and the transfer vector (pLGW) encoding viral sequences were sequentially diluted (10 fold, starting with $1 * 10^7$ molecules per reaction) [Salmon and Trono, 2001]. The samples (Qiagen Blood and Tissue Kit isolated DNA) were diluted 10 fold and measured in a qPCR along with the dilutions. The quantification was performed using two primer sets, the first one specific for RPPH1 to quantify the host genome content [Baer *et al.*, 1990] and the second one directed against the viral GAG gene on the transfer plasmid, which is integrated in the host genome during transduction.

3.6.9 Cleavage Assay

In the cleavage Assay Ago2,the only cleavage competent of the four Ago proteins, is incubated with a radioactively labelled target RNA sequence, resulting in a shorter cleaved RNA fragment, which can be visualised with autoradiography (see Meister *et al.* [2004]).

RNA target Sequence Preparation At first the coding sequence of the target RNA is amplified by PCR. The purified PCR product (80 $ng/\mu l$) was used in a T7 RNA-Polymerase reaction:

T7 reaction mix	NTP (0.2 M each)	10 μl
(100 μl)	DTT (1 M)	1 μl
	Triton X-100 (1%)	1 μl
	Spermidin (1 M)	0.2 μl
	$MgCl_2$	2.5 μl
	Tris-HCl pH 8.0 (1 M)	3 μl
	Template (1 $\mu g/100bp$)	25 μl
	Pyrophosphatase	0.2 μl
	T7 RNA Polymerase (5 mg/ml)	2 μl
	H_2O(DEPC)	as 100 μl

The RNA was transcribed for 4 hours at 37 °Cand was stopped by addition of DNaseI (0.2 μl; 40 $U/\mu l$) for 15 minutes at 37 °C. Subsequently the RNA was

denatured by 2× RNA LD (100 μl) and 5 minutes at 95 °C. The RNA was purified via an UREA-TBE Gel (8%) (Section 3.6.1). After a gel run for 1 hour at 350 V the RNA fragment was visualised by UV shadowing and cut from the gel. To elute the RNA form the gel, it was cut in small pieces and squeezed through a 0.5 ml tube, punctured by a G21 needle (0.8×40mm), by centrifugation at maximum speed for 5 minutes. The gel particles were resuspended in elution buffer (350 μl; 300 mM NaCl, 2 mM EDTA) and incubated on a rotating wheel at 4 °C overnight. On the next day the gel slurry was transferred into a SpinX Column (Costar) and centrifuged for 1 minute at maximum speed. The flowthrough was saved in a low bind tube (Eppendorf) and precipitated by 2.5 vol ethanol (100%). After centrifugation at 17000 $\times g$ for 45 minutes at 4 °C, the RNA pellet was washed with ethanol (80%) and finally dissolved in H_2O(20 μl). The RNA concentration (3.32 $\mu g/\mu l$ was measured and the molecular mass (340 $pmol$) was calculated.

Cap Labelling of the Target RNA In order to make the cleavage reaction visible, the target RNA sequence is cap labelled with ^{32}P-α-GTP. For labelling the following reaction mix is used:

Cap labelling reaction (20 μl)	Guanyltransferase buffer (400 mM Tris-HCl pH 8.0, 60 mM $MgCl_2$,150 mM NaCl, 20 mM Spermidine)	2 μl
	Ribolock	0.25 μl
	SAM (500 μM)	1 μl
	Guanlytransferase	2 μl
	^{32}P-α-GTP	2 μl
	traget RNA (40 pmol)	0.78 μl
	H_2O(DEPC)	ad 20 μl

The cap labelling was performed for 3 hours at 37 °C. The reaction was stopped by addition of 2×RNA LD (20 μl). The radioactive target RNA was then loaded on a 8% UREA gel (Section 3.6.1). The radioactive signal was detected by autoradiography, exposing an film for 5 minutes to the gel. The developed film was then placed under the gel and the labelled target RNA was cut from the gel. The RNA was extracted and precipitated as described before (Section 3.6.9).

Argonaute IP and Cleavage Assay HEK293 cells (10/15 cm plate) were once washed with ice cold PBS. Then the cells were treated with lysis buffer (200/400 μl, 25 mM Tris-HCl pH 7.5, 0.5% NP-40, 150 mM KCl, 2 mM EDTA, 1 mM NaF, 0.5 mM DTT, 0.5 mM AEBSF) and pipetted up and down. The lysate was incubated on ice for 5 minutes and centrifuged at 17000 $\times g$ for 10 minutes at 4 °C. The supernatant was saved into a new tube and frozen in N_2. Protein G Sepharose beads (40 μl per IP) were washed in PBS and incubated in PBS + BSA (5 mg/ml) on a rotating wheel for one hour at 4 °C. The protein lysates were thawn on ice: For V5-tagged proteins, the V5 antibody (2 μg) was added to the lysate and incubated for one hour. Followed by another hour of incubation in the presence with PBS washed Sepharose beads (40 μl). For Flag-tagged proteins, the lysate was incubated for two hours with PBS washed Flag beads (40 μl) at 4 °Con a rotating wheel. After binding of the target proteins, the beads were washed three times for 10 minutes on a rotating wheel. The samples were centrifuged at 700 $\times g$ for 2 minutes at 4 °C to collect the beads. A WB sample (25 %) was taken from the last washing step, pelleted and the protein was eluted by boiling the beads in 2× LD (5 μl) for 5 minutes at 95 °C. The rest of the sample (75 %) were used for cleavage assay:

Cleavage reaction (25 μl)	beads	15 μl
	TM buffer	7.5 μl
	Cap labelled RNA	2 μl
	H_2O(DEPC)	ad 25 μl

TM buffer	ATP (100 mM)	3.33 mM
	GTP (100 mM)	0.67 mM
	RNAsin (33 U/ml)	0.033 $U/\mu l$
	KCl(3 M)	333 mM
	$MgCl_2$ (2 M)	5 mM
	DTT (1 M)	1.67 mM
	H_2O	N.A.

The cleavage reaction was incubated for 1.5 hours at 30 °C. Beads were mixed every 10 minutes at 400 rpm. The reaction was stopped by addition of Proteinase K solution (200 μl, 300 mM NaCl 25 mM EDTA, 2% SDS, 200 mM Tris-HCl, pH

7.5, 0.2 mg/mlProteinase K [20 mg/ml]) and incubation at 65 °Cfor 15 minutes. The samples were extracted with PCI (1 Vol) 10 minutes shaking at 25 °C. The phase separation was achieved by centrifugation at 17000 $\times g$ for 5 minutes. The labelled RNA was precipitated with ethanol (2.5 Vol, 100%) overnight, dried and dissolved in H_2O(10 μl). For loading the RNA was boiled in RNA LD (10 μl) at 95 °Cfor 5 minutes. The cleavage products were separated on a sequencing gel (8%), dried on a gel dryer for 2 hours at 80 °C and visualised by autoradiography.

3.6.10 siRNA Binding Assay

For the Ago1-siRNA binding assay, siRNA#2 guide strand was radioactively labelled by PNK (Fermentas)-mediated 5′-phosphorylation using ^{32}P-γ-ATP (Hartmann Analytics). In addition, siRNA#2 guide and passenger strands were phosphorylated in the presence of cold ATP. Free ATP and PNK were removed from the labelling reactions by gel filtration with MicroSpin G25 columns. Concentrations of the purified siRNA strands were measured, and the radioactively labelled and unlabelled siRNA#2 guide strand was mixed in a ratio of 1:10, respectively. Subsequently, half of the mixture was pre-annealed with an equal amount of siRNA#2 passenger strand. The other half was diluted with an equal volume of water. To test siRNA#2 guide strand and duplex interactions with Ago1, 10 nM of purified His-Ago1 was incubated with 0.5 nM of either siRNA#2 guide strand or duplex in buffer (50 mM HEPES pH 7.5,300 mM NaCl, 10% glycerin, 2 mM $MgCl_2$, 0.5 mg/ml BSA) for 90 min at 24 °C. As control samples, the siRNA single- and double-strands were incubated in the absence of His-Ago1. Afterwards, His-Ago1 was immunoprecipitated from the binding reaction for 60 min by the addition of anti-Ago1 antibody loaded Protein G Sepharose 4 Fast Flow. The IP samples were washed two times with 1 mL of binding buffer lacking BSA and another two times with 1 mL PBS containing 0.01% Tween. Thereafter, the samples were Proteinase K digested as described above. Samples were diluted in 5 mL Unisafe I emulsifying scintillator (Zinser Analytic) and measured on a LS6500 scintillation counter (Beckmann).

3.7 Bioinformatic Methods

For bioinformatic analysis of NGS data commonly used programmes and open source software was used. Where custom solutions were needed, python programmes were written to perform the required task. The data was received a .fastq files encoding for read ID, sequence and phred quality score for each called base of the sequence. The ChIP data contained base qualities encoded as Phred+64, the newer DAMID data sets were encoded with Phred+33 (see http://en.wikipedia.org/wiki/FASTQ_format).

3.7.1 Quality Control by FastQC

In order to get quick, first impression of the data, the quality of $1 * 10^6$ reads were assessed using FastQC [Andrews]. This software displays for example: the per base quality,

per base quality:	helps to identify a trend to low quality bases, often found at start and end of the reads
per base sequence content:	informs about a bias to certain sequences at specific positions, i.e. for DAMID the signature of the PCR primer could be identified
Sequence duplication levels	shows if clonal reads are observed, which normally is based on low sample complexity and overamplification by PCR
overrepresented sequences	checks for recurring sequences in the reads,i.e. detection of adapter sequences form Illumina or PCR primers or adapters form DAMID

3.7.2 Adapter Removal

Contaminating adapters at the 5′ and 3′ ends of reads caused by library preparation and the DAMID procedure were removed using custom made python scripts (www.python.org) called 5p-primerem.py and 3pRC-primerrem.py. These custom scripts required as input the .fastq files and a text file containing all the contaminating adapter sequences, including the DAMID primer sequence: *GGTCGCGGCCGAGGATC*, a truncation at the 5′ end of it and one mismatch mutation at any position. The python scripts can be found in the appendix (5.6)

3.7.3 Trimming

For quality trimming and paired-end consistency the trimmomatic software was used [Bolger *et al.*, 2014]. Quality trimming is performed to remove low quality bases from the start or end of the read, which potentially create mismatches with the reference genome and therefore reduce the quality of the mapping of the read. This improves the subsequent alignment with Bowtie2. The program was invoked using the following parameters:

```
TrimmomaticPE: Started with arguments:\\
-phred33 inputfile-read1.fq.gz inputfile-read2.fq.gz read1-output_paired.fq.gz read1-
    output_unpaired.fq.gz read2-output_paired.fq.gz read2-output_unpaired.fq.gz LEADING
    :26 TRAILING:26 SLIDINGWINDOW:5:20 MINLEN:25
```

Leading	cut bases at the start of read with quality below threshold
Trailing	as leading but for end of the read
Slidingwindow	a slidingwindow of 5 bases must have an average quality higher than threshold
Minlen	the minimal length required to be taken forward

3.7.4 Read Alignment with Bowtie2

For read alignment the Bowtie2 package of Langmead *et al.* [2009] was used. The indexes for *h.sapiens*, hg19, and for *m.musculus*, mm10, were taken from *http://bowtie-bio.sourceforge.net/bowtie2/index.shtml*. Genome alignments were performed on the Athene HPC-Cluster of the University of Regensburg using the following code to invoke the program.

```
Bowtie2: Started with arguments:
bowtie2 -p 8 --phred33 -q -D 20 -R 3 -N 0 -L 20 -i S,1,0.50 --rdg 5,10 --rfg 5,10 -t -X
    700 --no-mixed --no-discordant --no-unal -x hg19 -1 read1-output_paired.fastq -2
    read2-output_paired.fastq -S Bowite-output.sam
```

p	cpu nodes
phred33	quality scores used in the .fastq file
q	a slidingwindow of 5 bases must have an average quality higher than threshold
D	number of failed seed extensions before bowtie2 moves on
R	number of times bowtie2 will "re-seed" reads with repetitive seeds
N	number of mismatches allowed in a seed alginment during multiseed alignment
L	length of seed substrings to align during multiseed alignment
i	function governing the interval between seed substrings during multiseed alignment
rdg	gap penalties for reads
rfg	gap penalties for reference
t	print wall-clock time required to load index and align reads
k	number of valid alignments looked for, if more than one is found
X	maximum insert length
x	reference genome
1,2	input: forward and reverse reads as tow separate .fastq files
S	output as .sam files

For more information on Bowtie2 the settings,
see *http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml*

3.7.5 GATC-genome

In order to select for reads resulting from DamID only and not from non specific amplification of sequences closely related to the DAMID-PCR primer, a GATC genome was constructed. A python script was created scanning each chromosome for the GATC sequence, producing one file per chromosome containing the coordinates of all GATCs existing in the sequence from the hg19 genome build. The python script can be found in the appendix (5.6).

3.7.6 R Analysis

The data analysis was performed with R and the DESeq package. Scripts are supplied in the appendix (5.6).

Chapter 4

Results

The main function of Ago proteins in the cell is the well characterized silencing of mRNA expression in the cytoplasm [Huntzinger and Izaurralde, 2011]. However, during the last decade also nuclear functions of Ago proteins have been described for plants, yeast, worms and flies [Gagnon and Corey, 2012; Huang and Li, 2014]. In these studies, Ago proteins were implicated in DNA methylation, heterochromatin formation and transposon silencing processes. But also human Ago proteins were shown to be capable of TGS when artificially recruited to gene promoters by exogenous small RNAs (see Table 1.1). In addition, several studies were published applying Ago-specific ChIP-ChIP or ChIP-Seq experiments under different conditions, which suggested chromatin interactions of Ago proteins and their involvement in regulation of senescence, splicing or gene activation [Benhamed *et al.*, 2012; Ameyar-Zazoua *et al.*, 2012; Allo *et al.*, 2009; Huang *et al.*, 2013a]. However, endogenous Ago-chromatin interaction sites in somatic cells as well as the functional mechanisms are still controversially discussed. Therefore, we set out to identify gene loci showing Ago-chromatin association mediated by endogenous targeting mechanisms.

4.1 Biochemical Analysis of Argonaute Localisation

As a first step the cytoplasmic and nuclear localisation pattern of Ago proteins, as described by [Robb *et al.*, 2005; Rüdel *et al.*, 2008; Ohrt *et al.*, 2008], was reproduced applying cellular fractionation to HeLa S3 cells (see Figure 4.1). In addition to

the accumulation of Ago2 in the cytoplasm also localisation to the nucleus could be confirmed. The nuclear pool of Ago2 was further precipitated with an Ago2 specific antibody (4E1 hybridoma supernatant) to identify the Ago2 specific band in the WB (α-Ago2 (11A9)).

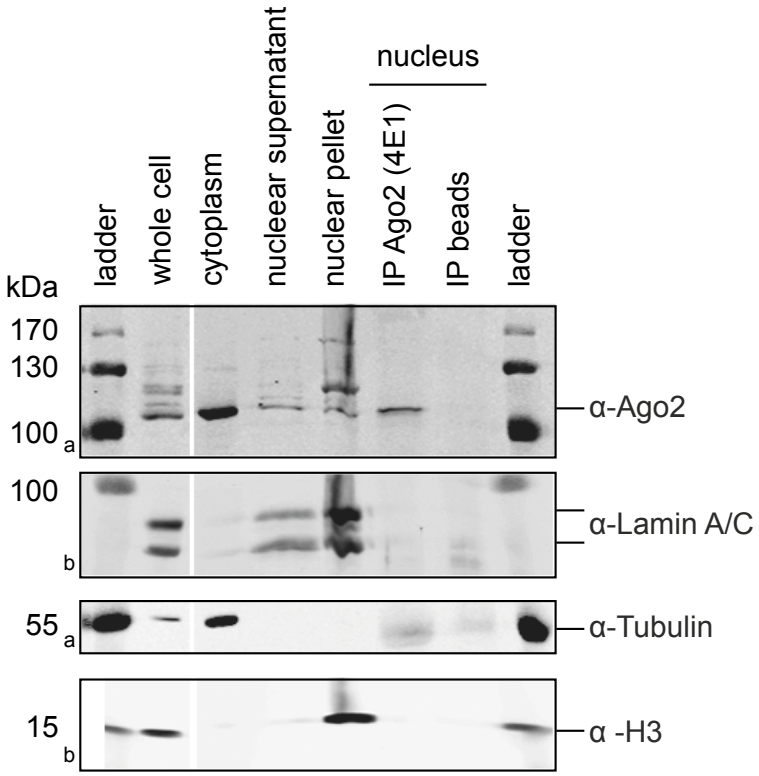


Figure 4.1: ***Cellular Fractionation of HeLa S3 cells.***
HeLa S3 cells (10^9) were harvested washed and lysed in hypotonic buffer. After douncing (Pestle B 10 strokes) the nuclei were extensively washed and lysed in Triton X-100 (0.5%) containing buffer. There is a strong Ago2 signal in the cytoplasm and a weaker one in the nuclear supernatant and pellet. Both fractions are quite clean according to the nuclear lamin A/C and cytoplasmic tubulin controls. Nuclear Ago2 was successfully precipitated with the Ago2 (4E1 hybridoma supernatant) lacking a contaminating band described for the 11A9 antibody used by Rüdel et al. [2008]. Loading: ~20 µg per lane. WB: α-Ago2 (11A9). Samples were loaded in equal amounts and order on two 10% SDS-PAGE gels (a, b). Further analysis by WB was performed in parallel.

In a second experiment the chromatin association of Ago2 was examined similar to an approach recently described by Cernilogar *et al.* [2011]. In this assay the cells were lysed with a Triton X-100 containing buffer to remove the cytoplasm and non chromatin associated proteins (S1) from the chromatin fraction (P1) (Fig-

ure 4.2). Then, the chromatin was treated with DNase and salt, to release its associated proteins into the soluble fraction (S2). The nuclear matrix components were pelleted by centrifugation (P2) (see Figure 4.2). In this assay Ago proteins could be detected in the S1 fraction but also in the chromatin associated P1 and S2 fraction. This finding implies a potential involvement of Ago proteins in chromatin associated processes ranging from regulation of transcription or splicing to an influence on chromatin structure.

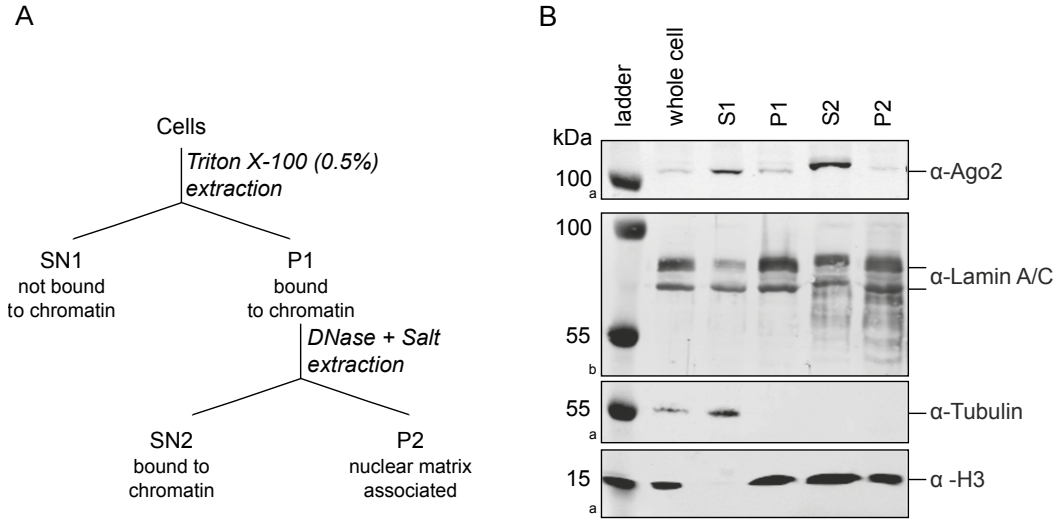


Figure 4.2: ***Analysis of the chromatin association of Ago proteins.***
A) *Outline representing the different fractionation steps of the chromatin association assay.* ***B)*** *Western blot probing the different fractions for Ago2 and control proteins. HeLa cells* $(4,4 * 10^8)$ *were lysed applying a Triton X-100 comprising buffer. The resulting nuclei were washed and treated with DNase to degrade the chromatin and eventually releasing associated components into the supernatant. Ago2 was found together with histone H3, a tight bound chromatin marker, in the chromatin containing fractions P1 and S2. The cytoplasmic pool of Ago proteins was found in the first supernatant (S1). The soluble protein marker tubulin was only found in the lysate and soluble fraction. Lamin A/C was used to detect the nuclear matrix. Loading: ~20* μg *per lane. Samples were loaded in equal amounts and order on two 10% SDS-PAGE gels (a, b). Further analysis by WB was performed in parallel.*

4.2 Analysis of Argonaute Recruitment to Chromatin by ChIP

Information on endogenous triggers for Ago-chromatin association, recruitment mechanisms or its role in the nucleus is sparse and needs further independent

validation ([Kim *et al.*, 2008; Hansen *et al.*, 2011; Benhamed *et al.*, 2012; Huang *et al.*, 2013a]). As a first experimental approach to identify possible target regions on the chromatin ChIP-sequencing was established and utilized.

4.2.1 Sonication Conditions

The first step in establishing a ChIP protocol in the lab is to find proper sonication and crosslinking conditions for the cell type used. Therefore, a timecourse experiment with HeLa cells was performed. Following the ChIP protocol (Section 3.6.4), cells ($\sim 3 * 10^7$) were crosslinked for 10 minutes with formaldehyde (1%) at RT, which forms covalent bonds between proteins and nucleic acids at a distance of about 2 Å [Das *et al.*, 2004]. This was shown to be sufficient, to crosslink Ago proteins with chromatin [Janowski *et al.*, 2006]. After preparation of nuclei and lysis (1.2 ml), the suspension was sonciated and samples (20 μl) were taken at different time points (0, 5, 10, 15, 20, 25, 30, 35 minutes). Then the crosslink was reversed by incubation at 65 °C and the DNA was precipitated and analysed by agarose gel electrophoresis. A shortening of the fragment length over time was observed, which reaches a halt at treatment times of 25 to 30 minutes (Figure 4.3 A). Therefore, a sonication time of 30 minutes which corresponds to a main DNA fragment size of 100 to 300 nt was chosen for the ChIP-Sequencing experiment.

4.2.2 Control of ChIP Conditions with the Transcription Factor UBF

The ChIP procedure was controlled by a test experiment using the upstream binding factor (UBF), an rDNA TF [Prieto and McStay, 2007]. UBF binds throughout the rDNA repeat and leads to formation of a specialized chromatin structure, found in nucleolar organizer regions (NORs). The binding of UBF in complex with the essential transcription factor SL1 is needed for efficient RNA polymerase I transcription initiation [Friedrich *et al.*, 2005]. For this ChIP analysis, H1299 cells ($\sim 3 * 10^7$) were crosslinked for 10 minutes with formaldehyde (1%) at RT, nuclei were prepared and the lysate was sonicated for 25 minutes. For the IP, protein G sepharose beads were prepared as described in Section 3.5.2, including a blocking step with salmon sperm DNA (500 $\mu g/ml$) and BSA (100 $\mu g/ml$) for two hours at 4 °C. Commercial antibodies specific for UBF and RNA Pol-II (phospho S2) were used in the ChIP and were bound to the blocked beads for

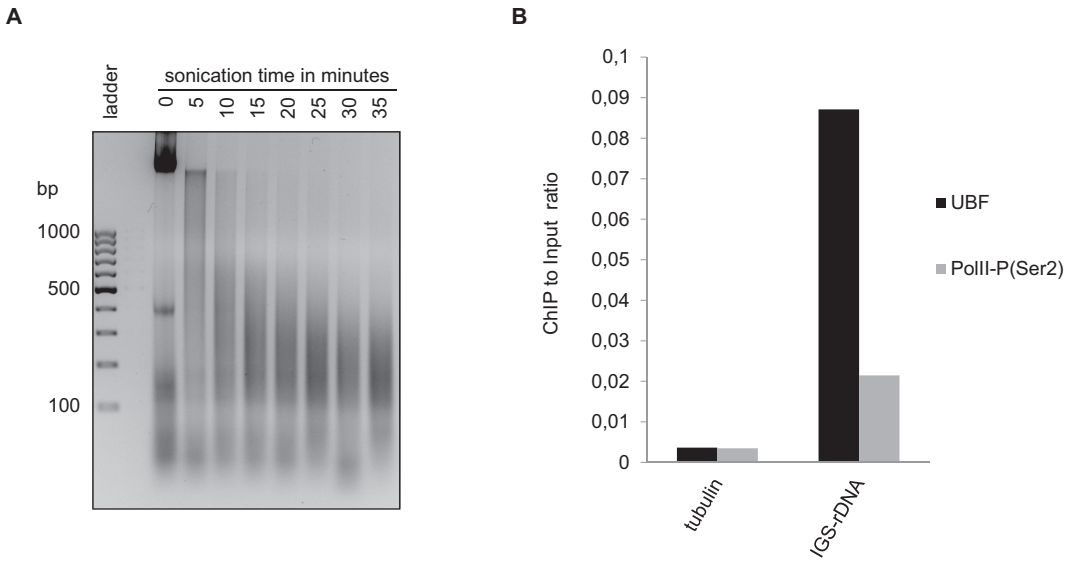


Figure 4.3: ***Establishing of a ChIP experiment.***
A*) HeLa cells (*$\sim 3 * 10^7$*) were crosslinked for 10 minutes with 1% formaldehyde at RT. Nuclei were prepared, lysed and sonicated for the indicated times. The samples (20* μl*) were reverse-crosslinked and purified. Then the DNA (2* μg*) was analysed by agarose gel electrophoresis (2 %).* ***B****) H1299 cells were crosslinked for 10 minutes with 1% formaldehyde at RT. Nuclei were prepared and used for ChIP experiments with antibodies specific for UBF and RNA PolII (phospho S2). An enrichment of UBF at the IGS was observed compared to RNA Pol-II. Both proteins showed no association with the control locus (tubulin-*β *promoter). One of two independent experiments is shown.*

two hours at 4 °C. RNA PolII has been shown to be phosphorylated at serine 2 of the CTD repeats during the middle phase of transcript elongation, but not during the initiation of transcription at the TSS [Kim *et al.*, 2010]. The purified DNA from the ChIP and the input sample were analysed by qPCR. As expected, there was no accumulation of either of the proteins at the control locus tubulin-β promoter (TSS). However, the intergenic spacer region of the rDNA (IGS-rDNA) was clearly enriched for UBF (see Figure 4.3 B). Thus, a successful ChIP of the known chromatin binding protein UBF indicates that the protocol is established and can be used for analysis of Ago proteins.

4.2.3 Analysis of Argonaute Precipitation under ChIP Conditions

The first Ago ChIP was published by Janowski *et al.* [2006], for which they used commercial antibodies against Ago1 and Ago2 designed and tested for WB application (Merck, Germany). A test on the specificity of their antibodies under the rather harsh ChIP conditions could, however, not be found in their study. Since antibodies are key for the planned experiments and these reagents frequently cause unspecific results, the ChIP sequencing experiment was designed to be based on a diverse set of antibodies. Several high specificity antibodies against

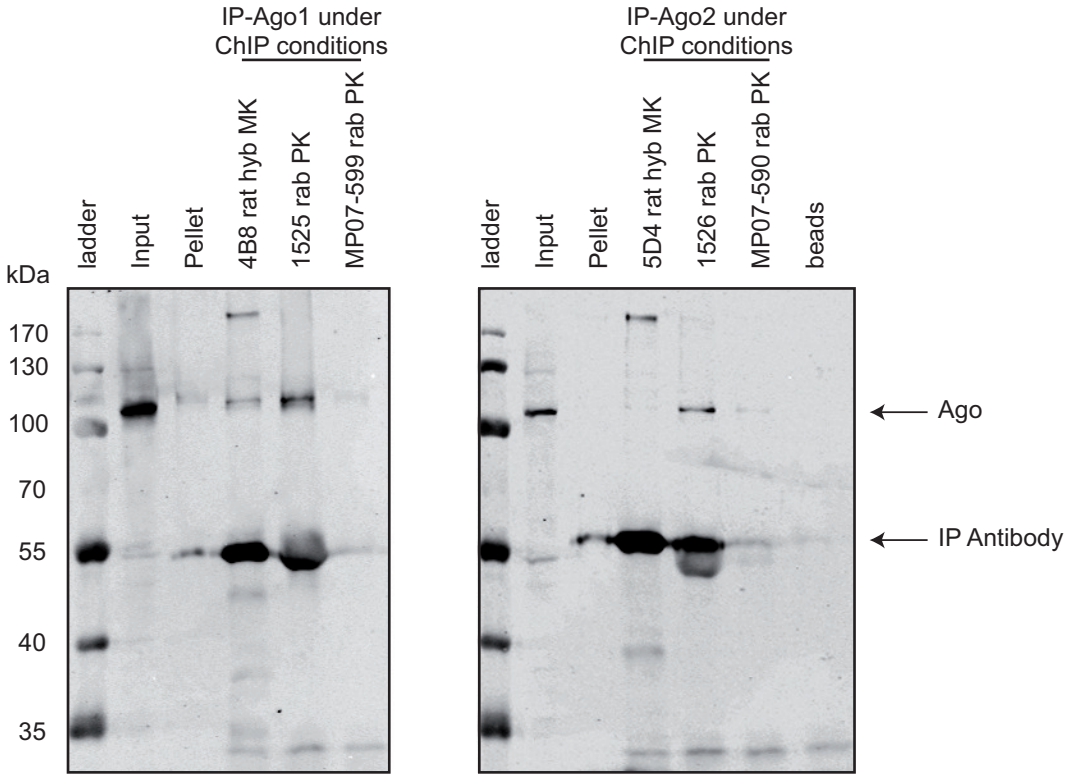


Figure 4.4: ***Immunoprecipitation of Ago1 and Ago2 under ChIP conditions analysed by WB.***

HeLa cells were crosslinked for 15 minutes with 1% formaldehyde at RT. Nuclei were prepared and lysed by sonication for 10 minutes. The two Ago paralogues were precipitated by protein-G sepharose beads (GE healthcare) blocked with salmon sperm DNA (500 µg/ml) and BSA (100 µg/ml) using three different antibodies for each protein. A fraction (4%) of the eluted chromatin complexes and the input (1%) was analysed by WB using antibodies specific for Ago1 (1C9) and Ago2 (11A9). All three antibodies seemed to precipitate Ago proteins, however, the sera 1525 and 1526 showed the highest affinities.

both Ago1 and Ago2 are available in the lab allowing for ChIP experiments with three different antibodies, including the ones already published [Janowski *et al.*, 2006]. The antibodies chosen for the ChIP of Ago1 consist of a hybridoma supernatant (rat) 4B8, the rabbit serum 1525 and the commercial antibody 07-599 of Merck (Germany). The set of antibodies specific for Ago2 comprises the hybridoma supernatant 5D4, the rabbit serum 1526 and the commercial antibody 07-590 of Merck (Germany). Given that Ago proteins do not contain a DNA binding domain, it was assumed that Ago proteins bind to chromatin mediated by a nascent transcript (compare Table 1.1), which would in turn be recognized by the associated miRNA, or by an unknown protein. Therefore, a longer crosslinking time (15 minutes) was tested for the experiment on the pull-down efficiencies of the different antibodies. The concentrated commercial antibodies (5 μg) were directly added to the diluted ChIP lysate and incubated overnight. On the next morning the antibody Ago complexes were captured by an incubation with blocked protein G sepharose beads for two hours at 4 °C. All other antibodies (100 μl of serum for rabbit pAB or 1 ml of hybridoma supernatant for rat mAB) were bound overnight to blocked protein-G sepharose beads on the previous day. Then, they were incubated with the diluted ChIP lysate overnight. The ChIP conditions and antibody amounts were chosen comparable to Janowski *et al.* [2006], except that 1/10 of the beads was used for analysis of the precipitated proteins by WB.

The western blot analysis of the anti-Ago1 immunoprecipitation under ChIP conditions suggested a low affinity of the commercial antibody, which improved for the hybridoma supernatant and was highest for the rabbit serum (Figure 4.4, left). For Ago2, the hybridoma supernatant showed a weak immunoprecipitation efficiency, which was slightly better for the purified commercial antibody. The rabbit serum, in turn, bound the highest amount of Ago2 (Figure 4.4, right). The fragmentation pattern of input DNA after sonication was also analysed on a 2% agarose gel. The extended crosslinking time of 15 minutes had an adverse effect on the sonication efficiency and was therefore abandoned.

4.2.4 Possible Recruitment of Ago to Chromatin Examined by ChIP-Sequencing

In order to map endogenous chromatin interaction sites of Ago proteins a ChIP-Sequencing experiment was performed. HeLa cells ($\sim 3 * 10^7$ per IP) were formaldehyde crosslinked for 10 minutes. After isolation of the nuclei, the lysate was sonicated for 30 minutes. For immunoprecipitation, the antibody amounts were kept as before, only the amount of the commercial antibodies was increased to 10 μg per IP. The precipitated DNA was depleted for contaminating RNA or protein and submitted to Fasteris (Plan-les-Ouates, Switzerland) for library preparation followed by NGS.

4.2.5 Data Statistics from the NGS Experiment

Fasteris processed the samples as paired-end library including barcodes sequences for multiplexing. Four samples were mixed (GPK1-4; GPK5-8) and measured on a HiSeq 2000 (Illumina) in one lane of a flow cell for 100 sequencing cycles. The base quality of the reads, which is the probability that the corresponding

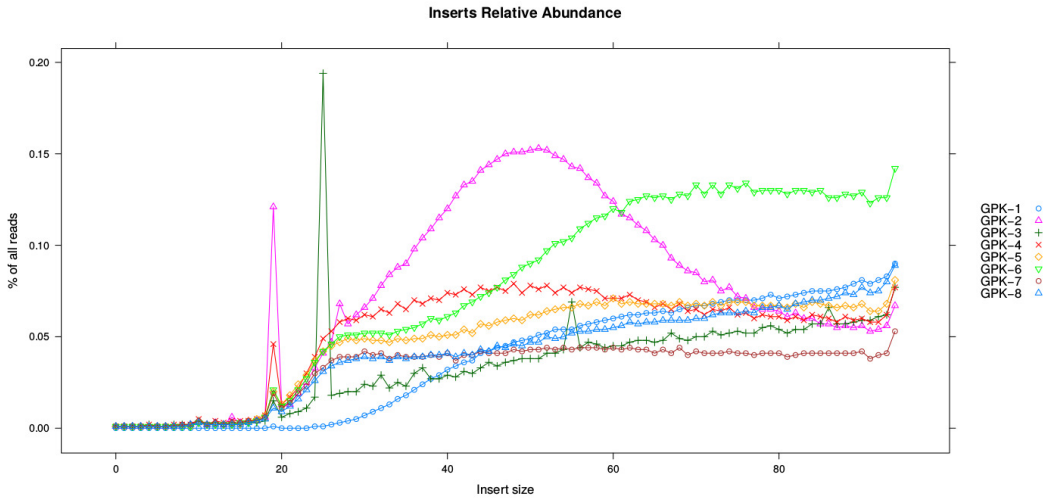


Figure 4.5: ***Insert length against read abundance plot of ChIP-Seq.***
All reads were searched for occurrence of the 3′-adapter sequence. The sequence was trimmed off and the length of the resulting inserts was plotted against their relative abundance. The abundance was estimated as a fraction of all insert containing reads. A peak at the 20th base is observed for all libraries beside the input library, which was supplied at a higher concentration.

base call is incorrect, was reported as Phred+64 score. The attributing of the reads to the different libraries based on their indexes succeeded in 98 % of the reads. Furthermore, the quality of the reads was in 83% of the cases equal or higher 30, which is 1 of 1000 bases is incorrect or there is one mismatch in 1 of 10 reads (length 100).
About 2% of the reads contained an adapter sequence, which was removed by the company. In addition, a few reads from each library contained the 3′-adapter from the 20th base on (Figure 4.5). This is a sign for the low DNA concentration of the libraries according to Fasteris.

Sample Name	Library code	Index	File name	Reads
Input	GPK-1	CGCATC	110121_SN365_B_s_4_seq_GPK-1.txt	23907169
A1-4B8	GPK-2	CTGAGG	110121_SN365_B_s_4_seq_GPK-2.txt	11116009
A1-Rabbit	GPK-3	GTCGCT	110121_SN365_B_s_4_seq_GPK-3.txt	13992230
A1-Millipore	GPK-4	TTTCTT	110121_SN365_B_s_4_seq_GPK-4.txt	11358686
			Total:	60374094
A2-5D4	GPK-5	TGATAC	110121_SN365_B_s_5_seq_GPK-5.txt	12537320
A2-Rabbit	GPK-6	GGTCGC	110121_SN365_B_s_5_seq_GPK-6.txt	14489939
A2-millipore	GPK-7	GACAGA	110121_SN365_B_s_5_seq_GPK-7.txt	12361720
Beads	GPK-8	AACGCG	110121_SN365_B_s_5_seq_GPK-8.txt	16649529
			Total:	56038508

Table 4.1: ***NGS run information on the HeLa ChIP-Seq experiment.***
The high-throughput sequencing experiment was performed on two lanes of a flow cell and yielded 24 Mio reads for the input and 11 to 14 Mio reads for the different Ago-specific antibodies.

The output of the NGS run is summarized in Table 4.1. The highest read number was achieved for the input sample, which was supplied in sufficient amounts. The amount of reads gained from the NGS for all other samples was a lot lower and ranged from 11 to 16 million. This is due to the low amount of DNA normally precipitated in a ChIP experiment, resulting in quite limited starting material for library preparation. This is also supported by the absence of a peak at 20 bp for the input sample in Figure 4.5, which is present for all immunoprecipitated samples.

4.2.6 Results from the Bioinformatic Data Analysis

The bioinformatic analysis was performed by Eugene Berezikov (Groningen, Netherlands). After quality control, the reads were mapped to the human genome

name	ID	Input	Not aligned	Aligned exactly 1	Aligned multiple	Alignment rate
Input	GPK-1	22744602	668389	17217782	4858431	97.06%
A1-4B8	GPK-2	9463861	1003795	5890907	2569159	89.39%
A1-Rabbit	GPK-3	12504515	1139482	7447732	3917301	90.89%
A1-Millipore	GPK-4	9935458	1599115	5074154	3262189	83.90%
A2-5D4	GPK-5	11110703	945361	6565332	3600010	91.49%
A2-Rabbit	GPK-6	12696692	1139457	7482646	4074589	91.03%
A2-millipore	GPK-7	10828620	1342803	6638182	2847635	87.60%
Beads	GPK-8	14846584	4834654	6944353	3067577	67.44%

Table 4.2: ***Sequencing depth and mapping of the ChIP-Seq experiment.***
The sequencing experiment yielded about 23 Mio reads for the input sample and between 9.4 and 14.8 Mio reads for IP-Samples. The alignment with the human genome was with 83 to 97% quite good for input and IP-Samples and reduced for the beads control.

(hg19) by using the genomatix software (see Table 4.2). The mapping efficiency ranged form 83 to 97 % and resulted in 9.4 to 14.8 million reads for the IP-Samples, which could be used in the subsequent analysis. The peak calling was performed with MACS [Liu, 2014] with the model fold values (10, 30) and a p-value cut off of 0.00001.

overlap	A1-4B8	A1-Rabbit	A1-Millipore	All peaks
A1-4B8	578	30	42	8
A1-Rabbit		1106	110	8
A1-Millipore			3699	8

A) *The number of total and overlapping peaks for Ago1*

overlap	A2-5D4	A2-Rabbit	A2-Millipore	All peaks
A2-5D4	2863	71	69	10
A2-Rabbit		871	17	10
A2-Millipore			715	10

B) *The number of total and overlapping peaks for Ago2*

Table 4.3: ***ChIP Peak Analysis by MACS.***
For all samples peaks were determined by MACS using the empty beads sample as control. The resulting peaks were compared for overlaps. Unfortunately, there were only a few conserved peaks between antibodies, which even decreased, when all three antibodies were included

When the ChIP samples were compared to the bead control for enriched sites, several hundred peaks were found for each sample (see Table 4.3). Given the low number of peaks especially for ChIP samples from the antibodies A1-4B8 and A2-Rabbit, the p-value cut off seems rather strict. In addition, the overlap of

peaks between the different antibodies is a very stringent criterion, leading to selection of only 8 common peaks for Ago1 and 10 for Ago2. Thus, one could think about relaxing the p-value cut off to increase the pool of peaks for the selection of overlaps.
For the few overlapping peaks from all three antibodies genomic annotation was retrieved, to identify a potential common pattern (Table 4.4, *see next page*). The annotation demonstrated, that the peak caller identified repeat associated regions. A fraction of the peaks localised to centromeric alpha satellites. Manual inspection of the peak profile in the genome bowser relative to input reads and beads control convinced of false positive binding for the respective sites.

In summary, no strong enrichment of Ago1 and Ago2 at the chromatin compared to the bead control could be detected under the applied experimental conditions. However, due to the low number of overlapping peaks between the different utilized antibodies also no pattern analysis for binding sites could be used. Additionally, weak interaction sites may have been missed by our analysis due to the limited sequencing depth and the lack of biological replicates for each ChIP-experiment. Other possible explanations of these negative results could be a low precipitation efficacy of chromatin-bound Ago by the utilized antibodies, the limited fractionation efficiency of crosslinked cells, causing cytoplasmic Ago proteins to be the dominant species in the ChIP lysate or high background noise, due to the use of non-purified antibodies (4B8, 1525, 5D4 and 1526) in the IP. In order to address these different questions we pursued two strategies. First, to improve our results regarding antibody performance in ChIP experiments and to reduce background noise, antibodies specific for Ago2 were generated and/or purified against recombinant Ago2. Second, an antibody independent technique for chromatin protein interaction, called DNA adenine methyltransferase identification (DAMID), was applied.

			MACS pvalue			repeat			gene	epigenetic mark	TF
source	start	end	A1-4B8	A1-Rabbit	A1-Millipore	Family	class	Name			
chr11	102601691	102602063	136.26	149.74	67.69	Simple_repeat	Simple_repeat	(CCCA)n		H3K27ac	TCF12
chr12	38017825	38018746	71.96	56.86	63.08	centr	satellite	ALR/Alpha			
chr21	46760203	46760651	63.67	61.74	66.49	Simple_repeat	Simple_repeat	(TCCA)n	LINC00316		
chr5	49438963	49439618	62.78	69.89	214.4	centr	satellite	ALR/Alpha			
chr7	71673995	71674645	67.67	70.98	83.69	Simple_repeat	Simple_repeat	(TTCC)n	CALN1		
chrX	61695711	61699642	63.58	83.89	100.63	centr	satellite	ALR/Alpha			
chrX	61696038	61699642	68.34	83.89	100.63	centr	satellite	ALR/Alpha			
chrX	61760501	61761127	71.5	51.85	186.94	centr	satellite	ALR/Alpha			

A) *Ago1 peaks*

			MACS pvalue			repeat			gene	epigenetic mark	TF
source	start	end	A2-5D4	A2-Rabbit	A2-Millipore	Family	class	Name			
chr20	26260782	26261233	122.85	76.05	70.89	centr	Satellite	ALR/Alpha			
chr3	90492038	90492626	57.11	68.1	53.89	centr	Satellite	ALR/Alpha			
chr8	43834378	43838882	324.4	101.7	127.74	centr	Satellite	ALR/Alpha			
chr9_gl000199_random	103585	108513	337.54	57.72	146.22	centr	Satellite	ALR/Alpha			
chr9_gl000199_random	103602	110944	137	57.72	146.22	centr	Satellite	ALR/Alpha			
chrX	58566985	58575520	57.13	50.92	132.11	centr	Satellite	ALR/Alpha			
chrX	58566985	58576276	77.91	50.92	132.11	centr	Satellite	ALR/Alpha			
chrX	58566985	58577230	91.85	50.92	132.11	centr	Satellite	ALR/Alpha			
chrX	61694941	61697568	125.19	57.44	74.17	centr	Satellite	ALR/Alpha			
chrX	61694941	61701789	264.29	57.44	74.17	centr	Satellite	ALR/Alpha			

B) *Ago2 peaks*

Table 4.4: ***Annotation for overlapping ChIP peaks of three different antibodies for each Ago protein.***

The overlapping peaks from the three antibodies used for the Ago1 ChIP experiment (A) and Ago2 ChIP experiment (B) were annotated. Additional information on overlapping repeats, genes, TF or histione modifications was added based on UCSC.

4.3 Generation and Purification of Ago Antibodies

The generation and purification of antibodies against a target protein relies on its recombinant expression. The Ago2 protein was the first time successfully expressed in *E. coli* by Rivas *et al.* [2005] as a GST-fusion protein in HSP90 expressing bacteria. However, after several unsuccessful trials in *E. coli*, SF21 insect cells were tested by our group for Ago expression. The cDNA sequences of Ago1 to Ago4 were cloned into the pFastBac-HTA vector (Invitrogen) by J.Pfaff and L.Weinmann.

4.3.1 Expression of Ago Proteins in SF21 Insect Cells

The insect cell expression system utilizing a bacmid encoded Autographa californica nuclear polyhedrosis virus in combination with SF21 cells to produce the protein of interest was used [Fitzgerald *et al.*, 2006]. It is based on the life cycle of the genetically modified baculovirus, which consists of an early phase (0-6 hours post infection (hpi)), a late phase (18-36 hpi) and a very late phase (24-96 hpi) [Invitrogen, 2002]. During the early phase, the virus gains control over the cell and prepares for viral DNA replication. Subsequently, DNA replication starts and membrane coated virus vesicles bud from the cell, to infect neighbouring tissue. The last phase leads to very high expression of p10 and polyhedrous genes. The corresponding proteins surround the viral genome and form, so called polyhedral inclusion bodies, which are released by cell lysis [Okano *et al.*, 2006]. The proteins to be expressed are cloned under the control of the p10 or polyhedrous promoter and accumulate therefore during the very late phase of infection. The pFastBac-HTA vectors carrying the Ago cDNA were transformed into DH10B MultiBac-YFP [Fitzgerald *et al.*, 2006]. Additionally, this bacmid co-expresses YFP under the polyhedrous promoter and thus facilitates screening and titration experiments. After reconstitution of the virus, the initial virus stock (V_0) was generated (see Section 3.2.6). The expression of HIS-TEV-Ago by cells infected with V_0 was confirmed by SDS-PAGE stained with coomassie and also by western blotting using Ago2-specific antibodies. Subsequently, V_1 was produced and titrated for large scale infections. The procedure and workflow for expression of recombinant HIS-TEV-Ago2 is illustrated in Figure 4.6. The titration was monitored over a time of 72 hours after the day of proliferation arrest (DPA) by SDS-PAGE and quantification of the YFP fluorescence (Figure 4.6). Measuring the fluorescence of

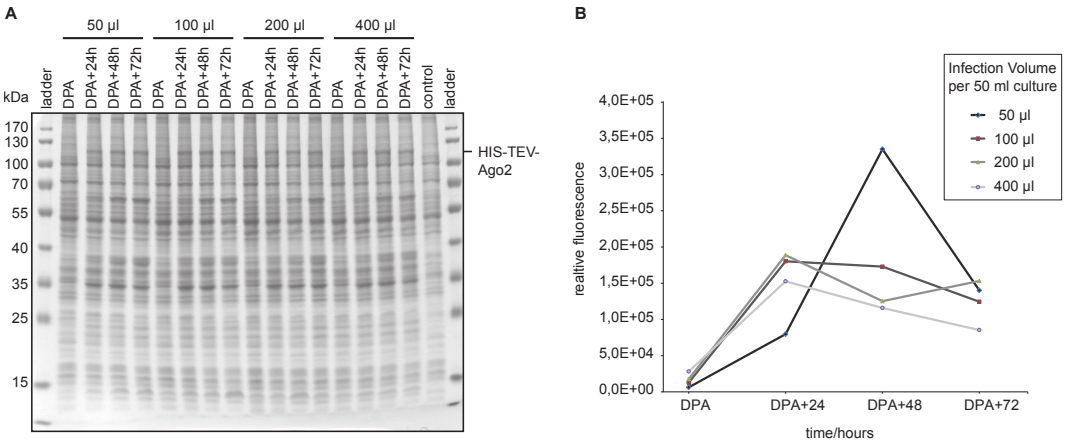


Figure 4.6: V_1 *Virus titration for Ago2.*
To determine the optimal virus concentration for large scale infections, a test experiment was performed. Insect cells were infected with different volumes of genetically modified baculovirus and were monitored for 72 hours after growth arrest. ***A****) Analysis of Ago2 expression by SDS-PAGE and coomassie staining under different infection conditions. Ago2 was strongly expressed at 48 hours in all treatments.* ***B****) The viral infection was monitored using the co-expressed YFP. Here, the over-infection by higher volumes became obvious. For the smallest volume, still leading to growth arrest (50 µl), a "burst of expression" was observed for YFP, as described for p10 and polyhedrin genes [Okano et al., 2006].*

the co-expressed YFP in cell lysates helped to detect the burst of transcription in the very late phase of infection, which was described for proteins under control of the very late promoters, p10 and polyhedrin [Okano *et al.*, 2006](Figure 4.6 B). As all infection volumes led to growth arrest after 24 hours, the smallest infection volume was chosen for the large scale infection to prevent negative effects of over-infection. Furthermore, according to the YFP quantifications, the highest amount of the recombinant expressed proteins was observed at 48 hours. Given that huge amounts of cells started to die at 72 hours DPA, 48 hours seemed to be the optimal time point for harvesting the cells.

4.3.2 Purification of Ago Proteins

The SF21 cells derived from the ovaries of *Spodoptera frugiperda* are very sensitive to detergent. In case the target protein is solubly expressed, gentle lysis conditions involving 0.1% NP-40, vortexing and Benzonase treatments are sufficient to acquire high quality protein. As Ago proteins were soluble in the cytoplasm, gentle lysis was applied. Following clarification by centrifugation and filtration, the

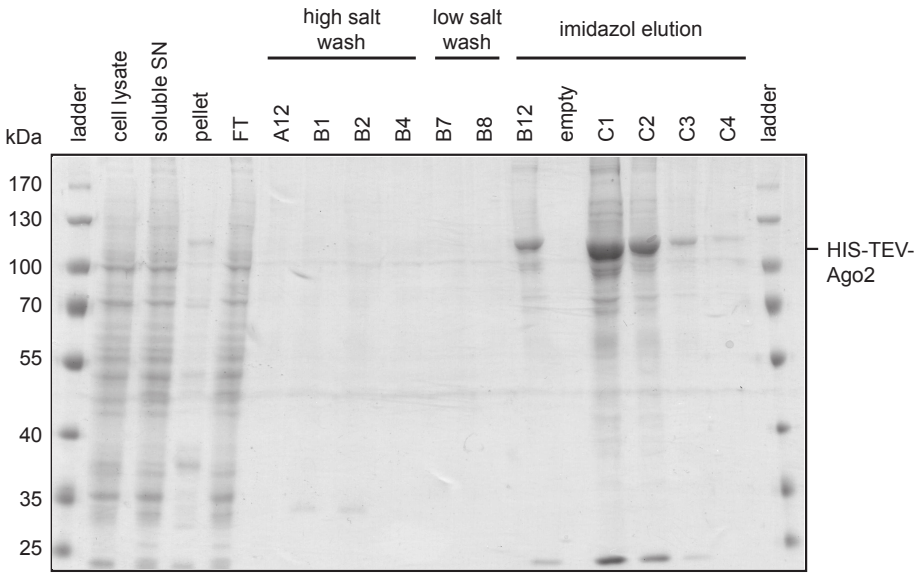


Figure 4.7: ***HIS-TEV-Ago2 purification with Ni-IMAC-column.***
HIS-TEV-Ago2 was expressed in SF21 cells. The clarified lysate was then applied to a Ni-IMAC-column, connected to an Äkta purifier system. After three washing steps with different salt concentrations, the protein was eluted with an imidazol containing buffer.

lysate was loaded on a Ni-IMAC-column, connected to an Äkta purifier system. The Ni-IMAC-column was washed with a high salt buffer (2 *M* NaCl) to remove nucleic acids, which are in part non-specifically bound, due to Ago2's high pI value of 9.32. For the elution step, several schemes were tried. An elution gradient, did not lead to pure fractions of the protein but rather to its dilution. A two step protocol starting with 3% elution buffer for 10 column volumes, followed by a steep gradient to 50% elution buffer, led to a dirty first peak, which was followed by a clean, but minor second peak. To generate more Ago in the clean fraction an elution was performed with 50% elution buffer to recover all HIS-TEV-Ago2. The first purification round yielded already a quite pure protein, see Figure 4.7 fractions B12 to C4.

After TEV cleavage, the sample was loaded again on a Ni-IMAC-column. Nearly all of the co-purified proteins stayed on the column, while Ago2, deprived of its 6×-HIS-tag, accumulated in the flow-through (Figure 4.8 A). After concentration of the protein, the sample was loaded on a gelfiltration column (Superdex 16/60). Ago2 eluted in two peaks: one around 39 to 64 *ml* (fractions A1-A5) and a second one around 65 to 80 *ml* (Figure 4.8 B, fractions A6-A8).

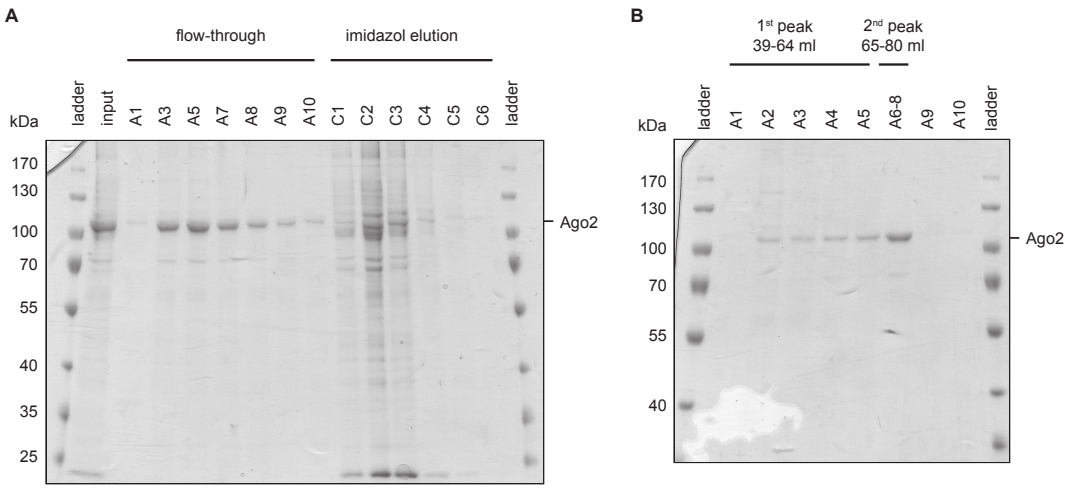


Figure 4.8: ***Ago2 purification by Ni-IMAC- and gelfiltration column.***
A*) The HIS-TEV-Ago2 protein eluted from the first Ni-IMAC purification was dialysed overnight in TEV protease containing buffer. The protein, now released from the HIS-tag, was loaded on a Ni-IMAC-column, to remove the HIS-tag and the TEV protease. Ago2 was recovered in the flow-through and concentrated by a membrane based concentrator.* ***B****) Then the protein was loaded on a gelfiltration column. Two peaks were observed in the 280 nm absorption profile. Both peaks contained Ago2. The first peak with a high 260 nm absorption comprised presumably nucleic acid bound Ago2, whereas the second had a lower 260 nm reading.*

The fractions of the second peak were again concentrated in a centrifugal concentrator and used for rabbit polyclonal antibody production by Davids Biotechnology (Regensburg, Germany).

4.3.3 Side Project on Recombinant Expression of PIWI Proteins

In addition to Ago1 and Ago2, I also expressed other members of the Argonaute protein family, the Piwi proteins in SF21 insect cells. Until today, biochemical approaches characterizing human PIWI proteins lack far behind the genetic information gained. Therefore, trials for recombinant expression of Piwi proteins were initiated. Based on the positive results for Ago proteins, human Piwis were cloned into pFASTBac HTA by A.Frohn, a former graduate student. The reconstitution of baculoviruses expressing Hili, Hiwi and Hiwi2 worked nicely. The viral titres were also good enough to infect host cells and led to the expected growth arrest. However, the protein production was really low and batch purification of these constructs with an affinity matrix did not yield protein amounts, convinc-

ing of a successful large scale production. Given that Piwi proteins play a role in germ cell development, the overexpression of these proteins in an ovary derived cell line (SF21) might not be tolerated. This may lead to the observed repression of expression.

4.3.4 Test for Specificity of the Rabbit Sera

The rabbit sera against full length Ago2 were tested on wildtype as well as lysates from FH-Ago1 to FH-Ago4 overexpressing HEK-293 cells. The lysates were loaded four times, in sets of five, on a SDS-PAGE gel. After blotting, the WB membrane was cut in four pieces and either pre-immune serum or immunized serum was applied (Figure 4.9 A). The serum of rabbit 1 showed signals above background at about 75 and 110 *kDa* in all lanes. Ago migrates at a size of 100 *kDa*on an SDS-PAGE gel and the FH-tag adds another 5 *kDa*. For rabbit 2 samples weaker signals were observed. Only in the lanes loaded with FH-Ago1 and 2 a prominent band at approximately the right position was observed. Next, the antibodies were tested against recombinant Ago2 (Figure 4.9B). Even though there

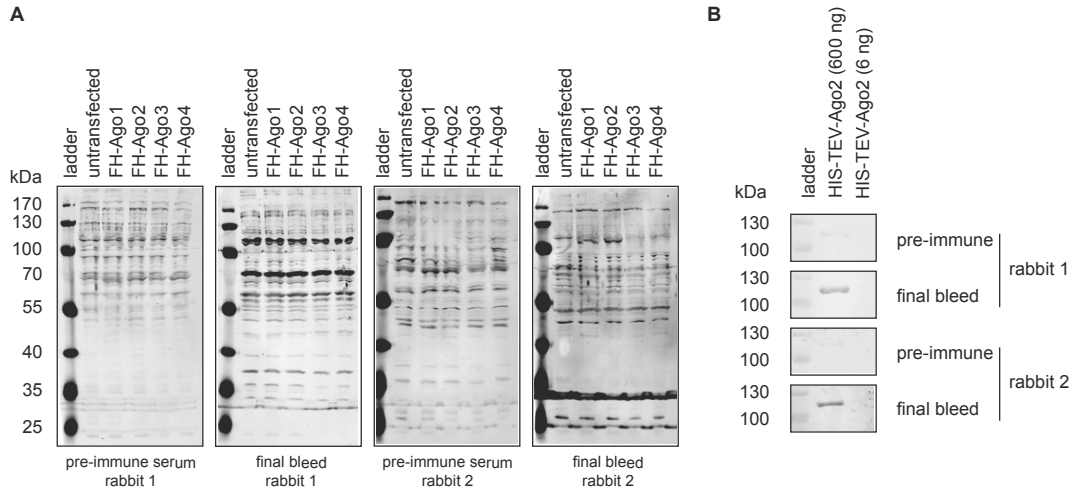


Figure 4.9: ***Test on Ago2 specificity of rabbit serum.***
Rabbit serum (pre-immune and final bleed) of two rabbits immunized with full-length Ago2, was used as primary antibody for WB in a dilution of 1:10. ***A****) The lanes contained HEK-293 cell lysates from untransfected cells and cells transfected with plasmids carrying FH-Ago1 to FH-Ago4. Serum 1 showed a signal for Ago1 through Ago4 and another presumably unspecific band at around 75 kDa on the WB. The serum of rabbit 2 was specific for Ago1 and Ago2.* ***B****) Recombinant HIS-TEV-Ago2 was loaded and a WB was decorated with the sera of rabbit 1 and 2. Here both sera detected Ago2.*

were 600 *ng* of recombinant protein loaded, the WB signals were weak. However, both antibody sera recognized recombinant Ago2.

All in all, these data suggest that the antibody produced by rabbit 1 recognizes Ago1 to 4 and the one from rabbit 2 seems to be more specific for Ago1 and 2. As both antibodies were not entirely specific for Ago2 but for several Ago proteins and the second rabbit serum seemed to give a weaker signal in WB, the serum of rabbit 1 was chosen for further purification.

4.3.5 Affinity Chromatography of Antibodies Using the Target Protein

There are two ways commonly used to purify antibodies by affinity purification from serum. Either the serum is incubated with Protein-A/G, which binds to the Fc part of any IgG antibody, or with the target protein containing the epitopes, which are recognized by the variable domains of the antibody. In order to purify Ago2 antibodies specifically from rabbit serum, HIS-TEV-Ago2 was isolated from insect cells and purified by a Ni-IMAC-column followed by gelfiltration on a Superdex 16/60. The resulting protein (1 *mg*) was covalently coupled to CNBr column material (1*g*) (see Section 3.5.16). Then, Ago2-specific antibodies were purified from the serum 1526, taken form a rabbit, immunized with an Ago2 N-terminal peptide (serum already present in tha lab), and the serum of rabbit 1 from Davids Biotechnology, (see above; for method see Section 3.5.17). The purification process was analysed by SDS-PAGE is shown for the Ago2 antibody from serum 1526 (Figure 4.10 A). Purification of the newly generated antibody led to similar results (data not shown).

The detection limit in western blot of the purified 1526-Ago2-antibody was also tested. Recombinant HIS-TEV-Ago2 protein and HEK-293 cell lysate were loaded on a SDS-PAGE gel and Ago2 was detected by WB using the 1526 antibody (Figure 4.10 B). Here the antibody gave a good signal for recombinant HIS-TEV-Ago2. The signal for endogenous Ago2 was rather low but still detectable. Since the main aim of this project was to characterize potential Ago-chromatin interactions, the Ago2 binding efficiency under ChIP conditions was tested next.

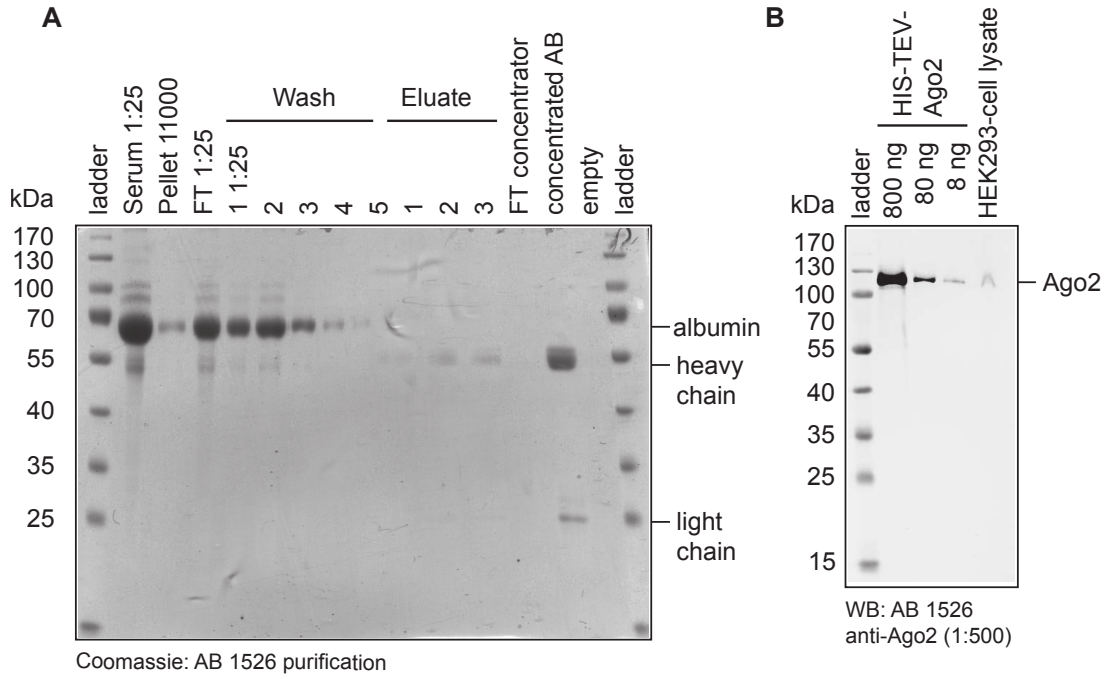


Figure 4.10: ***Antibody purification by Ago2 affinity column.*** ***A****) For antibody (1526) purification, the rabbit serum (10 ml) was cleared by centrifugation and incubated with the HIS-TEV-Ago2-CNBr-column rotating at 4°C for 2 hours. After several washing steps the antibody was eluted with acidc elution buffer and concentrated. The purification was analysed by SDS-PAGE.* ***B****) The detection limit of the antibody (1526) in WB was tested with purified Ago2 protein and HEK293 lysate.*

4.3.6 Efficiency of Immunoprecipitation under ChIP Conditions

After affinity purification of the antibodies against their target protein, the binding efficiency under ChIP conditions was tested (Section 3.6.4) and compared with other antibodies (Figure 4.11). MRC-5 cells used in subsequent DAMID experiments were crosslinked with 1% formaldehyde for 10 minutes at RT. The cells were lysed by sonication for 10 minutes at 4 °C and cellular debris was pelleted by high speed centrifugation. For the IP, the following purified Ago2-specific antibodies were used: 11A9 (5 μg rat mAB), 1526 (5 μg rabbit pAB), rabbit 1 Davids (4 μg rabbit pAB), 2A8 (5 μg mouse mAB) and rabbit IgG (5 μg) as negative control. All antibodies precipitated a band at a size of Ago2, detected by the 11A9 antibody (Figure 4.11). The band from the 11A9-IP migrated somewhat faster. This might be due to the high amount of protein bound by the antibody, leading to overloading the gel.

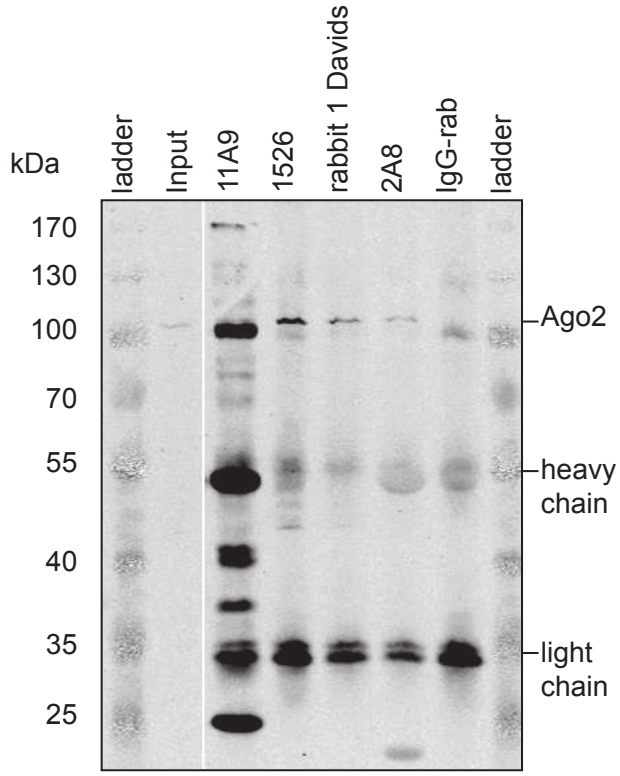


Figure 4.11: ***Comparison of the pull-down efficiency of several Ago2 antibodies under ChIP conditions.***

MRC-5 cells were crosslinked with formaldehyde (1%) and sonicated. The clarified lysate was diluted ten fold in ChIP dilution buffer and was incubated for 2 hours at 4°C in the presence of different Ago2 antibodies. Then Protein G agarose beads were added for another hour at 4°C. After several washing steps, the beads were eluted with SDS-PAGE loading dye. Input and the IPs were loaded on a 10% SDS-PAGE gel and were analysed by WB. Ago2 was detected with the 11A9 antibody. It appears that all antibodies show an Ago2 band. However, there seems to be a slight difference in migration speed of Ago2 for the different lanes. Input, the 1526, the Rabbit 1 Davids and the 2A8 antibody showed a slightly slower running behaviour compared to the Ago2 band recognized by 11A9. The 11A9 antibody yielded also other signals apart from Ago2. A distinct band at 155 kDa, which was described for the antibody and might be the SMRC1 protein, a member of the DNA remodelling complex SWI/SNF, could not be found [Rüdel, 2011]. All the faster migrating bands were not further evaluated in Rüdel [2011].

The 155 *kDa* protein, SMRC1, which is a member of the SWI/SNF complex and was described to be pulled-down in IP's of the 11A9 antibody [Rüdel, 2011], was not observed (Figure 4.11). Unfortunately, both pan-Ago antibodies (rabbit 1 Davids and 2A8) precipitated less Ago2 than the Ago2-specific antibodies 11A9 and 1526.

In conclusion, a set of two Ago2-specific purified antibodies: 11A9 (ProteinA purified, E.Kremmer, Munich, Germany) and 1526 (protein specific purification) as well as a purified pan-Ago-specific antibody: rabbit 1 Davids (proteins specific purification) are available for further ChIP experiments. All of these successfully immunoprecipitated Ago2 under ChIP conditions and will hopefully result in strong signal to noise ratios in subsequent validation experiments of peaks detected by DAMID (see below).

4.3.7 Side Project on Ago1 Loading

Parallel to Ago2, also Ago1 was expressed in insect cells and successfully purified. Having the Ago1 recombinant protein at hand, a question regarding the affinity for single or double stranded siRNA of Ago1 could be addressed. With these experiments I contributed to the publication by Petri *et al.* [2011]. RNA interference (RNAi) is a post transcriptional gene silencing mechanism found in many eukaryotes. Its effector complex is the RNA induced silencing complex (RISC), comprising an Ago protein and a single-stranded RNA, the guide strand. The RISC targets complementary RNAs resulting in cleavage and eventually degradation of the respective RNA. In human cells this process is invoked by transfection of a small double stranded RNA molecule with a length of ~21 nt, called siRNA [Meister and Tuschl, 2004]. Following transfection, siRNAs are loaded into all four human Ago proteins Liu *et al.* [2004]. While Ago2 efficiently removes one of the two RNA strands due to its cleavage activity [Matranga *et al.*, 2005; Rand *et al.*, 2005; Leuschner *et al.*, 2006], this mechanism is absent from Ago1, 3 and 4, leaving them bound to the siRNA duplex [Liu *et al.*, 2004; Meister *et al.*, 2004].

For *Aquifex aeolicus* Ago Yuan *et al.* [2005] tested the binding affinity for different small single and double stranded nucleic acid molecules and found a strongly diminished binding to double stranded RNA as compared to single stranded RNA. Thus, we wanted to test if human cleavage deficient Ago proteins also show a higher affinity to single-stranded RNA as compared to siRNA duplexes. To this end, recombinant Ago1 was incubated with a radio-labelled single-stranded siRNA guide or double stranded siRNA. Ago1 was precipitated using Ago1-specific antibodies and washed several times. The bound RNA was eluted and the associated radioactivity was quantified using a scintillation counter.

Excitingly, Ago1 bound ten fold stronger to the guide strand than to the duplex

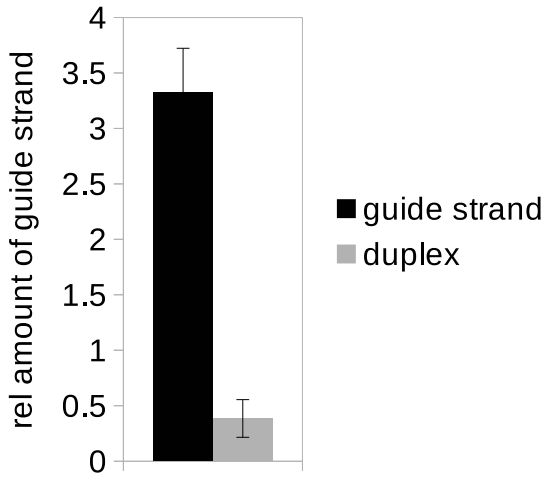


Figure 4.12: ***Analysis of Ago1 affinity for siRNA guide strand or duplex.***
Recombinant Ago1 was purified from insect cells and incubated with radioactively labelled siRNA guide strand or duplex. After precipitation of Ago1 with 4B8 antibody, immobilized on Protein G Sepharose, or empty control beads, the bound proteins were washed and the specifically bound siRNAs were eluted by Proteinase K digest. The radioactive signal of the IP and Input samples was quantified with a scintillation counter. The affinity of Ago1 for the guide strand was about ten fold higher than for the siRNA duplex. The average of three independent experiments is shown. Error bars, SD

RNA (see Figure 4.12). Inside the cell this means that siRNA duplexes dissociate from Ago1 with a high rate, while they are transformed into stable RISC complexes by Ago2. This may result in an increased loading of Ago2 with siRNAs compared to its relative cellular abundance, which can be further amplified by thermodynamically stabilised siRNA duplexes as described in Petri *et al.* [2011].

4.4 Identification of Ago-Chromatin Interaction Sites Applying DAMID

In previous experiments the interaction of Ago with chromatin was evaluated using ChIP. Based on the negative result of the ChIP assay, which is especially sensitive to cytoplasmic Ago contamination and the quality of the utilized antibodies, an antibody-independent technique, called DAMID, was employed. This *in vivo* method relies on the stable expression of a fusion protein consisting of

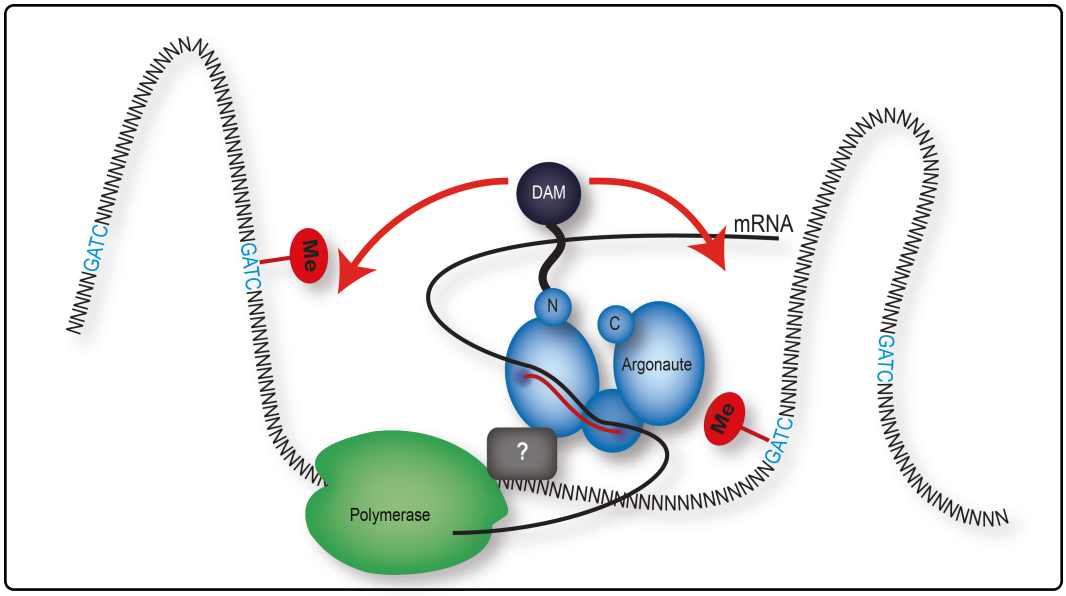


Figure 4.13: ***Model of the DAMID procedure.***
An N-terminal fusion of Dam with Ago is shown. The fusion protein is most likely recruited to the chromatin by complementary base pairing of the bound small RNA with a nascent transcript or by an unknown factor. Then, GATC stretches in the vicinity of the chromatin binding site are methylated by Dam, whereas the GATCs further up- or downstream are not modified. The specific methylation marks, absent from human DNA can be analysed by enzymatic steps of the DAMID protocol.

DNA Adenine Methyltransferase (Dam) from *E. coli* and the target protein (Figure 4.13). Notably, Dam is fully functional as N- and C-terminal fusion [Steensel and Henikoff, 2000]. Ago proteins, however, can only be modified at the N-terminus because the C-terminal carboxyl-group is required for proper folding. During the course of the experiment, a fraction of the fusion protein accumulates in the nucleus and marks native chromatin binding sites by preferential methylation of GATC residues at N6 of the adenine close to putative chromatin binding sites [Greil *et al.*, 2006] (Figure 4.13). Given that adenine methylation is absent from mammals, DAMID produces a unique mark, which is selected for by DpnI digestion (restriction site G^{me}ATC) and selective PCR. Subsequently, the DNA fragments are analysed by a high throughput method. Another critical point is the low expression level of the Dam-fusion protein, otherwise saturation of methylation levels are reached and a quantitative readout is impossible. This low level is achieved by employing a non induced heat shock promoter for control of Dam-fusion protein expression [Vogel *et al.*, 2007]. The low expression level is

an advantage of the technique, granting that the modified protein does not interfere with critical functions of the endogenous protein, which could compromise important cellular functions [Greil *et al.*, 2006]. Even though the target protein is present in low levels, it seems that the Dam-fusion protein can easily accomplish methylation of its target sites [Greil *et al.*, 2006]. This is probably due to the fact, that most chromatin modifying factors have residence times in the range of minutes on a specific locus and therefore quickly change their positions [Phair *et al.*, 2004]. Consequently, a stable expression of the fusion proteins for incubation times above 24 hours should yield a detailed report on all possible binding sites.

4.4.1 Design of the Ago-DAMID Experiment

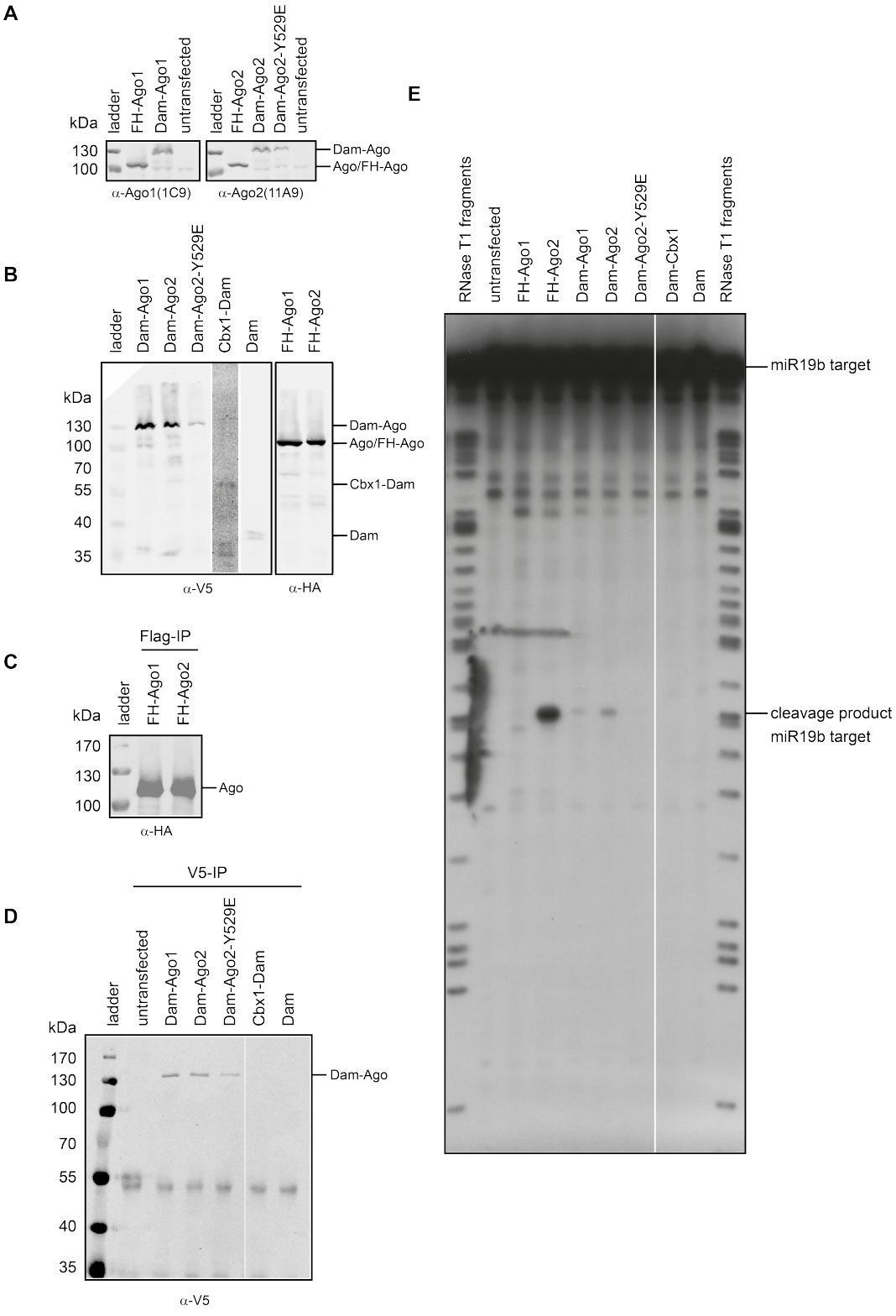
Similar to the ChIP experiment described above, putative chromatin interaction sites of Ago1 and Ago2 should be identified. To distinguish between miRNA or other small RNA-guided Ago1/2 recruitment to chromatin, the Ago2-Y529E mutant discovered by Rüdel *et al.* [2011], was included in the screen. The tyrosine to glutamate substitution in the MID domain at position 529 of this mutant protein creates a negative charge in the 5′-binding pocket of Ago2 and strongly diminishes small RNA binding due to repulsion of the 5′-phosphate. Furthermore, the stacking interactions of the de-localised pi-system of the hydroxyl-phenyl ring of tyrosine with the base at the 5′-position of the miRNA are also destroyed by this mutation. Since this mutant is correctly folded and even functional in cleavage assays with a 5′-methylated RNA, it is a perfect control for miRNA-independent chromatin interaction [Rüdel *et al.*, 2011]. In addition, the Dam protein alone was used to control for the different accessibility states of chromatin [Greil *et al.*, 2006]. Serving as positive control, Cbx1, heterochromatin binding protein HP1-β was included in the assay, for which DAMID has been performed and datasets are available for comparison Vogel *et al.* [2007].

4.4.2 Test on Dam-Ago2 Functionality by Cleavage Assay

In order to use Ago2 in a DAMID experiment, the correct folding and functionality of the Dam-Ago2 fusion protein needs to be asserted. The Ago subfamily of proteins consists of Ago1, Ago2, Ago3 and Ago4 [Carmell *et al.*, 2002]. However, only Ago2 is able to cleave RNA, when loaded with a complementary

small RNA [Meister *et al.*, 2004; Liu *et al.*, 2004; Hauptmann *et al.*, 2013]. Therefore, the cleavage assay used in the aforementioned publications, can be applied to test the catalytic activity of the Dam-Ago2 fusion protein. After cloning of the target proteins into lentiviral transfer vectors, the expression (Ago1, Ago2, Ago2-Y529E) and cleavage competence (Ago2) of the N-terminally tagged proteins were tested (Figure 4.14). The plasmids were transfected into HEK-293 cells (see Section 3.3.6). The expression of the fusion constructs was tested by SDS-PAGE and WB (Figure 4.14, A, B). Here, the size shift of the Dam-Ago fusion proteins compared to the Flag-HA-tagged (FH-)Ago used as control and endogenous Ago proteins is nicely shown. Then, the IP efficiencies were tested for the different constructs, were FH-tagged proteins were highly enriched compared to a rather intermediate level for Dam-Ago fusion proteins and low signals for the Dam and Cbx1 controls (Figure 4.14, C, D). This difference in signal intensity after IP is probably due to the lower expression level of the Dam-fusion proteins compared to the FH-tagged ones (Figure 4.14, A, B). Subsequently, the cell lysates were used for a cleavage assay (compare Section 3.6.9) [Meister *et al.*, 2004]. Here, the transfected Dam fusion proteins were immunoprecipitated from the cell lysate and incubated with an artificial target RNA. This target RNA is fully complementary to miR-19b, which is highly abundant in HEK-293 cells. For *in vitro* cleavage reactions, 5′-cap-labelled substrate RNA was incubated with different fusion proteins. The reaction was stopped by Proteinase K digest and the isolated RNA was size separated on a sequencing gel and visualised with auto-radiography (Figure 4.14, E). FH-Ago1 and FH-Ago2 were included as controls in the cleavage assay [Hauptmann *et al.*, 2013].

Figure 4.14 ***(on the next page): Test on expression and cleavage competence of Dam-Ago.*** *HEK-293 cells were transfected with Dam-fusion constructs and several controls. The expression of the fusion proteins was analysed by WB:* ***A****) using Ago-specific antibodies and* ***B****) V5- and HA-tag antibodies.* ***C, D****) The immunoprecipitation efficiencies of the different tagged fusion proteins were analysed.* ***E****) For the cleavage assay V5-tagged Dam-fusions and FH-tagged Ago1 and Ago2 were immunoprecipitated. The proteins were incubated with radioactively labelled RNA substrate and the cleaved fragments were size separated on a sequencing gel followed by auto-radiography. Flag-HA-Ago2 and Dam-V5-Ago2 were able to cleave the target-RNA, as expected. FH-Ago1 and the Dam fusion proteins of Ago2-Y529E, Dam-Cbx1 and Dam showed only background activity.*


A
ladder
FH-Ago1
Dam-Ago1
untransfected
ladder
FH-Ago2
Dam-Ago2
Dam-Ago2-Y529E
untransfected
kDa
130
100
Dam-Ago
Ago/FH-Ago
α-Ago1(1C9)
α-Ago2(11A9)
B
ladder
Dam-Ago1
Dam-Ago2
Dam-Ago2-Y529E
Cbx1-Dam
Dam
FH-Ago1
FH-Ago2
kDa
130
100
70
55
40
35
Dam-Ago
Ago/FH-Ago
Cbx1-Dam
Dam
α-V5
α-HA
C
Flag-IP
ladder
FH-Ago1
FH-Ago2
kDa
170
130
100
Ago
α-HA
D
V5-IP
ladder
untransfected
Dam-Ago1
Dam-Ago2
Dam-Ago2-Y529E
Cbx1-Dam
Dam
kDa
170
130
100
70
55
40
35
Dam-Ago
α-V5
E
RNase T1 fragments
untransfected
FH-Ago1
FH-Ago2
Dam-Ago1
Dam-Ago2
Dam-Ago2-Y529E
Dam-Cbx1
Dam
RNase T1 fragments
miR19b target
cleavage product
miR19b target

In summary, both Ago2 constructs were able to cleave the target RNA, complementary to miR-19b. The intensity difference is explained by the much stronger expression level of FH-Ago2 compared to a moderate to low expression of Dam-Ago2, which results in a higher cleavage rate for FH-Ago2. All other constructs were correctly expressed and, as expected, inactive in the cleavage assay. Thus, the Dam-Ago2 constructs are correctly folded and a meaningful DAMID experiment can be performed.

4.4.3 DAMID Experiment with MRC-5 cells

As Ago1 and Ago2 were both implied to have a nuclear function for example [Robb *et al.*, 2005; Janowski *et al.*, 2006], this project focused on these two members of the Ago subfamily. To elucidate their endogenous chromatin interaction sites, MRC-5 cells, primary human lung fibroblasts, were transduced with lentiviruses encoding for Dam-Ago1, Dam-Ago2, Dam-Ago2-Y529E, Dam-Cbx1 and Dam. After 72 hours of incubation time, the experiment was stopped and genomic DNA was isolated, following RNaseA treatment.
The viral titre was estimated using a PCR-based protocol (Section 3.6.8) and confirmed that the cells were infected with a multiplicity of infection (MOI)$<$1. This low MOI, which is less than one infectious virus per cell, reduces the risk of several Dam-fusion protein insertions in the genome of the cells and thus ensures the low expression level required for DAMID (Figure 4.15a). Subsequently, the DNA fragments, methylated by the Dam moiety, were enriched by enzymatic digest and GATC selective PCR (Section 3.6.5). A fraction of the amplified fragments was analysed on an agarose gel (2%). The different Dam-Ago proteins produced a strong band pattern overlaying a weak background smear. This suggests amplification of only a few selected target sequences or sequences of a certain length. In contrast, the controls, Dam and Dam-Cbx1, produced a smear ranging from about 100 bp to 3 kb, implying a vast amount of different amplified sequences.
The specificity of the DAMID procedure was additionally controlled by samples lacking DpnI or T4-DNA-Ligase. DpnI digestion specifically recognizes methylated GATC residues, while the T4-DNA-Ligase is required for fusion of a specific PCR adapter. Therefore, both samples should not contain PCR-amplifiable fragments, which is demonstrated in Figure 4.15a. As the range of the enriched DNA fragments was between 100 bp and 3kb and thus too long for applying NGS directly, the DNA was sonicated to a mean fragment length of 250 bp, which is

required for Illumina NGS. Then paired-end NGS-libraries were prepared from these shorter DNA pieces using the Illumina NGS library preparation kit. This treatment rendered the low complexity samples of the Dam-Ago experiments into higher complexity libraries, regarding the length of the sequences and to a minor part, the difference in start and/or end points. The overall sequence content, of course, was restricted to the few selected input sequences visible on the agarose gel for the Dam-Ago treated cells.

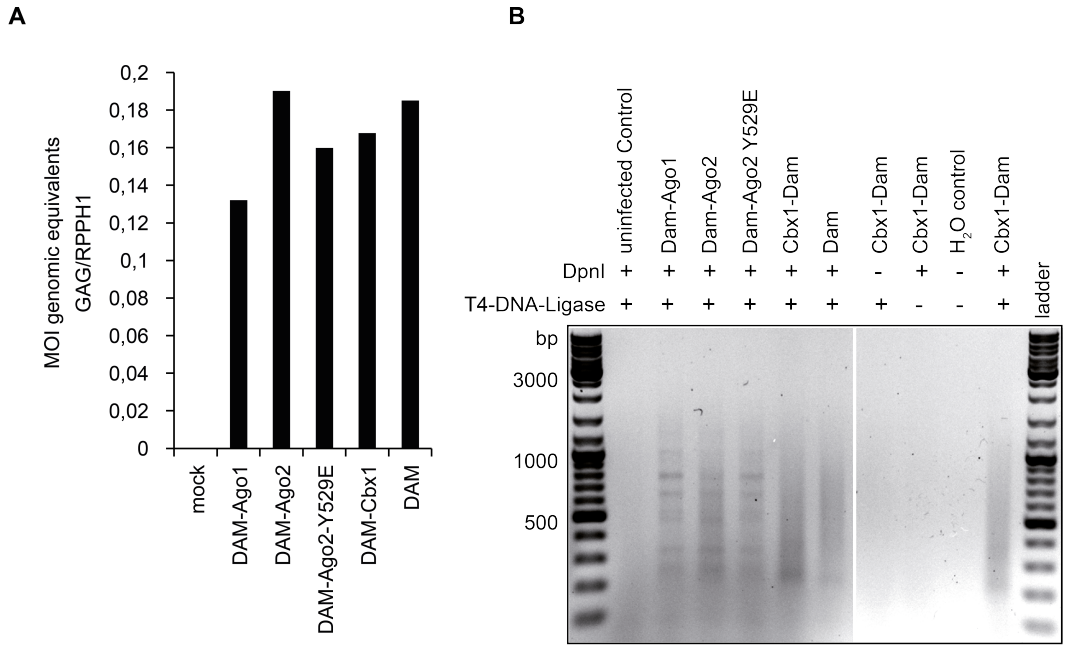


Figure 4.15: ***Preparation of methylated DNA fragments from DAMID.***
MRC-5 cells were transduced with lentiviruses encoding for Dam-fusion proteins. Following an incubation for 72 hours, the DNA was extracted from the cells. ***A****) The viral titre was estimated by qPCR, comparing the genome-integrated, viral DNA to an endogenous genomic locus. Here a MOI of 0.2 was found.* ***B****) The methylated DNA fragments, marking Ago chromatin interaction sites, were selected, enriched by PCR and analysed on an agarose gel (2%). Dam-Ago fusion proteins produced a weak background smear with several prominent bands, which was rather unexpected. The control proteins Dam and Dam-Cbx1, showed a homogeneous smear, indicating a broad range of different bound loci. DpnI and T4-DNA-Ligase controls confirmed the specificity of the method.*

The Illumina libraries were sent to the Kompetenz Zentrum für Fluoreszente Bioanalytik (KFB, Regensburg, Germany) for high-throughput sequencing. Due to technical problems at the KFB the libraries were finally sequenced by an Illumina Sequencing Facility (Keith Moon, Illumina Union City, CA, USA) on a HiSeq2000.

4.4.3.1 Analysis of the Band Pattern in Ago DAMID Experiments

As the band pattern observed for Dam-Ago treated cells was unexpected and implied a specific amplification of only few selected genomic loci, a deeper analysis of these bands was performed. Therefore, the DNA extracted from Dam-Ago treated cells was subjected to the DAMID protocol and loaded on an agarose gel. Some of the prominent bands observed for cells expressing Ago fusion proteins but not for control proteins were isolated from the gel. After gel extraction and purification of the DNA fragments, these were cloned into pGEM-T-easy vectors. Four different DNA bands, starting from the strongest, were selected for Dam-Ago1 and Dam-Ago2 and for each band eight bacterial clones were analysed by Sanger sequencing. Subsequently, the sequencing reads were mapped to the human genome using the blast algorithm [Altschul *et al.*, 1990]. The results are summarized in Table 4.5. Unfortunately, only few of the sequencing runs for Dam-Ago2 clones were successful due to technical problems at GATC Biotech AG (Konstanz, Germany).
There were several mitochondrial DNA sequences found in bacterial clones from Dam-Ago1 and Dam-Ago2 bands (Table 4.5). Furthermore, sequences from various chromosomes were identified, which did not converge to one gene locus or type of repetitive element. In order to get a better understanding of the sequence fragments found here, they could be compared with the NGS data of the Dam-Ago1 and Dam-Ago2 libraries. The high number of mitochondrial sequences cloned especially from Dam-Ago1 bands, suggests that the observed band pattern in Dam-Ago Samples of the DAMID experiment may derive from mitochondrial sequences.

Ago1 bands	comment	length seq frag	chrom	start	end	length aligned
Ago1-1_1	match	1043	chrM	1917	2917	1000
Ago1-1_2	match	1128	chrM	1940	2940	1000
Ago1-1_3	match	999	chrM	1966	2966	1000
Ago1-1_4	match	1009	chrM	10990	11800	810
Ago1-1_5	no sequence					
Ago1-1_6	match	924	chrM	11074	11870	796
Ago1-1_7	match	1169	chr2	154726508	154727508	1000
Ago1-1_8	match	1029	chrM	10965	11900	935
Ago1-2_1	match	571	chr1	50204840	50205365	525
Ago1-2_2	match	1139	chr6	56688282	56689282	1000
Ago1-2_3	no sequence					
Ago1-2_4	match	1025	chrM	12418	13318	900
Ago1-2_5	match	1041	chrM	12402	13302	900
Ago1-2_6	match	1056	chr19	867534	868434	900
Ago1-2_7	match	800	chrM	12279	13179	900
Ago1-2_8	no sequence					
Ago1-3_1	match	1124	chr19	21775122	21776022	900
Ago1-3_2	no sequence					
Ago1-3_3	no sequence					
Ago1-3_4	no sequence					
Ago1-3_5	match	1120	chr21	37578897	37579771	874
Ago1-3_6	match	640	chr3	19988849	19989349	500
Ago1-3_7	no sequence					
Ago1-3_8	match	1133	chr20	61070495	61071495	1000
Ago1-4_1	match	1097	chr1	167794091	167794851	760
Ago1-4_2	match	262	chrM	552	754	202
Ago1-4_3	no sequence					
Ago1-4_4	match	801	chr17	29257881	29258618	737
Ago1-4_5	match	823	chr1	567455	568212	757
Ago1-4_6	match	769	chr1	180126402	180127106	704
Ago1-4_7	match	849	chr1	568209	568988	779
Ago1-4_8	match	724	chr19	17598466	17599066	600
Ago2 bands	comment	length seq frag	chrom	start	end	length aligned
Ago2-1_1	no match	1204				
Ago2-1_2	no sequence					
Ago2-1_3	no match	1148				
Ago2-1_4	no match	1178				
Ago2-1_5	match	1162	chrM	1211	2347	1136
Ago2-1_6	no sequence					
Ago2-1_7	no sequence					
Ago2-1_8	no match	1076				
Ago2-2_1	no sequence					
Ago2-3_1	no sequence					
Ago2-3_2	match	1007	chr3	93636845	93637835	990
Ago2-3_3	match	1117	chr3	93636846	93637949	1103
Ago2-3_4	match	1104	chr16	89343144	89343370	226
Ago2-3_5	match	1067	chr3	93636842	93637816	974
Ago2-3_6	match	1152	chr21	46723557	46724574	1017
Ago2-3_7	match	1051	chr3	93636849	93637835	986
Ago2-3_8	no sequence					
Ago2-4_1	match	847	chr16	88620267	88621049	782
Ago2-4_2	match	725	chrM	7886	8599	713
Ago2-4_3	match	1169	chr3	170108506	170109348	842
Ago2-4_4	no sequence					0
Ago2-4_5	no match	42				
Ago2-4_6	match	1099	chr9	99678548	99679240	692
Ago2-4_7	match	1125	chr16	4512923	4513722	799
Ago2-4_8	match	1116	chr16	83554430	83555221	791

Table 4.5: ***Sanger analysis of Ago1 and 2 band pattern on Agarose Gels.*** *DNA resulting from the DAMID procedure was loaded on an agarose gel. The four most prominent bands of the Dam-Ago1 and Dam-Ago2 samples were extracted from the gel, cloned and sequenced. The resulting sequences were mapped to the human genome using the blast algorithm [Altschul et al., 1990].*

4.4.3.2 Bioinformatic Analysis of DAMID in MRC-5

The Dam-Ago libraries were generated in paired-end mode and sequenced on an Illumina HiSeq2000 with a read length of 100 *bp* for forward and reverse read. The samples were indexed with different primers during library preparation and could therefore be mixed and sequenced in parallel on several lanes of a flowcell.

Sample name	Library code	Index	Reads
Dam-Ago1	1	CGATGT	18516843
Dam-Ago2	2	TGACCA	23181289
Dam-Ago2-Y529E	3	ACAGTG	17882388
Dam-Cbx1	4	GCCAAT	20021720
Dam	5	CAGATC	19134936

Table 4.6: ***NGS-DAMID run information for MRC-5 Cells.***
The read numbers received were comparable for all libraries, setting a good basis for further analysis.

The received reads were of good quality and general information on the samples is shown in Table 4.6. For each sample at least 17 Mio reads were generated forming a good basis for subsequent analysis.

4.4.3.3 Data Analysis Using the Software Homer

The data from the NGS run was forwarded to the Institute for Functional Genomics at University of Regensburg (Engelmann and Wang) for bioinformatic data analysis. The reads were aligned to the human reference genome using the alignment software Bowtie [Langmead *et al.*, 2009]. The overall alignment rate was good for all samples (Table 4.7).

Sample Name	Mapping Rate
Dam-Ago1	98.75%
Dam-Ago2	98.54%
Dam-Ago2-Y529E	98.87%
Dam-Cbx1	97.39%
Dam	98.67%

Table 4.7: ***Mapping rate of MRC-5 DAMID samples against the human reference genome.***
Using the read alignment software Bowtie the data from the NGS analysis was mapped to the human genome with good alignment rates for all samples.

Sample Name	Total	Total size:1kB
Dam-Ago1	48864	6289
Dam-Ago2	42882	866
Dam-Ago2-Y529E	73217	8828
Dam-Cbx1	75569	1723
Dam	NA	NA

Table 4.8: ***Peak calling with the software package homer.*** *The aligned reads were analysed with the ChIP-Seq analysis package homer. The control sample Dam was used as input control and enrichment was assessed for the different samples. A huge number of peaks was identified for all samples. As a filter only peaks with a size around 1 kB were taken forward.*

Applying the software homer, which is designed for ChIP analysis, peaks were called using the reads from the DAMID control experiment Dam as input sample (Salk Institute: http://homer.salk.edu/homer/chipseq/). In theory, the Dam protein should be able to randomly GATC-methylate all accessible DNA and thus create fragments from these loci across all cells used for the experiment. As the Dam-fusion proteins are bigger and sterically hindered compared to Dam, the non-specific methylation rate should be reduced and most of the methylation events should occur at the *in vivo* targeting site in the treated cells. This causes an increased methylation ratio of targeted regions for Dam-fusion proteins compared to Dam integrating over the respective cell population. The results of the peak calling are summarized in Table 4.8. Between 42000 and 75000 potential peaks were identified by homer. As a first filter only peaks of a length around 1 *kB* were taken forward for validation.
Homer leaves the user with two scores, the peak score and the focus ratio (see Table 4.9, 4.10). Both values were used to sort the data and top peaks were analysed manually using the UCSC genome browser in combination with .bedGraph files depicting the read distribution and accumulation at a peak site. Here, the focus ratio showed to be the better value to identify peaks for subsequent validation. The top 40 peaks defined by homer sorted according to their focus ratio are shown for Dam-Ago1 and Dam-Ago2 in the Tables 4.9, 4.10 (*see next pages*).

Table 4.9: ***Dam-Ago1 enriched sites in MRC-5 based defined by homer analysis.***

PeakID	Chr	Start	End	Peak Score	Focus Ratio/Region Size	Detailed Annotation	Distance to TSS	Gene Name	Gene Description
chr13-1	chr13	110076091	110077091	1917.7	0.998	Intergenic	362323	IRS2	insulin receptor substrate 2
chr2-7	chr2	227586531	227587531	564.5	0.998	Intergenic	76475	IRS1	insulin receptor substrate 1
chr2-396	chr2	81893223	81894223	100.6	0.998	Intergenic	-1190204	LOC1720	dihydrofolate reductase pseudogene
chr3-32	chr3	43270543	43271543	151	0.997	Intergenic	-56961	SNRK	SNF related kinase
chrX-97	chrX	110659857	110660857	87.5	0.997	Intergenic	-4897	DCX	doublecortin
chr2-14	chr2	50815831	50816831	217.7	0.995	intron	-241437	NRXN1	neurexin 1
chr10-1675	chr10	37890956	37891956	50.3	0.994	5' UTR	403	MTRNR2L7	MT-RNR2-like 7
chrY-148	chrY	4212392	4213392	35	0.994	Intergenic	-655375	PCDH11Y	protocadherin 11 Y-linked
chr5-4705	chr5	94543067	94544067	30.6	0.993	intron	76712	MCTP1	multiple C2 domains, transmembrane 1
chr8-2	chr8	68494085	68495085	536	0.992	intron	164035	CPA6	carboxypeptidase A6
chr1-4118	chr1	221966792	221967792	43.8	0.992	CT-rich|Low_compl	-51776	DUSP10	dual specificity phosphatase 10
chr8-2439	chr8	111945105	111946105	40.5	0.992	Intergenic	-958646	KCNV1	potassium channel, subfamily V, member 1
chr1-6126	chr1	77436570	77437570	33.9	0.992	intron	103884	ST6GALNAC5	ST6 sialyltransferase
chrX-1	chrX	125606067	125607067	1060.1	0.991	Intergenic	80275	DCAF12L1	DDB1 and CUL4 associated factor 12-like 1
chr2-2	chr2	49456387	49457387	1026.1	0.986	L1PA6|LINE|L1	-75221	FSHR	follicle stimulating hormone receptor
chr5-1263	chr5	5396482	5397482	62.4	0.986	Intergenic	-25825	KIAA0947	KIAA0947
chr6-3545	chr6	99897761	99898761	38.3	0.984	L1PA5|LINE|L1	-25054	SFRS18	splicing factor, arginine/serine-rich 18
chr5-158	chr5	105888655	105889655	113.8	0.979	Intergenic	1117441	EFNA5	ephrin-A5
chr8-1833	chr8	32870843	32871843	47	0.976	Intergenic	291992	NRG1	neuregulin 1
chr18-38	chr18	2841795	2842795	124.7	0.967	Intergenic	-4733	EMILIN2	elastin microfibril interfacer 2
chr8-749	chr8	98377694	98378694	68.9	0.966	L1M4c|LINE|L1	-88018	TSPYL5	TSPY-like 5
chr5-1623	chr5	122517735	122518735	56.9	0.959	intron	93394	PRDM6	PR domain containing 6
chr3-1572	chr3	162283992	162284992	59.1	0.958	Intergenic	736597	LOC647107	hypothetical LOC647107
chr2-3	chr2	132140525	132141525	838	0.954	Intergenic	-19294	WTH3DI	RAB6C-like
chr6-4009	chr6	25318077	25319077	35	0.946	intron	38921	LRRC16A	leucine rich repeat containing 16A
chr14-494	chr14	61794054	61795054	68.9	0.942	intron	6039	PRKCH	protein kinase C, eta
chr6-1167	chr6	37579956	37580956	64.5	0.939	MLT1L|LTR	85310	MDGA1	MAM domain containing glycosylphosphatidylinositol anchor 1
chr22-109	chr22	33290638	33291638	84.2	0.938	intron	94336	TIMP3	TIMP metallopeptidase inhibitor 3
chr3-1809	chr3	149489809	149490809	55.8	0.938	L2a|LINE|L2	-20023	COMMD2	COMM domain containing 2
chr6-965	chr6	6300853	6301853	68.9	0.936	intron	19571	F13A1	coagulation factor XIII, A1 polypeptide
chr11-1473	chr11	34911091	34912091	52.5	0.936	intron	-26086	PDHX	pyruvate dehydrogenase complex, component X
chr4-5	chr4	65473450	65474450	549.2	0.933	Intergenic	-198772	TECRL	trans-2,3-enoyl-CoA reductase-like
chr1-1984	chr1	244953427	244954427	61.3	0.933	L1M4|LINE|L1	-44712	FAM36A	family with sequence similarity 36, member A
chr17-571	chr17	39332427	39333427	70	0.932	TTS	1533	KRTAP4-2	keratin associated protein 4-2
chr6-1311	chr6	131605285	131606285	62.4	0.928	Intergenic	34486	AKAP7	A kinase (PRKA) anchor protein 7
chr14-610	chr14	92260307	92261307	64.5	0.924	Tigger1|DNA	42042	TC2N	tandem C2 domains, nuclear
chr4-1112	chr4	181033791	181034791	63.5	0.924	Intergenic	1046011	NCRNA00290	non-protein coding RNA 290
chr2-1932	chr2	194682789	194683789	61.3	0.923	Intergenic	1068718	PCGEM1	prostate-specific transcript 1 (non-protein coding)
chr1-1600	chr1	174663741	174664741	66.7	0.922	L1PA3|LINE|L1	-180443	RABGAP1L	RAB GTPase activating protein 1-like
chr4-1102	chr4	138087516	138088516	63.5	0.922	L1MC4|LINE|L1	365613	PCDH18	protocadherin 18

Table 4.10: ***Dam-Ago2 enriched sites in MRC-5 called by the homer package.***

PeakID	Chr	Start	End	Peak Score	Focus Ratio/Region Size	Detailed Annotation	Distance to TSS	Gene Name	Gene Description
chr13-1	chr13	110076099	110077099	1438.8	0.997	Intergenic	362315	IRS2	insulin receptor substrate 2
chr2-6	chr2	227586536	227587536	475.8	0.997	Intergenic	76470	IRS1	insulin receptor substrate 1
chrX-33	chrX	110659857	110660857	51.8	0.997	Intergenic	-4897	DCX	doublecortin
chrX-1	chrX	125606119	125607119	722.9	0.995	Intergenic	80223	DCAF12L1	DDB1 and CUL4 associated factor 12-like 1
chr2-665	chr2	81893219	81894219	47.6	0.992	Intergenic	-1190208	-	
chr3-97	chr3	43270538	43271538	65.3	0.99	Intergenic	-56966	SNRK	SNF related kinase
chr2-18	chr2	50815834	50816834	144.9	0.986	intron	-241440	NRXN1	neurexin 1
chr5-179	chr5	105888658	105889658	56.8	0.984	Intergenic	1117438	EFNA5	ephrin-A5
chr8-2	chr8	68494067	68495067	375	0.969	intron	164053	CPA6	carboxypeptidase A6
chr2-2	chr2	49456415	49457415	651.2	0.96	L1PA6 | LINE | L1	-75249	FSHR	follicle stimulating hormone receptor
chr2-4	chr2	132140540	132141540	541.9	0.96	Intergenic	-19309	-	
chr22-106	chr22	33290641	33291641	50.4	0.956	intron	94339	TIMP3	TIMP metallopeptidase inhibitor 3
chr9-108	chr9	88163007	88164007	58.9	0.942	intron	193437	AGTPBP1	ATP/GTP binding protein 1
chr18-10	chr18	2841794	2842794	85.9	0.933	Intergenic	-4734	EMILIN2	elastin microfibril interfacer 2
chr11-1	chr11	10531028	10532028	958	0.926	Promoter-TSS	-805	MTRNR2L8	MT-RNR2-like 8
chr18-47	chr18	13408375	13409375	61.1	0.921	CT-rich | Low_compl	190089	C18orf1	chromosome 18 open reading frame 1
chr5-285	chr5	22352176	22353176	51.8	0.919	intron	210215	-	
chr14-95	chr14	71583995	71584995	57.5	0.918	L1MEg | LINE | L1	210373	PCNX	pecanex homolog (Drosophila)
chr7-381	chr7	130383990	130384990	47.6	0.914	Intergenic	-13084	TSGA13	testis specific, 13
chr12-229	chr12	62047960	62048960	52.6	0.908	LTR49-int | LTR	538091	FAM19A2	family with sequence similarity 19 (chemokine (C-C motif)-like)
chr7-237	chr7	102647413	102648413	53.3	0.906	L1MEc | LINE | L1	67102	FBXL13	F-box and leucine-rich repeat protein 13
chr4-318	chr4	122506681	122507681	48.3	0.904	Intergenic	110966	ANXA5	annexin A5
chr4-285	chr4	148944895	148945895	49.7	0.903	intron	291942	ARHGAP10	Rho GTPase activating protein 10
chr5-4	chr5	93904882	93905882	316.7	0.902	intron	48927	KIAA0825	KIAA0825
chr8-185	chr8	19026901	19027901	54	0.902	MER65A | LTR	-143680	SH2D4A	SH2 domain containing 4A
chr18-157	chr18	51894000	51895000	48.3	0.902	L1M2 | LINE | L1	9329	C18orf54	chromosome 18 open reading frame 54
chr16-217	chr16	85896585	85897585	51.8	0.899	AluSx | SINE | Alu	-35689	IRF8	interferon regulatory factor 8
chr10-290	chr10	116570467	116571467	52.6	0.895	Intergenic	-10536	FAM160B1	family with sequence similarity 160, member B1
chr12-319	chr12	115001744	115002744	49.7	0.895	MLT1B | LTR	119725	TBX3	T-box 3
chr1-436	chr1	172768354	172769354	54	0.894	HERV9-int | LTR	140669	FASLG	Fas ligand (TNF superfamily, member 6)
chr10-641	chr10	93226069	93227069	43.3	0.892	L1MC2 | LINE | L1	56473	HECTD2	HECT domain containing 2
chr11-338	chr11	95442201	95443201	49.7	0.891	L2a | LINE | L2	80253	FAM76B	family with sequence similarity 76, member B
chr3-1	chr3	96336302	96337302	303.2	0.89	Intergenic	-196623	EPHA6	EPH receptor A6
chr17-159	chr17	3478627	3479627	56.1	0.89	L1MC4 | LINE | L1	17044	TRPV1	transient receptor potential cation channel, subfamily V
chr2-8	chr2	88124092	88125092	313.2	0.889	intron	694	RGPD2	RANBP2-like and GRIP domain containing 2
chr18-2	chr18	45379231	45380231	180.4	0.888	intron	77195	SMAD2	SMAD family member 2
chr4-74	chr4	142336110	142337110	66.8	0.887	MLT1H | LTR	194561	ZNF330	zinc finger protein 330
chr22-178	chr22	28146967	28147967	45.5	0.886	L2c | LINE | L2	50019	MN1	meningioma (disrupted in balanced translocation) 1
chr16-1252	chr16	46641401	46642401	31.2	0.884	intron	13410	SHCBP1	SHC SH2-domain binding protein 1
chr1-470	chr1	226036888	226037888	53.3	0.883	intron	24386	EPHX1	epoxide hydrolase 1, microsomal (xenobiotic)

Subsequently, the peaks were classified in genomic elements and the frequencies were compared to the human genome (build hg19) (see Figure 4.16). The peaks from the Dam-Ago1 and 2 transduced cells showed enrichment for the class of LINE elements and intronic regions. A comparison of Dam-Ago2 with Dam-Ago2-Y529E hints to a potentially small RNA-dependent recruitment to SINE elements, which are, however, not enriched compared to the frequency of the element in the human genome (hg19). The enrichment of the class of LINE elements is seen for Dam-Ago2 and Dam-Ago2-Y529E suggesting either non-specific or small RNA-independent association of Dam-Ago2 with the respective chromatin (Figure 4.16). As mentioned before, we observe a strong enrichment of Dam-Cbx1 for SINE elements compared to the frequency in the human genome (hg19).

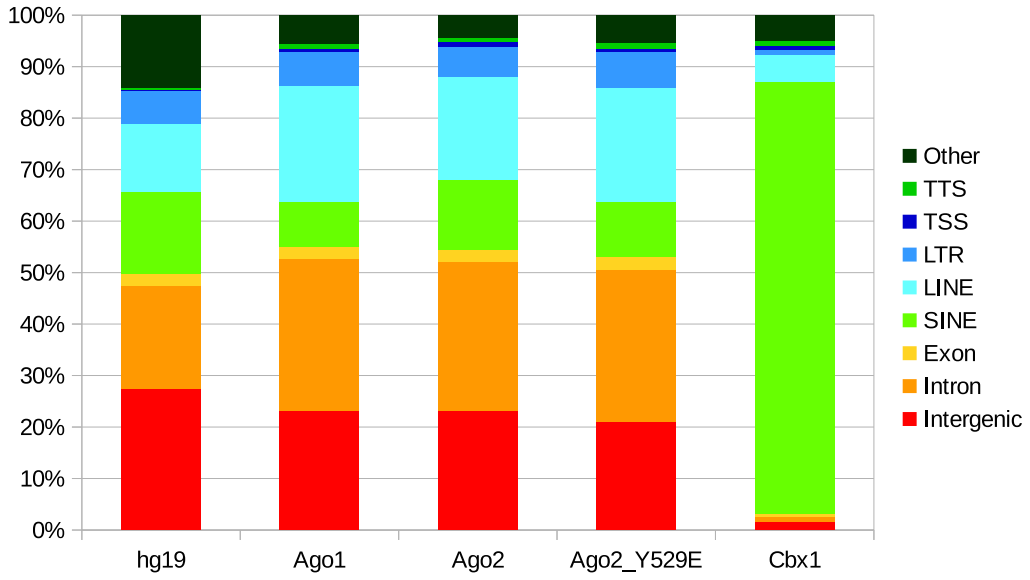


Figure 4.16: ***Classification of detected peaks into genomic elements.***
Comparison of the frequencies of peaks in certain genomic classes of Dam-Ago1 and Dam-Ago2 with the human genome revealed a potential enrichment for the class of LINE elements and introns. Dam-Cbx1 was highly associated with SINE elements.

The positive control Dam-Cbx1 was included in our analysis and yielded 1723 peaks. To assess the quality of our experiment we compared our data with the original publication [Vogel *et al.*, 2006]. Ten of our peaks overlapped with exonic regions of genes, seven of these were also found enriched in the Dam-Cbx1 of MCF7 cells [Vogel *et al.*, 2006]. When a broader set of peaks was examined, which also containing gene introns, 66 of 79 peaks were found enriched in the

reference data set [Vogel *et al.*, 2006]. Given that the DAMID-Seq experiment detected Cbx1 association sites genome wide whereas the study of Vogel *et al.* [2006] analysed only genes and their respective transcripts, the total overlap is reduced. In addition, we find strong enrichment of Cbx1 at repetitive elements, which are not properly captured by the experiments of Vogel *et al.* [2006] and therefore further diminish the number of peaks spanning protein coding genes (compare Figure 4.16). In summary, these findings suggest that we successfully established the DAMID technique.

4.4.3.4 Validation of Homer Peaks for Ago2

For validation only peaks derived from Dam-Ago2 were analysed. The peaks were manually inspected using the UCSC genome browser in combination with .bedGraph files depicting the read distribution and accumulation at a peak site. Based on this, two to three peaks form different genomic elements as for example exon, intron, SINEs and LINEs were chosen, which scored high in the bioinformatic analysis and showed no accumulation of reads in the Dam-Ago2-Y529E sample. Subsequently, a total of 24 primer sets specific for these peaks were designed for Ago2-ChIP (11A9 antibody) validation experiments followed by qPCR analysis (Figure 4.17). For a subset of peaks also control primers were designed in the vicinity of the peak, located 2000 or 5000 *bp* upstream of the respective peak.

The promoter region and transcription start site was tested using the primer sets chr10-248 and chr8-271. Here, no enrichment compared to the bead control was observed. Exons were analysed with the primer sets chr1-628, chr3-47 and chr3-47-2000 (upstream control). Both showed a signal slightly higher than their respective bead control. However, the non targeted locus chr3-47-2000 was for Ago2 and empty beads relatively more enriched compared to its input than chr1-628 and showed a similar enrichment for Ago2 over beads as several target peaks. The peak chr3-47 was only 1.5 fold enriched over its unspecific control locus and the empty bead control. Introns were probed with the primer sets chr20-189 and chr20-189-5000 (upstream control) and chr6-557 (not shown). Here again, only 1.5 fold enrichment compared to empty beads were observed. The $3'$-UTR comprising peaks were examined with the primer sets chr16-73, chr19-32 and chr2-244, which showed no enrichment compared to the control.

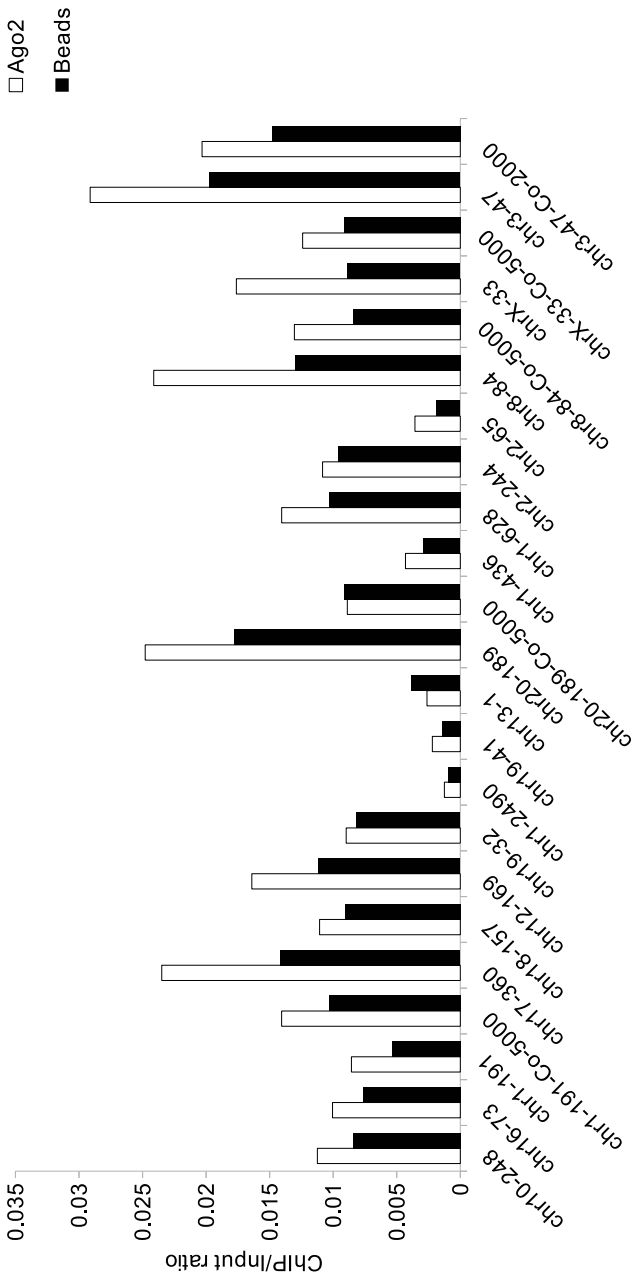


Figure 4.17: ***ChIP experiment of several potential target loci.*** *FA-Crosslinked chromatin was prepared from MRC-5 cells and incubated with Ago2-antibody (11A9) or empty beads as control. The precipitated DNA was analysed by qPCR with respective primer sets. In order to control for specificity of a target locus, another primer set was designed for some sites binding several kb upstream of the target site. However, the enrichment observed was often less than twofold which might either hint to a rather indirect interaction of Ago2 with the chromatin, loss of signal in the background noise level or a lack of identification of valid peaks.*

Also repeat associated regions and transpoable elements were tested. SINE elements were analysed with the primer sets chr1-191, chr1-191-5000 (upstream control), chr1-2490, chr2-82 (not shown) and chr2-8 (not shown). Here, no enrichment was found for the target regions, unexpectedly, the control region chr1-191-5000 yielded higher signal than the respective target site. LINE elements were assessed with the primer sets chr17-360, chr18-157, chr19-41. Here 1.5 fold enrichment over the bead control was found for the peaks chr17-360 and chr19-41, while chr18-157 did not show positive association. Long terminal repeats and other repeat elements were analysed with the primer sets chr1-436, chr2-65, chr18-47 (CT-rich, not shown), chr3-119 (hTA-Charlie, not shown). All of these peaks displayed 1.5 to 2 fold enrichment but gave lower ChIP/input ratios than upstream controls from other loci.
At last, also intergenic regions were tested with the primer sets chr12-169, chr13-1, chr8-84, chr8-84-5000 (upstream control) and chrX-33. Whereas chr12-169 and chr13-1 showed no enrichment, chr8-84 and chrX-33 enrichment levels ranged between 1.5 to 2 fold.
In summary, no strong enrichment of Ago2 at the validation sites (at the most 2 fold) could be shown (compare Figure 4.17). In addition, no genomic element was consistently enriched over bead controls and the non-specific "upstream" primer sets. Unfortunately, optimization experiments using the 1.5 and 2 fold enriched sites failed to increase the signal to noise ratio. Therefore, no further validation of this analysis was undertaken.

4.4.3.5 GATC-Centered Analysis of DAMID Data

Inspired by the thesis of Zhou [2012], who analysed chromatin regulation using ChIP and a modified DamID protocol, an alternative analysis method for the DAMID data set was developed. In their methods Zhou [2012] described a peak caller, which counts the reads per GATC site and then sums up the read counts for all GATC sites per 5 kB stretches in the genome. This GATC-specific analysis could reduce background noise from non-GATC containing fragments non-specifically amplified during the PCR step of the DAMID experiment. Those fragments deriving from repetitive elements not containing GATC sites were found in the read coverage-based analysis by homer. However, no code or programme was available from their side. Therefore, a similar approach based on counts of reads overlapping GATC sites was pursued, because these sites are directly

modified by the Dam-fusion proteins. In the course of this analysis, the library composition, read trimming and analysis of the resulting count tables using the R software and the Bioconductor package DESeq, was performed.

4.4.3.6 Library Composition and Alignment

The quality control of the reads using FastQC [Andrews] returned good overall quality and, as expected, several contaminating sequences resulting from the DAMID-PCR primer. Bolger *et al.* [2014] found that the number of reads properly aligning to the genome, especially under strict conditions, is tremendously improved when contaminating sequences and bases of bad quality are trimmed. Therefore, the DAMID-PCR primer sequence, which does not match with the human genome and can be found at either end of the reads in forward or reverse orientation, had to be removed.
This placement of the primer sequence is caused by the sonication of the adapter-flanked, methylated fragments of the DAMID experiment preceding library preparation. Therefore, three types of reads are expected in the libraries: those containing the DAMID-PCR primer at the 5′-end, the ones containing the reverse complement of the primer at the 3′-end and lastly reads which arise from the middle of the methylated fragments carrying no primer sequence. As no programme was publicly available for this task, custom made python scripts were developed for the removal of DAMID-based contaminating sequences (Section 3.7.2). The programmes removed all sequences matching the DAMID-PCR primer, while a 5′-truncation of one base and sequences containing exactly one mismatch were allowed. The resulting sequences were quality trimmed at the 5′- and 3′-ends

Sample Name	Input Read Pairs	Both Surviving	Forward only	Reverse only	Dropped
Dam-Ago1	18516843	13583063	930541	2873297	1129942
Dam-Ago2	23181289	16264668	1244820	4096875	1574926
Dam-Ago2-Y529E	17882388	13062886	854161	2929537	1035804
Dam-Cbx1	20021720	12301959	965335	4922841	1831585
Dam	19134936	12379749	969645	4101014	1684528

Table 4.11: ***DAMID-preprocessing with Trimmomatic.***
Reads from the DAMID Libraries, were depleted of contaminating DAM-PCR-Primer. The resulting reads were loaded into Trimmomatic for quality trimming, removal of reads shorter than 25 bp and checking for read pairs. Roughly 65 to 70 % of the reads were then used for alignment.

and reads of a length below 25 *bp* were removed using the Trimmomatic software [Bolger *et al.*, 2014]. In the last step, the forward and reverse files containing each one read of a pair were checked for consistency (Table 4.11). Reads having no paired mate were removed from the downstream analysis.

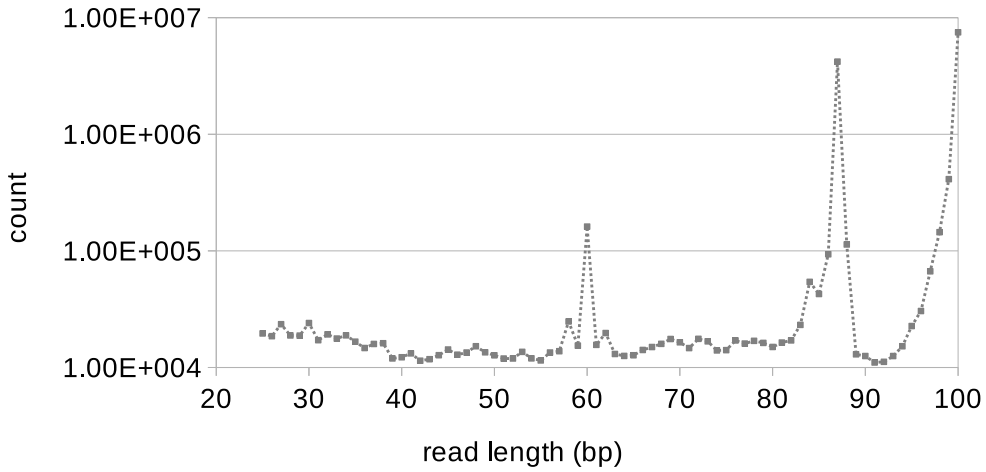


Figure 4.18: ***Length distribution of NGS data after quality trimming.***
Three peaks were observed in the distribution profile. The first marks sequences containing the fully DAMID-Adapter, the second consists of sequences, where the DAMID-PCR primer was removed and the third comprises all reads internal to DAMID fragments or those escaping the adapter/primer removal scripts.

The resulting length distribution of the reads was plotted (Figure 4.18). Three major peaks were produced: 1. the peak at 60 *bp*, resulting from the removal of the DAMID-Adapter; 2. the peak at 87 *bp*, based on the deletion of the DAMID-PCR primer sequence; 3. the peak at 100 *bp*, consisting of reads not carrying a DAMID-PCR primer sequence or truncations and mutations of this sequence, not recognized by the custom python scripts.
After preprocessing, about 60 to 70% of the reads were used for alignment against the human genome. (Dam-Ago1: 73%, Dam-Ago2: 70%, Dam-Ago2-Y529E: 73%, Dam-Cbx1: 61%, Dam: 65 %). The alignment was performed using Bowtie2, which supports paired-end alignments [Langmead *et al.*, 2009]. With an overall alignment rate of 90 % (except for Cbx1 75%), nearly all reads could be assigned a position in the genome (Table 4.12). However, also a substantial fraction of the reads mapped at more than one position.

Sample Names	Input Pairs	Aligned Exactly	Aligned More than Once	Sum of Aligned
Dam-Ago1	13583063	6887834	5926546	12814380
Dam-Ago2	16264668	9188295	5809520	14997815
Dam-Ago2-529E	13062886	7846576	4580777	12427353
Dam-Cbx1	12301959	7023116	2254000	9277116
Dam	12379749	10327232	1001407	11328639

Table 4.12: ***Read alignment using Bowtie2.***
The reads from NGS were aligned using Bowtie2. The overall alignment rate was over 90 % for all samples apart from Cbx1, where only 75 % of read pairs could be aligned. Of the aligned reads between 50 % and 60 % of read pairs aligned exactly, except for Dam, for which 83% of reads could be aligned exactly.

The aligned reads were further processed into .bam files, which were used for downstream analysis. Here, some of the multiple aligned reads were also taken forward, if one of mates of a pair mapped at a unique position. Then possible PCR duplicates were removed with samtools rmdup (remove duplicates) option [Li *et al.*, 2009b], leading to the aligned unique reads in Table 4.13. For Dam-

Sample Name	All Aligned Reads	Aligned Reads Unique
Dam-Ago1	6910145	4856874
Dam-Ago2	9220488	7097486
Dam-Ago2-Y529E	7874048	6230759
Dam-Cbx1	7070474	6272561
Dam	10364235	8535734

Table 4.13: ***Aligned and unique reads classified by samtools.***
The reads from the different samples were analysed with samtools rmdup, which compares the outer coordinates of all reads and removes duplicates. Unfortunately, up to 30% of the read pairs of Dam-Ago libraries were duplicates and mapped to exactly the same position.

Ago transduced cells, the duplication level ranged from 20 to 30%. The controls contained fewer duplicates ranging from 12 to 18% for Dam-Cbx1 and Dam, respectively.

The ratio of unique reads to PCR duplicate-containing reads was plotted per chromosome for each sample. Concentrating on chromosomes 1 to Y, it seems that for all samples the over-amplification leading to PCR duplicates was limited (compare Figure 4.19). However, for the mitochondrial chromosome (chrM), a high level of duplicate reads (~80 %) was observed in the Dam-Ago samples, which was reduced for Dam (~65 %) and on background level for the nuclear localised Dam-Cbx1 (~30 %) (Figure 4.19). If the fraction of removed reads are real PCR duplicates or result from DNA fragments actually present in the sample, can not be ruled out.

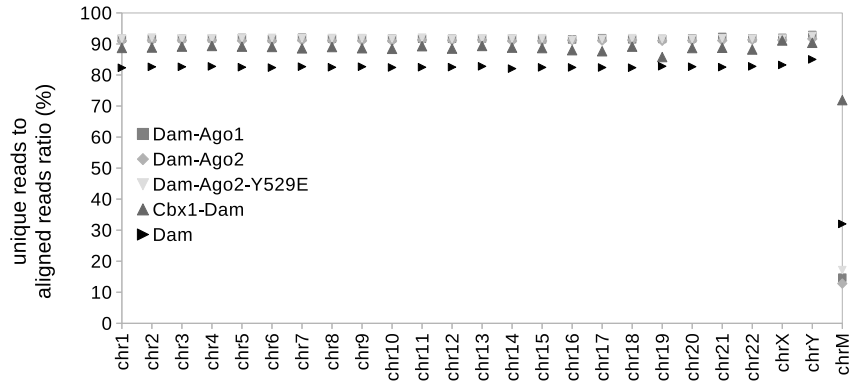


Figure 4.19: ***Analysis of PCR-duplicates.***
The number of unique reads was plotted against the number of aligned reads (unique + duplicate reads). For the human genome only about 10 % of the read pairs align at multiple locations. For the mitochondrial genome nearly 80% of the reads from the Dam-Ago libraries were removed by the "rmdup" option of samtools.

4.4.3.7 Analysis of Aligned and Unique Reads

Further analysis and identification of enriched regions was performed using R and the DESeq package. In order to get a first impression of the distribution of the data, reads were plotted per chromosome (*bp*) for all aligned (Figure 4.20 A) and for unique reads (Figure 4.20 B). All samples showed similar read numbers per chromosome for the nuclear encoded DNA. A striking difference was observed for the mitochondrial chromosome. Here, Dam-Ago reads were enriched at least 20 fold over Dam and nearly 200 fold compared to Dam-Cbx1 even for the unique reads, hinting at a potential mitochondrial function of Ago. This is analysed in Section 4.5. In the next step, it had to be determined, if Ago is enriched at specific regions of the genome.

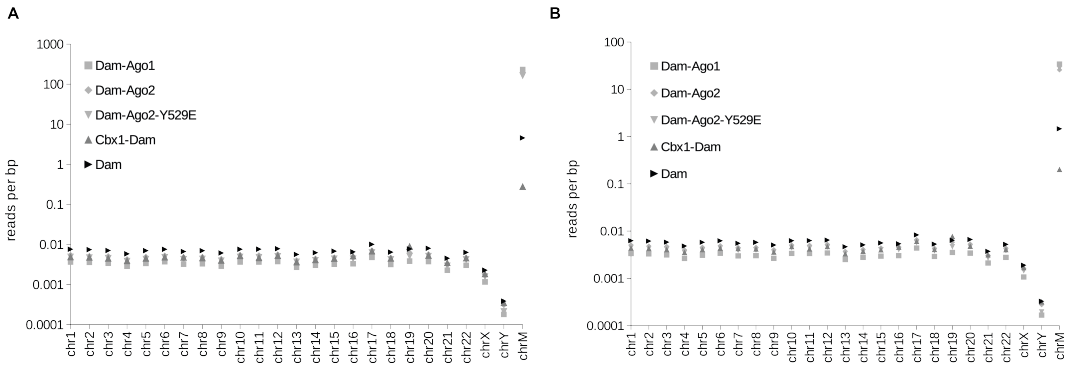


Figure 4.20: ***Distribution of reads per chromosome.***
A*) All aligned reads were plotted per chromosome (bp).* ***B****) All unique reads were plotted per chromosome (bp). Comparing both analyses, it became apparent, that despite the high duplication level of the data, mitochondria derived sequences were strongly enriched for samples transduced with Dam-Ago.*

Analysis Based on Aligned Reads

The first analysis was performed based on all aligned reads. To select for reads resulting from Dam methylation, only reads containing GATC had to be counted. Therefore, a GATC genome was constructed. This consists of the positional information of all GATCs found per chromosome (Section 3.7.5). Then an R script was used to count all the reads overlapping a GATC. The count table for all chromosomes was loaded into DESeq and analysed [Anders and Huber, 2010]. For a peak to be selected for analysis, however, reads in the Dam-Ago as well as in the Dam control were required, to estimate a proper difference between the samples.

Unfortunately, no statistically significant hits could be achieved. This is due to many regions without coverage by the Dam control. Furthermore, missing replicates resulted in a lack of data to estimate the dispersion of the count values at a given position and to calculate a false discovery rate needed for differential peak assessment. This means, that there are no ways to distinguish a random association from a targeted binding event causing a high amount of false positives under the top peaks further complicating the identification of promising candidate peaks.
Therefore, all regions enriched over the Dam control with a log2-fold change above two were selected and treated as potential peaks. A list of the GATC sites most enriched for Dam-Ago1 and Dam-Ago2 is presented below (Table 4.14, 4.15). A full peak list can be found on the CD accompanying the thesis, deposited at the library of the University of Regensburg. Remarkable at the peak list for Dam-Ago proteins is the accumulation of mitochondrial peaks as the strongest enriched ones. This supports the theory, that the band pattern observed during the DAM-ID experiment (Figure 4.15) might be due to mitochondrial localisation of Dam-Ago.

Table 4.14: ***Ago1 enriched sites in MRC-5 based on all aligned reads.***

Nr	chrom	start	end	Dam-Ago1	Dam-Ago2	Dam-Ago2-Y529E	Dam	DAM-Cbx1	log2Fold Change	repclass	external_gene_id	GO_domain	source
1	chrM	6905	6908	1462	1316	948	20	2	6.01	NA	MT-CO1	mito electron transport, cytochrome c to oxygen	exon
2	chrM	3065	3068	181882	129058	116570	3194	206	5.65	NA	MT-RNR2	NA	exon
3	chrM	10255	10258	26836	18091	16841	531	32	5.48	NA	MT-ND3	mito electron transport, NADH to ubiquinone	exon
4	chrM	2898	2901	63413	52998	41774	1491	96	5.23	NA	MT-RNR2	NA	exon
5	chrM	6461	6464	1527	1439	1062	36	2	5.23	NA	MT-CO1	mito electron transport, cytochrome c to oxygen	exon
6	chrM	3660	3663	129517	148049	92283	3189	178	5.17	NA	MT-ND1	mito electron transport, NADH to ubiquinone	exon
7	chr3	173920133	173920136	79	10	7	2	4	5.13	NA	NLGN1	positive regulation of synaptic vesicle exocytosis	intron
8	chrM	15592	15595	371766	432392	268561	9696	496	5.08	NA	MT-CYB	mito electron transport, ubiquinol to cytochrome c	exon
9	chrM	14260	14263	13456	9305	7926	373	29	5.00	NA	MT-ND6	mito electron transport, NADH to ubiquinone	exon
10	chrM	953	956	87630	107689	60339	2460	123	4.98	NA	MT-RNR1	NA	exon
11	chrM	2998	3001	13791	11984	10574	395	24	4.95	NA	MT-RNR2	NA	exon
12	chrM	14870	14873	7253	7675	5372	211	20	4.93	NA	MT-CYB	mito electron transport, ubiquinol to cytochrome c	exon
13	chr1	226144392	226144395	128	5	5	4	2	4.82	SINE/Alu	NA	NA	intergenic
14	chrM	15061	15064	95270	137286	63684	3111	227	4.76	NA	MT-CYB	mito electron transport, ubiquinol to cytochrome c	exon
15	chrM	15358	15361	299093	385348	225816	9825	478	4.75	NA	MT-CYB	mito electron transport, ubiquinol to cytochrome c	exon
16	chrM	1229	1232	44012	67860	28494	1449	56	4.75	NA	MT-RNR1	NA	exon
17	chrM	7659	7662	1419	1505	1075	48	2	4.71	NA	MT-CO2	mito electron transport, cytochrome c to oxygen	exon
18	chr3	104509849	104509852	59	2	27	2	4	4.70	SINE/MIR	NA	NA	intergenic
19	chr2	99926073	99926076	58	6	7	2	2	4.68	SINE/Alu	C2orf15	NA	intron
20	chrM	742	745	8276	8688	6122	297	22	4.62	NA	MT-RNR1	NA	exon
21	chrM	7860	7863	1460	1488	991	54	2	4.58	NA	MT-CO2	mito electron transport, cytochrome c to oxygen	exon
22	chr11	46969371	46969374	51	3	9	2	2	4.49	NA	C11orf49	NA	intron
23	chr15	56452380	56452383	48	3	2	2	2	4.41	NA	RFX7	regulation of transcription, DNA-dependent	intron
24	chr17	55702830	55702833	47	8	6	2	4	4.38	NA	MSI2	stem cell development	intron
25	chr7	20391826	20391829	46	8	4	2	8	4.35	NA	ITGB8	ganglioside metabolic process	intron
26	chr16	67627611	67627614	44	9	4	2	2	4.28	NA	CTCF	regulation of histone methylation	intron
27	chr3	11724872	11724875	42	3	9	2	7	4.21	LTR/ERVL	VGLL4	regulation of transcription, DNA-dependent	intron
28	chr18	56463767	56463770	41	4	5	2	6	4.18	NA	NA	NA	intergenic
29	chr5	15016403	15016406	41	2	7	2	5	4.18	LINE/L1	NA	NA	intergenic
30	chr5	131065958	131065961	40	2	11	2	2	4.14	NA	CTC-432M15.3	signal transduction	intron
31	chr7	4726099	4726102	40	5	11	2	7	4.14	NA	FOXK1	muscle organ development	intron
32	chr2	99926386	99926389	79	3	16	4	5	4.13	SINE/MIR	C2orf15	NA	intron
33	chr11	34494094	34494097	39	18	14	2	4	4.11	SINE/Alu	NA	NA	intergenic
34	chr11	56983432	56983435	39	13	7	2	2	4.11	LINE/CR1	NA	NA	intergenic
35	chr14	61992453	61992456	39	4	28	2	6	4.11	NA	PRKCH	protein phosphorylation	exon
36	chr6	76487248	76487251	39	2	24	2	3	4.11	LINE/L2	MYO6	regulation of secretion	intron
37	chr10	33300400	33300403	38	4	13	2	6	4.07	SINE/MIR	RP11-462L8.1	NA	intron
38	chr4	169238967	169238970	38	3	3	2	2	4.07	LINE/L2	DDX60	positive regulation of RIG-I signaling pathway	intron
39	chr11	111400325	111400328	37	6	2	2	5	4.03	LINE/L1	C11orf88	NA	intron
40	chr15	63520453	63520456	37	2	7	2	2	4.03	LTR/ERVL-MaLR	RAB8B	positive regulation of cell projection organization	intron

Table 4.15: ***Ago2 enriched sites in MRC-5 based on all aligned reads.***

Nr	Chrom	Start	End	Dam-Ago1	Dam-Ago2	Dam-Ago2-Y529E	Dam	Dam-Cbx1	Log2-Fold Change	Repclass	External Gene_ID	GO_Domain	Source
1	chrM	6905	6908	1462	1316	948	20	2	5.64	NA	MT-CO1	mito electron transport, cytochrome c to oxygen	exon
2	chrM	1229	1232	44012	67860	28494	1449	56	5.14	NA	MT-RNR1	NA	exon
3	chrM	3660	3663	129517	148049	92283	3189	178	5.13	NA	MT-ND1	mito electron transport, NADH to ubiquinone	exon
4	chrM	15592	15595	371766	432392	268561	9696	496	5.07	NA	MT-CYB	mito electron transport, ubiquinol to cytochrome c	exon
5	chrM	15061	15064	95270	137286	63684	3111	227	5.06	NA	MT-CYB	mito electron transport, ubiquinol to cytochrome c	exon
6	chrM	953	956	87630	107689	60339	2460	123	5.05	NA	MT-RNR1	NA	exon
7	chrM	3065	3068	181882	129058	116570	3194	206	4.93	NA	MT-RNR2	NA	exon
8	chrM	6461	6464	1527	1439	1062	36	2	4.92	NA	MT-CO1	mito electron transport, cytochrome c to oxygen	exon
9	chrM	15358	15361	299093	385348	225816	9825	478	4.89	NA	MT-CYB	mito electron transport, ubiquinol to cytochrome c	exon
10	chrM	14870	14873	7253	7675	5372	211	20	4.78	NA	MT-CYB	mito electron transport, ubiquinol to cytochrome c	exon
11	chrM	2898	2901	63413	52998	41774	1491	96	4.75	NA	MT-RNR2	NA	exon
12	chrM	10255	10258	26836	18091	16841	531	32	4.69	NA	MT-ND3	mito electron transport, NADH to ubiquinone	exon
13	chrM	7659	7662	1419	1505	1075	48	2	4.57	NA	MT-CO2	mito electron transport, cytochrome c to oxygen	exon
14	chrM	2998	3001	13791	11984	10574	395	24	4.52	NA	MT-RNR2	NA	exon
15	chrM	742	745	8276	8688	6122	297	22	4.47	NA	MT-RNR1	NA	exon
16	chr2	46623686	46623689	3	57	11	2	5	4.43	LINE/L2	NA	NA	intergenic
17	chrM	7860	7863	1460	1488	991	54	2	4.38	NA	MT-CO2	mito electron transport, cytochrome c to oxygen	exon
18	chrM	14260	14263	13456	9305	7926	373	29	4.24	NA	MT-ND6	mito electron transport, NADH to ubiquinone	exon
19	chr2	24996479	24996482	5	45	2	2	5	4.09	NA	NA	NA	intergenic
20	chr3	52303949	52303952	22	41	10	2	3	3.95	NA	WDR82	histone H3-K4 methylation	intron
21	chr12	110881838	110881841	13	40	6	2	2	3.92	LINE/L2	ARPC3	regulation of actin filament polymerization	intron
22	chr17	44229979	44229982	10	39	2	2	2	3.88	NA	KANSL1	histone H4-K8 acetylation	intron
23	chr6	17973924	17973927	10	39	29	2	2	3.88	Low_complexity	KIF13A	cargo loading into vesicle	intron
24	chr1	184143844	184143847	3	73	7	4	12	3.79	LINE/L1	NA	NA	intergenic
25	chr10	126634199	126634202	26	36	3	2	5	3.77	NA	RP11-298J20.3	protein deubiquitination involved in catabolic process	exon
26	chr3	45085853	45085856	28	34	9	2	12	3.68	LINE/L2	NA	NA	intergenic
27	chr6	11014221	11014224	7	33	3	2	6	3.64	Low_complexity	ELOVL2	linoleic acid metabolic process	intron
28	chr14	73533654	73533657	14	32	3	2	4	3.60	LTR/ERVL-MaLR	RBM25	regulation of alternative mRNA splicing, via spliceosome	intron
29	chr3	128747165	128747168	2	32	11	2	10	3.60	SINE/MIR	EFCC1	NA	intron
30	chr16	28871802	28871805	6	31	8	2	11	3.55	Low_complexity	SH2B1	regulation of DNA biosynthetic process	intron
31	chr2	191352497	191352500	9	31	2	2	8	3.55	LINE/L2	MFSD6	transmembrane transport	intron
32	chr8	134923501	134923504	6	31	5	2	2	3.55	LINE/L1	NA	NA	intergenic
33	chr1	8963004	8963007	2	29	4	2	9	3.45	LTR/ERVL	NA	NA	intergenic
34	chr8	37802033	37802036	3	29	8	2	11	3.45	LTR/ERVK	NA	NA	intergenic
35	chr9	75191128	75191131	5	29	3	2	4	3.45	NA	TMC1	vestibular reflex	intron
36	chr5	124179260	124179263	4	43	5	3	3	3.44	NA	NA	NA	intergenic
37	chr12	68799719	68799722	2	28	2	2	3	3.40	NA	RP11-81H14.2	NA	intron
38	chr16	67337287	67337290	24	42	28	3	14	3.40	NA	KCTD19	protein homooligomerization	intron
39	chr18	8834988	8834991	6	28	3	2	17	3.40	SINE/Alu	NA	NA	intergenic
40	chr5	10822093	10822096	3	28	38	2	10	3.40	SINE/Alu	NA	NA	intergenic

Subsequently, the peaks were classified in genomic elements and the frequencies were compared to the ones in the human genome (build hg19) (Figure 4.21). The peaks from the Dam-Ago1 and 2 transduced cells showed enrichment for the class of SINE elements, LTR transposons and exons, whereas intergenic sequences appeared to be reduced. Only a few peaks mapped to the intergenic space compared to the relative abundance in the human genome. The enrichment of SINE elements, LTR transposons and exons across all peaks in Ago1 and 2 samples (Figure 4.21) is in good agreement with the peaks showing the strongest log2-fold enrichment (Table 4.14, Table 4.15). In the latter, especially the mitochondrial peaks are supported by high read counts (median count value 13456 for Ago1 and 18091 for Ago2), whereas the genomic peaks show comparably low read numbers with a median count value of 19 for Ago1 and 17 for Ago2. Therefore, also the classification into genomic elements derived from the Dam-Ago1 and Dam-Ago2 peaks is based on low read numbers (median count value 18) and has to be interpreted very carefully.

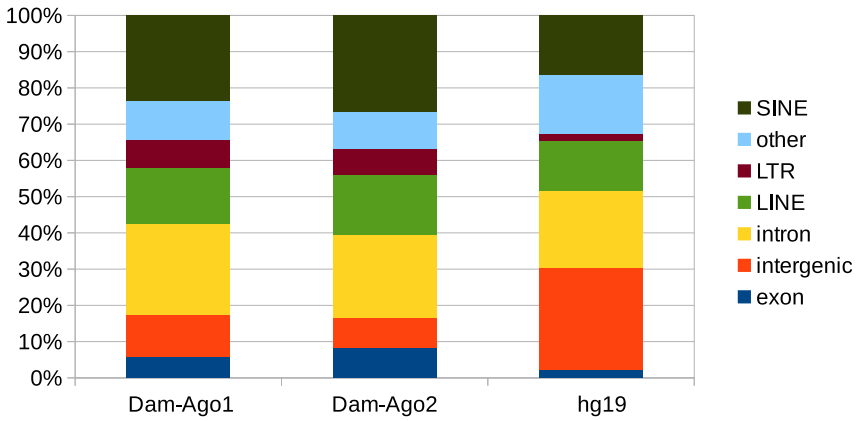


Figure 4.21: ***Comparison of the frequencies of genomic elements between samples and the human genome.***
The peaks of all samples, which showed a log2-fold change higher two, were classified in genomic elements and compared to the frequencies of the respective elements in the human genome. Dam-Ago1 and 2 seemed to be enriched in SINE elements and exons. Only a few peaks mapped to intergenic positions.

Analysis Based on Unique Reads[1]

In order to generate a complete picture a second analysis was performed on the GATC overlapping reads after removal of potentially duplicated sequences. Again, a count table was generated and analysed with DESeq [Anders and Huber, 2010]. In this analysis some of the peaks reached estimated p-values close to the commonly accepted 0.05. In order to get a good overview, values with the strongest log2-fold changes ($p < 0.07$) are shown for Dam-Ago1 and Dam-Ago2 in the following (Table 4.16, 4.17). Unfortunately, only eight peaks from the Dam-Ago1 experiment and 40 for Dam-Ago2 met these criteria. For both samples the analysis based on only unique reads identified mitochondrial regions as the strongest enriched ones. This is in good agreement with the observed band pattern from the DAMID enrichment PCR (Section 4.4.3.1). For these bands, Sanger sequencing revealed a high frequency of mitochondrial derived sequences.
However as described above, none of these peaks are statistically significant. Due to the lack of replicates, the dispersion of the count values at a given position could not be evaluated. Further, no good estimation of p-values and correction for multiple testing, which is the calculation of a false discovery rate, could be performed. Therefore, we can not decide on true and false positive peaks among the lists presented. Nevertheless, these lists are a very conservative evaluation of the data and therefore could contain some true positive peaks.

[1]This analysis was performed after my contract ended at the department of Prof. Dr. G. Meister, therefore no experimental validation of the identified peaks was possible.

Table 4.16: ***Ago1 enriched sites in MRC-5 based on unique reads.***

Nr	chrom	start	end	Ago1	Dam	log2FoldChange	pval	repclass	external_gene_id	GO_domain	source
1	chrM	8593	8596	195	8	4.8297227351	0.050152579	NA	MT-ATP6	mitochondrial ATP synthesis coupled proton transport	exon
2	chrM	11923	11926	181	8	4.7222383084	0.0498456418	NA	MT-ND4	ATP synthesis coupled electron transport	exon
3	chrM	3694	3697	149	8	4.4415609418	0.0523342731	NA	MT-ND1	mitochondrial electron transport, NADH to ubiquinone	exon
4	chr12	98704008	98704011	79	5	4.2042450746	0.0646079682	LINE/CR1	NA	NA	intergenic
5	chr4	40858022	40858025	60	4	4.1292830169	0.0624407956	NA	APBB2	extracellular matrix organization	intron
6	chrX	139846408	139846411	58	4	4.0803734165	0.0637794038	NA	LINC00632	NA	intron
7	chr11	94421703	94421706	54	4	3.9772799235	0.0666314286	SINE/Alu	RP11-867G2.8	NA	intron
8	chr10	506102	506105	50	4	3.8662486111	0.0698391939	NA	DIP2C	metabolic process	intron

Table 4.17: ***Ago2 enriched sites in MRC-5 based on unique reads.***

Nr	chrom	start	end	Ago2	Dam	log2FoldChange	pval	repclass	external_gene_id	GO_domain	source
1	chrM	8593	8596	220	8	5.103	0.063	NA	MT-ATP6	mitochondrial ATP synthesis coupled proton transport	exon
2	chrM	3694	3697	135	8	4.399	0.066	NA	MT-ND1	mitochondrial electron transport, NADH to ubiquinone	exon
3	chrM	11923	11926	124	8	4.276	0.068	NA	MT-ND4	ATP synthesis coupled electron transport	exon
4	chr1	184143844	184143847	59	4	4.205	0.050	LINE/L1	NA	NA	intergenic
5	chr1	92455156	92455159	44	4	3.781	0.057	SINE/Alu	BRDT	positive regulation of transcription during meiosis	intron
6	chr11	67848051	67848054	44	4	3.781	0.057	Simple_repeat	CHKA	CDP-choline pathway	intron
7	chr11	67847691	67847694	43	4	3.748	0.058	NA	CHKA	CDP-choline pathway	intron
8	chr5	73244162	73244165	42	4	3.714	0.059	NA	NA	NA	intergenic
9	chr10	78891499	78891502	41	4	3.679	0.059	LTR/ERV1	KCNMA1	ion transport	intron
10	chr11	77399603	77399606	41	4	3.679	0.059	NA	RSF1	nucleosome positioning	intron
11	chr3	172493002	172493005	40	4	3.644	0.060	SINE/Alu	ECT2	neurotrophin TRK receptor signaling pathway	intron
12	chr2	227600363	227600366	39	4	3.607	0.061	NA	IRS1	positive regulation of glucose metabolic process	exon
13	chr7	155081352	155081355	39	4	3.607	0.061	LINE/L2	NA	NA	intergenic
14	chr18	26743744	26743747	38	4	3.570	0.061	NA	NA	NA	intergenic
15	chr3	12814379	12814382	38	4	3.570	0.061	NA	NA	NA	intergenic
16	chr14	74223034	74223037	37	4	3.531	0.062	NA	ELMSAN1	regulation of transcription, DNA-dependent	intron
17	chr6	111284812	111284815	37	4	3.531	0.062	NA	GTF3C6	tRNA transcription from RNA polymerase III promoter	intron
18	chr1	16123629	16123632	36	4	3.492	0.063	SINE/Alu	NA	NA	intergenic
19	chr11	15856172	15856175	36	4	3.492	0.063	LINE/L2	NA	NA	intergenic
20	chr17	9273483	9273486	36	4	3.492	0.063	SINE/Alu	STX8	vesicle fusion	intron
21	chr20	49923566	49923569	35	4	3.451	0.064	NA	NA	NA	intergenic
22	chr3	123547309	123547312	35	4	3.451	0.064	LTR/ERVL-MaLR	MYLK	tonic smooth muscle contraction	intron
23	chr9	95387833	95387836	35	4	3.451	0.064	NA	IPPK	inositol phosphorylation	intron
24	chr17	30651373	30651376	34	4	3.409	0.065	NA	RHBDL3	proteolysis	exon
25	chr2	217565669	217565672	34	4	3.409	0.065	NA	AC007563.5	NA	intron
26	chr3	177242889	177242892	34	4	3.409	0.065	NA	LINC00578	NA	intron
27	chr3	182948563	182948566	34	4	3.409	0.065	LINE/CR1	MCF2L2	regulation of Rho protein signal transduction	intron
28	chr7	95916177	95916180	34	4	3.409	0.065	NA	SLC25A13	malate-aspartate shuttle	intron
29	chr1	62927916	62927919	33	4	3.366	0.066	LINE/L1	DOCK7	activation of Rac GTPase activity	intron
30	chr1	244467303	244467306	33	4	3.366	0.066	LTR/ERV1	NA	NA	intergenic
31	chr10	28263453	28263456	33	4	3.366	0.066	SINE/Alu	ARMC4	NA	intron
32	chr11	109512516	109512519	33	4	3.366	0.066	NA	NA	NA	intergenic
33	chr13	20535643	20535646	33	4	3.366	0.066	NA	ZMYM2	regulation of transcription, DNA-dependent	intron
34	chr3	189873818	189873821	33	4	3.366	0.066	NA	NA	NA	intergenic
35	chr4	169638802	169638805	33	4	3.366	0.066	NA	PALLD	cytoskeleton organization	intron
36	chr5	32423224	32423227	33	4	3.366	0.066	SINE/MIR	ZFR	multicellular organismal development	intron
37	chr1	43995062	43995065	32	4	3.322	0.067	NA	PTPRF	transmembrane receptor protein tyrosine phosphatase signaling pathway	intron
38	chr1	57294791	57294794	32	4	3.322	0.067	NA	NA	NA	intergenic
39	chr1	153865521	153865524	32	4	3.322	0.067	NA	GATAD2B	regulation of transcription, DNA-dependent	intron
40	chr10	3549103	3549106	32	4	3.322	0.067	SINE/Alu	NA	NA	intergenic

4.4.3.8 Validation of DAMID Results

In the above analyses, based on either all aligned or only on unique reads, mitochondrial peaks emerged as highly enriched. This correlates with the high frequency of mitochondrial sequences observed in direct cloning experiments from the strong bands of the DAMID enrichment PCR.

Therefore, specific primer sets were designed for validation of the Ago2 association with mitochondrial DNA by ChIP experiments. Similar to the first validation approach primer sets were picked using the UCSC genome browser together with indexed .bam files representing the number of observed reads at each genomic position. Two Ago2-specific purified antibodies (11A9 and 1526) were used for ChIP validation (Figure 4.22).

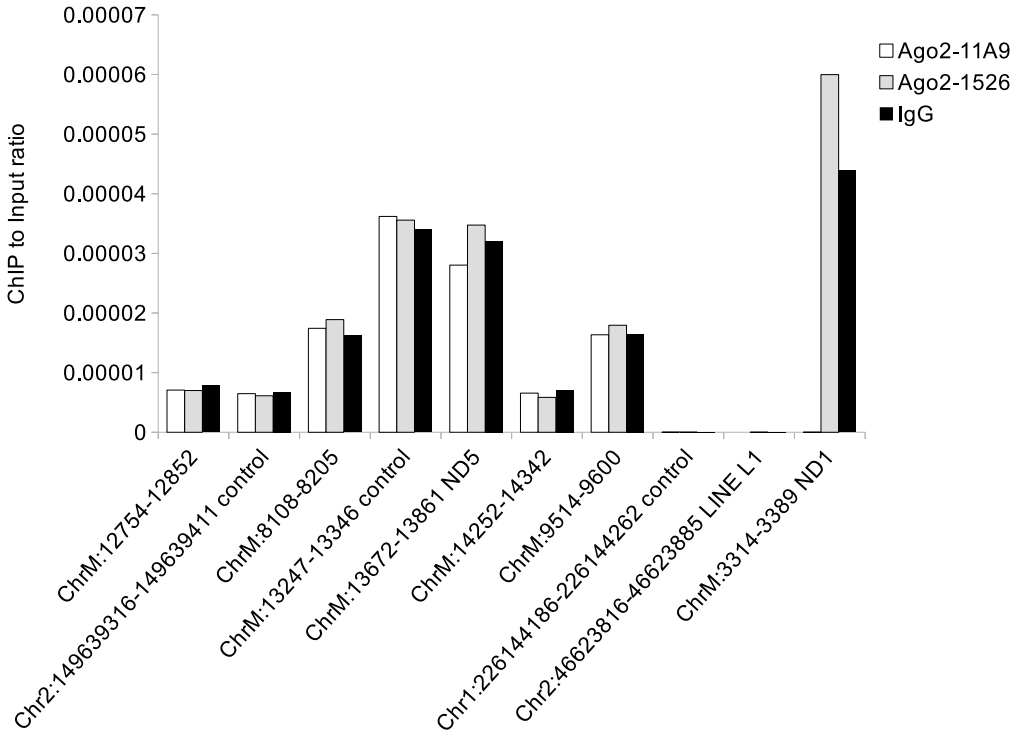


Figure 4.22: ***Validation of Dam-Ago2 association with mitochondrial DNA by ChIP.*** *MRC-5 cells were crosslinked with formaldehyde for 10 minutes at RT and sonicated. The resulting lysate was cleared by centrifugation and used for ChIP using two Ago2-specific antibodies (rat mAB:11A9, rabbit pAB: 1526) and rabbit IgG as control. Several mitochondrial and human genomic DNA specific primers sets were used for qPCR amplification following PCI extraction of the precipitated DNA. However, no locus showed enrichment compared to IgG control. One of several optimisation experiments shown.*

Several mitochondrial loci were tested for Ago2 association in a row of ChIP optimisation experiments applying several Ago2-specific antibodies on lysates of crosslinked MRC-5 cells. However, non of the tested loci showed any enrichment compared to the IgG background control. The high amounts of mitochondrial DNA in the input samples, which were readily amplified, suggest that lysis conditions were also efficient for mitochondria and sufficient crosslinked DNA was present in the ChIP lysates used for Ago2 IP. In addition, crosslink and sonication conditions were controlled by agarose gel electrophoresis of input DNA. These results imply that no direct interaction of Ago2 with mitochondrial DNA can be trapped applying formaldehyde crosslink followed by ChIP.

4.4.3.9 Summary of the MRC-5 DAMID Experiment

In summary, no direct interaction of Ago2 with human chromatin, which was analysed in Section 4.4.3.3, or mitochondrial DNA, examined in Section 4.4.3.5, could be demonstrated. Even though a set of 18 genomic (Figure 4.17) and 7 mitochondrial loci (Figure 4.22) were tested by ChIP experiments, no peak could be validated.
The observed GATC-methylation in mitochondria, which was supported by high read counts in Dam-Ago samples, can also be interpreted as positional information, thus supporting mitochondrial localisation of Ago proteins. In this case, mtDNA methylation is just a by-product due to the limited space in mitochondria and the open accessible mitochondrial DNA. In human lung fibroblasts about 300 mitochondria are found, each of which contains 2-3 DNA copies [Robin and Wong, 1988]. The circular mitochondrial DNA has a size of ~16 kB and contains 24 GATC sites. Furthermore, the level of mitochondrial DNA methylation measured for Dam-Cbx1 was even lower than that for the Dam control, which marks as entirely nuclear protein the level of non specific mitochondrial methylation. Both observations suggest that the high level of GATC methylation observed in Dam-Ago treated cells might be due to a specific import of Ago proteins into mitochondria. This hypothesis is tested in Section 4.5.

4.4.4 DAMID Experiment with Mouse Embryonic Stem Cells

The validation of potential chromatin interaction sites of Ago2 in MRC-5 cells failed, suggesting non-specific Ago chromatin interactions in these assays. Nev-

ertheless, Ago2 might associate with chromatin in earlier steps of development. For LINE L1 elements, which are reactivated by hypomethylation in the early blastocyste, it was shown, that Dicer and Ago2 are required to restrict retrotransposition [Ciaudo *et al.*, 2013]. Knock out of Dicer or Ago1 to 4 led to increased retrotransposition of L1 elements and a differentiation defect in Dicer -/- mESCs. The repression of L1 transcripts seems to be RNAi-mediated based on 22 *nt* endo-siRNAs derived from sense and anti-sense transcripts of L1, which are at least in part processed by Dicer [Ciaudo *et al.*, 2013]. By partially restoring the Ago2 level in Dicer -/- mESC, which is normally strongly diminished, the differentiation ability of Dicer -/- mESCs could be rescued [Ciaudo *et al.*, 2013]. This suggests that Ago2 inhibits the detrimental effect of retrotransposition and enables the silencing of the coding sequence of L1 elements. However, the mechanism of L1 transcriptional silencing remains to be elucidated.

Therefore, a second DAMID experiment with the human Ago2 as bait in mouse embryonic stem cells (mESCs), was performed to examine, if Ago2 plays a role in transcriptional silencing of repetitive genomic elements. In collaboration with Prof. Dr. C. Ciaudo (ETH Zürich) stable Dam or Dam-Ago2 expressing mESCs were created and chromatin association sites were analysed using DAMID. Human Ago2 differs from the mouse protein only by 7 amino acids within the first 35 residues at the N-terminus, therefore we assume a similar target range for the exogenous protein. Another advantage of mESC as model system is that cell differentiation can be induced which may lead to a change in binding site specificities of Ago2. These variations in the binding pattern of Ago2 will strengthen the analysis and interpretation of discovered target sites.

Cells from the male E14 WT mESC line (129/Ola background) [Hooper *et al.*, 1987] were stably transfected with constructs carrying Dam-Ago2 or only Dam as inserts. The constructs contained also the HSP promoter controlling the expression of the Dam fusion proteins, granting the low expression level required for DAMID. Following screening for hAgo2 expression, two independent clones for each construct were chosen (Dam-Ago2: clone 4 and D, Dam: clone 30b and F) and expanded before the methylated DNA was extracted. This work was contributed by C. Ciaudo (ETH Zürich). The isolated DNA was subsequently shipped to Regensburg and DAMID was performed (see Section 3.6.5).

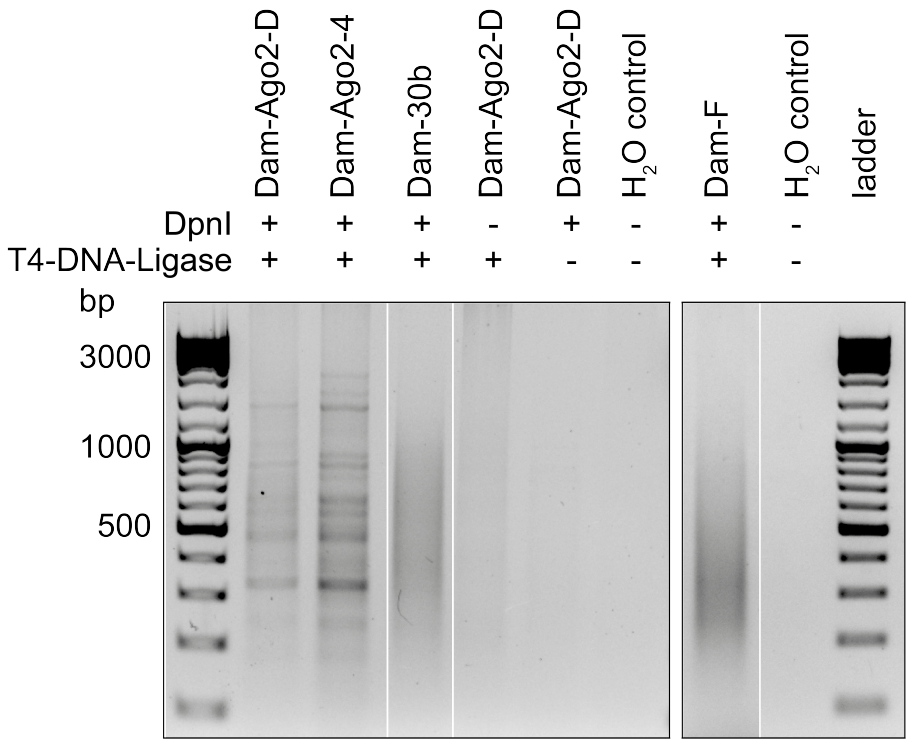


Figure 4.23: ***Examination of the DNA fragments of the mESC DAMID experiment.***
The DNA was extracted from the different stable mESCs and prepared according to the DAMID procedure. After selection for methylated DNA fragments, the samples were loaded on an agarose gel (2 %). The Dam control cell lines led to enrichment of a broad range of target sequences, as proven by a smear. The Dam-Ago2 expressing cells showed a band pattern reminiscent of the one observed for MRC-5 cells. The specificity of the assay was controlled by samples not treated with either DpnI or T4-DNA-Ligase.

After GATC-selective PCR amplification, the DNA fragments methylated by Dam or Dam-Ago2 were analysed on an agarose gel (Figure 4.23). The DAMID experiment was controlled by samples lacking DpnI or T4-DNA-Ligase (compare Section 4.4.3). Omitting these steps prevented the formation of PCR-amplifiable fragments confirming the specificity of the method. Dam yielded a smear as expected for an amplification of a broad range of targets. On the contrary, Dam-Ago2 in mESCs resulted again in the band pattern already observed for Dam-Ago2 in MRC-5 cells. Then, the replicates were fragmented by sonication to about 250 *bp* and NGS libraries were prepared with the NEB Next Library preparation kit. High-throughput sequencing was performed by the Kompetenz Zentrum für Fluoreszente Bioanalytik (KFB, Regensburg, Germany).

4.4.4.1 Bioinformatic Analysis of DAMID in mESCs

As the libraries contained different barcodes, they were pooled and sequenced in one lane on an Illumina HISeq 1500 machine. The results from the NGS run are shown in Table 4.18.

Sample Name	Index	Yield (Mbases)	Reads	% of Raw Clusters per Lane	% Perfect Index Reads	% of => Q30 Bases	Mean Quality
Dam_30b	ACTTGA	5,884	58,844,556	17.80	99.04	93.38	35.98
Dam_Ago2_4	ATCACG	5,204	52,036,500	15.74	99.17	95.09	36.64
Dam_Ago2_D	TTAGGC	6,505	65,050,460	19.68	99.22	95.18	36.75
Dam_F	GATCAG	4,9	49,004,476	14.82	99.12	92.97	35.78

Table 4.18: ***Run information of the Illumina HiSeq.***
The NGS yielded high read numbers for all samples. Also the quality of the reads was nearly perfect with a mean Phred score of 36, which is a probability that 1 in 4000 bases is called wrong.

The numbers of reads acquired for each sample ranged from 50 to 65 million reads, with each library generating between 15 and 20 % of the sequencing clusters on the lane of the flowcell. Nearly all reads could be assigned to their library (99%) and the quality of the reads was explicitly good. At least 93 % of the reads had quality values equal or higher than the Q30 score, which is 1 in 1000 bases is wrongly called. In the next step, contaminating sequences had to be removed and quality trimming was performed, as it greatly enhances the mapping efficiency of aligners [Bolger *et al.*, 2014].

4.4.4.2 Library Composition and Alignment

As described above, the read quality and sequencing biases were assessed using FastQC (Section 3.7.1)[Andrews], which reported good quality of reads but a high duplication rate. Then, the DAMID-PCR primer sequences and sequencing adapters were removed by custom python scripts (see Section 3.7.2). The reads were checked for consistency of pairs, minimal read length (25 bp) and quality using Trimmomatic (Section 3.7.1) [Bolger *et al.*, 2014]. After these selection steps, the resulting length distribution of the reads was analysed (Figure 4.24).

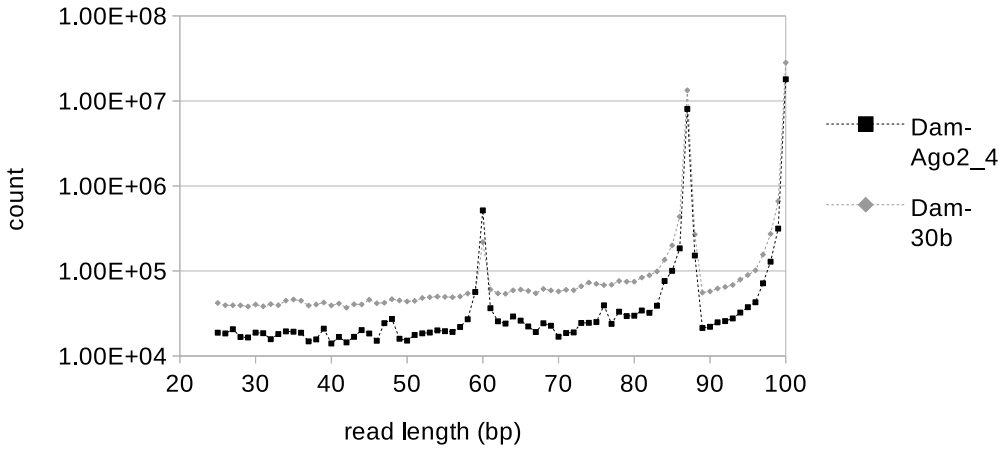


Figure 4.24: ***Length distribution of mESC NGS data after quality trimming.***
Three peaks were observed in the distribution profile, consistent with the data for MRC-5. The first marks sequences containing the fully DAMID-Adapter, the second comprises sequences, where the DAMID-PCR primer was removed and the third consists of all reads internal to methylated fragments or those escaping the adapter/primer removal scripts.

Three major peaks were produced: 1. the peak at 60 bp, caused by the depletion of the DAMID adapter sequence; 2. the peak at 87 bp, due to the removal of the DAMID-PCR primer sequence; 3. the peak at 100 bp, consisting of reads not carrying a DAMID-PCR primer sequence or truncations and mutations of this sequence being not recognized by the python scripts (Figure 4.24).
The results from the trimming are summarized in Table 4.19. Nearly all reads (at least 95 %) survived the primer removal and quality trimming. However, one third of the reads contained a removable DAMID-PCR primer sequence as can be seen in the length distribution signature of the DAMID adapter and primer at 60 and 87 bp, respectively (Figure 4.24).

Sample Name	Input	Surviving Both	Surviving F	Surviving R	Dropped
Dam_30b	29422278	28120075 (95,57%)	829464 (2,82%)	314931 (1,07%)	157808 (0,54%)
Dam_F	24502238	23294495 (95,07%)	778890 (3,18%)	276694 (1,13%)	152159 (0,62%)
Dam_Ago2_4	26018250	25051661 (96,28%)	596399 (2,29%)	262247 (1,01%)	107943 (0,41%)
Dam_Ago2_D	32525230	31487411 (96,81%)	657336 (2,02%)	1281824 (0,87%)	98659 (0,30%)

Table 4.19: ***Summary of the trimming process.***
The reads removed of contaminating adaptor and primer sequences were loaded into the trimming programme. After a check for quality and paired mates, the reads were mapped.

The alignment of the trimmed reads to the mouse genome (build mm10) was performed using bowtie2 (see Section 3.7.4) [Langmead *et al.*, 2009], with the only difference, that instead of one reported alignment, up to five valid alignments were recorded (Table 4.20). Here, the overall alignment rate was the lowest for Dam_Ago2_D. The low rate of exactly aligned read pairs already observed for Dam-Ago2 in MRC-5 (56 %) (compare Table 4.12) was even more severe in mESCs with only 20 % of pairs. This is in contrast to the Dam control reads of which at least 66 % were assigned to exact genomic positions. However, given that the mESC cells derived from the mouse strain 129/Ola and the mouse genome build (mm10) is based on the C57BL/6J strain, a reduced mapping rate is expected.

Sample Name	Paired	Aligned Exactly Once	Aligned Multiple Times	Aligned 0 Times	Overall Alignment Rate	Exact Alignment Rate
Dam_30b	28120075	21068368	5157670	1894037	93%	75%
Dam_F	23294495	15305718	5735976	2252801	90%	66%
Dam_Ago2_4	25051661	5649204	17033104	2369353	90%	23%
Dam_Ago2_D	31487411	6225457	20042102	5219852	83%	20%

Table 4.20: ***Bowtie2 alignment to the mouse genome.***
Trimmed reads were mapped to the mouse genome (build mm10) with a good overall alignment rate ranging between 83 and 93 %. However, the number of exactly mapping reads for both Dam_Ago2 samples is remarkably low when compared to the Dam control samples.

As up to five valid alignments were reported per read pair, a python script was created sorting the reads in exactly aligned reads, reads mapping to chrM, reads mapping putatively to chrM (based on comparison with the data of MRC-5 cells) and a last set of multiple aligning reads called repetitive. A summary of the sorting process can be found in Table 4.21.

Sample Name	Aligned Pairs	Exactly Aligned	ChrM Assigned	Putative ChrM	Repetitive	Total Repetitive Pairs
Dam_30b	26226038	23677576	103758	519542	1925162	18553154
Dam_F	21041694	17959258	145982	670160	2266294	22077582
Dam_Ago2_4	22682308	13102643	1457365	6003177	2119123	20891320
Dam_Ago2_D	26267559	14847216	1082752	5752629	4584962	45323234

Table 4.21: ***Sorting of mESC DAMID reads after alignment.***
After applying bowtie2 with the setting to return up to five valid alignments, the read pairs were sorted according to their mapping behaviour and, if not unique, due to their location.

For the Dam control samples 23.7 and 17.9 Mio paired reads aligned exactly. Pairs mapping to the mitochondrial DNA were less than 4% and about 2 Mio (10%) paired reads aligned to repetitive sequences. For Dam_Ago2 only 13.1 and 14.8 Mio paired reads could be assigned to one genomic position. A large number of paired reads, 1.4 and 1.0 Mio, mapped exactly to mitochondrial DNA, while 6 and 5.7 Mio paired reads mapped to multiple positions, with the best alignment score for chrM. The class of repetitive sequences contained 2.1 and 4.5 Mio paired reads (Table 4.21). In the subsequent analysis only exactly matching read pairs and read pairs for which one of the pair could be exactly assigned were used.
The FastQC analysis of the data had already revealed a high number of duplicate reads. This rises the question, if the large amount of duplicates is due to over-amplification during library preparation or results from the PCR step in the DAMID protocol which is used for specific signal amplification. Given that in the course of the Dam_Ago2-specific DAMID experiment only a limited number of target sequences were amplified by PCR, this step could also cause the observed amount of duplicate reads. To properly assess both scenarios, two data sets were created with the samtools package [Li *et al.*, 2009b], one carrying all duplicates and a second with duplicates removed (Table 4.22). Especially for Dam_Ago2 samples the numbers of paired reads dropped to only 14% and 28% of aligned pairs, when duplicates were removed. Whereas the control samples contained 88% and 81% unique reads.

Sample Name	Aligned Reads	Unique Reads
Dam_30b	23781334	20771640
Dam_F	18105240	14076653
Dam_Ago2_4	14560008	1821198
Dam_Ago2_D	15929968	3387090

Table 4.22: ***Aligned and unique reads from mESC as determined by samtools.***
The reads from the different samples were analysed with samtools rmdup, which compares the outer coordinates of all reads and removes duplicates. Unfortunately, 72 to 86% of the paired reads from Dam_Ago2 samples were duplicates mapping to exactly the same position.

Then, the ratio of uniquely to all aligned reads was plotted per chromosome (Figure 4.25). In contrast to the MRC-5 cells, where 10% of the duplicate reads derived from the human genome (Figure 4.19), several mouse chromosomes gave rise to high duplication rates comprising chromosome 1, 2, 9, 12 and 14 (Figure 4.25). If the set of mouse chromosomes generating high amounts of PCR-duplicates in

DAMID experiments is a specific effect or caused by non-specific Dam_Ago2 association, needs to be examined. In depth analysis of read distribution and the underlying DNA sequence may help to explain this finding.

Regarding mitochondrial DNA, there were about ~80 % duplicates observed for the Dam control in mESCs compared to ~65 % of duplicate reads identified in MRC-5 cells. The numbers of duplicate reads for Dam-Ago2 in MRC-5 (see Figure 4.19) and Dam_Ago2 in mESCs (Figure 4.25) ranged for both experiments at roughly 90 %. Here, it seems that the generation of stable cell lines led to a slightly higher level of duplicate reads for the Dam control.

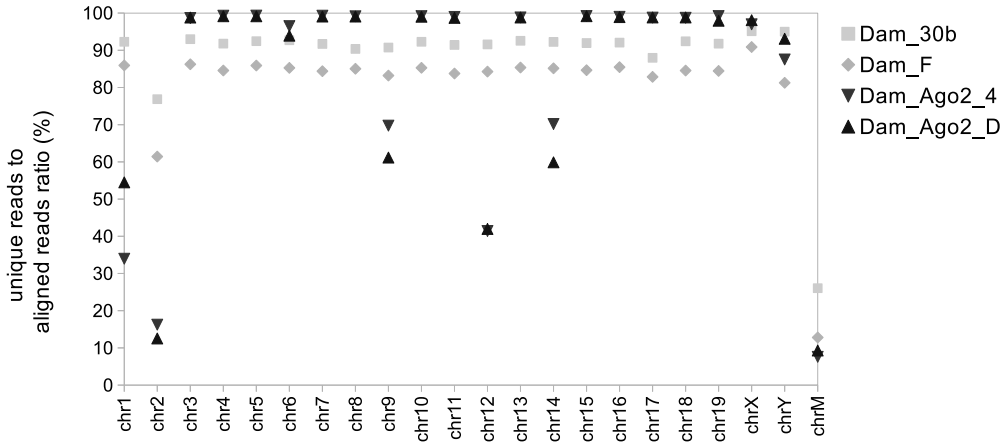


Figure 4.25: ***Analysis of PCR-duplicates.***
The number of unique reads was plotted against the number of aligned reads (unique + duplicate reads). For the mouse genome the chromosomes 1, 2, 9, 12 and 14 contained high rates of duplicates, while the other chromosomes showed rather low duplication rates. For the mitochondrial genome nearly 90% of the reads from the Dam_Ago libraries were removed by the "rmdup" option of samtools, while ~80 % were removed for the Dam control.

In order to assess the read coverage of the different mouse chromosomes, additionally, reads per base pair were calculated for each chromosome and plotted for all aligned and only unique read pairs (Figure 4.26). In this analysis the read coverage for chromosomes 1, 2, 9, 12, 14 and M was elevated for the data set comprising duplicate read pairs. In the unique dataset only chromosomes 2, 9, 14 and M showed an increased read coverage.

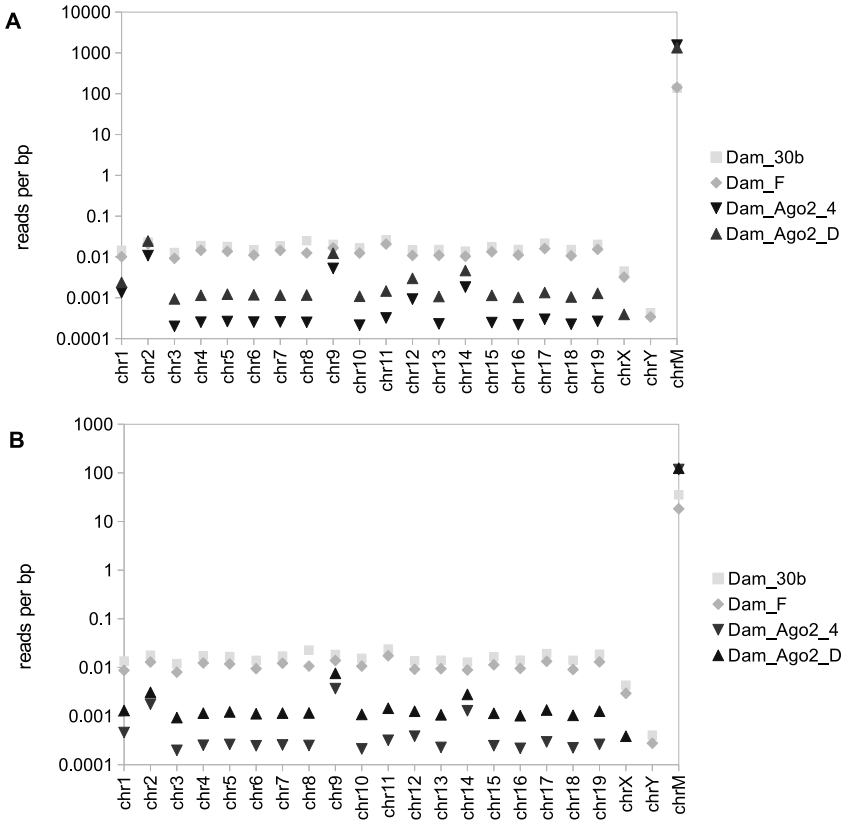


Figure 4.26: ***Distribution of reads per chromosome in mESCs.***
A*) All aligned reads were plotted per chromosome (bp).* ***B****) All unique reads were plotted per chromosome (bp). Comparing both analyses, it became apparent, that despite the high duplication level of the data, mitochondria derived sequences were strongly enriched for samples transduced with Dam_Ago2.*

In agreement with the data from MRC-5 high enrichment for mitochondrial sequences for Dam_Ago2 containing mESCs over the Dam controls was observed for aligned reads (~10 fold) and for unique reads (~5 fold) (Figure 4.26). This could hint to a potential mitochondrial localisation of Ago2 proteins which is assessed in Section 4.5. If the duplicated reads derive from certain loci on the specific chromosomes or from random positions needs to be further analysed.
To further analyse if duplicate reads are caused by the DAMID method and are therefore specific or if these reads are introduced by the PCR amplification step during library preparation, both data sets were used for downstream analysis. Based on the data sets comprising all aligned reads or only unique reads, reads overlapping a GATC sequence were counted applying an R script. These count tables were then loaded into the DESeq package. Here, only those locations were

selected in which all replicates yield a value above zero. Locations with zero counts can derive from a lack of enrichment or an escape from the NGS reaction and thus no conclusions can be drawn from these count values.
The correlation of the different samples was first examined by scatter plots drawn from log2 count values of the different samples (Figure 4.27). Here, the Dam controls showed some correlation, as did the Dam_Ago2 treated replicates.

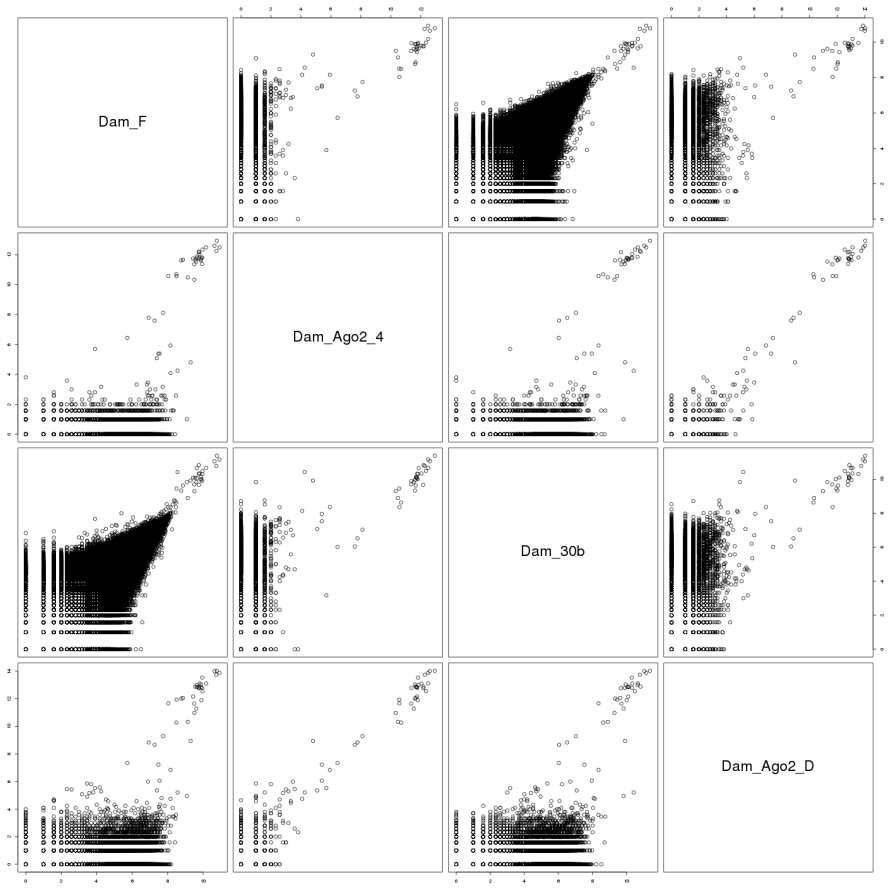


Figure 4.27: ***The correlation plot of control and Dam_Ago2 treated samples.***
The correlation of Dam and Dam-Ago2 treated samples was assessed by performing scatter plots based on log2 count values for all samples.

But no strong correlation was found between DamAgo2 expressing cells and controls when low to intermediate counts are examined. For very high counts, however, all samples showed clustering. These counts correspond to reads mapping to mitochondrial DNA. Here, the Dam controls display the same association pattern as Dam-Ago2, only the read numbers are increased by a factor of 10. This hints to a rather random association with mitochondrial DNA which only dif-

fers by the amount of methylation observed for Dam_Ago2 compared to Dam controls. As the higher methylation level could be caused by a mitochondrial localisation of Dam_Ago2, further experiments on a potential role of Ago2 in mitochondria were conducted (Section 4.5).

4.4.4.3 Analysis of Aligned Reads with DESeq

Based on the read count values form the DAMID-Seq experiment in mESCs containing also duplicate reads, an analysis with the DESeq package was performed [Anders and Huber, 2010]. After loading raw read counts into the software (Table 4.24, columns with sample names), DESeq performed library normalisation, which was modified by supplying the library sizes from the NGS run. Then, the average count values of all samples (baseMean), the average count value for replicates (baseMeanA, baseMeanB) of aligned reads and enrichment levels (foldChange, log2FoldChange) could be estimated for the replicates of treated and control samples (Table 4.24). Based on the count values and their variance among replicates of treated and control samples, p-values and false discovery rates (padj) were determined (Table 4.24). Using a padj value of 10 % as threshold, 470 peaks showing enrichment for Dam_Ago2 at the chromatin were identified. The resulting peaks were annotated using the Bioconductor package Biomart on the Ensembl database (gene_ids, GO_domain, source) and repeatmasker (http://repeatmasker.org, repclass) and the top 40 peaks ranked by their log2FoldChange can be found in the following tables. The first table shows the genomic annotation and classification (Table 4.23), while the second one holds information on the peak statistics (Table 4.24). Here, peaks with high counts were observed for many regions in the mitochondrial DNA, but also for some loci on chromosome 1, 2, 9, 10 and 18. This overlaps in part with the chromosomes showing especially high duplication rates (chromosomes 1, 2, 9, 12, 14 and M). For the chromosomes 10, 12, 14 or 18, however, the read counts of the significant loci, did not correlate with the rate of duplicate reads.
Therefore, it seems that the duplication rate does not correlate with the specific association of Dam_Ago2 proteins with chromatin. From this, we can conclude that at least a fraction of duplicate reads derive from over-amplification during the PCR step of the library preparation.

Table 4.23: ***Ago2 enriched sites in mESC based on all aligned reads.***

Nr	chrom	start	end	log2FoldChange	pval	padj	repclass	ensembl_gene_id	external_gene_id	GO_domain	source
1	chrM	2438	2441	4.27E+00	7.18E-10	1.28E-07	NA	ENSMUSG00000064339	mt-Rnr2	NA	exon
2	chr10	85026105	85026108	4.18E+00	4.67E-03	2.31E-02	NA	NA	NA	NA	intergenic
3	chr2	22588241	22588244	3.74E+00	1.15E-03	9.32E-03	NA	ENSMUSG00000083563	Gm13340	sensory perception of sound	exon
4	chr1	24615844	24615847	3.73E+00	1.17E-03	9.32E-03	NA	NA	NA	NA	intergenic
5	chrM	6731	6734	3.73E+00	1.17E-03	9.32E-03	NA	ENSMUSG00000064351	mt-Co1	aerobic respiration	exon
6	chrM	3223	3226	3.61E+00	3.37E-10	7.98E-08	NA	ENSMUSG00000064341	mt-Nd1	oxidation-reduction process	exon
7	chrM	6697	6700	3.52E+00	6.25E-07	1.27E-05	NA	ENSMUSG00000064351	mt-Co1	aerobic respiration	exon
8	chrM	13820	13823	3.51E+00	6.81E-09	4.33E-07	NA	ENSMUSG00000064368	mt-Nd6	oxidation-reduction process	exon
9	chrM	13598	13601	3.49E+00	7.32E-09	4.33E-07	NA	ENSMUSG00000064368	mt-Nd6	oxidation-reduction process	exon
10	chrM	11329	11332	3.49E+00	1.09E-09	1.56E-07	NA	ENSMUSG00000064363	mt-Nd4	mitochondrial electron transport, NADH to ubiquinone	exon
11	chr2	22587887	22587890	3.48E+00	1.60E-03	1.18E-02	NA	ENSMUSG00000083863	Gm13341	sensory perception of sound	exon
12	chrM	4276	4279	3.47E+00	3.60E-08	1.16E-06	NA	ENSMUSG00000064345	mt-Nd2	oxidation-reduction process	exon
13	chrM	3566	3569	3.45E+00	3.04E-09	2.33E-07	NA	ENSMUSG00000064341	mt-Nd1	oxidation-reduction process	exon
14	chrM	4065	4068	3.43E+00	3.04E-09	2.33E-07	NA	ENSMUSG00000064345	mt-Nd2	oxidation-reduction process	exon
15	chrM	6328	6331	3.43E+00	3.09E-07	7.08E-06	NA	ENSMUSG00000064351	mt-Co1	aerobic respiration	exon
16	chrM	1889	1892	3.31E+00	3.28E-09	2.33E-07	NA	ENSMUSG00000064339	mt-Rnr2	NA	exon
17	chrM	5988	5991	3.29E+00	1.70E-09	1.89E-07	NA	ENSMUSG00000064351	mt-Co1	aerobic respiration	exon
18	chrM	13925	13928	3.28E+00	2.32E-08	8.70E-07	NA	ENSMUSG00000064368	mt-Nd6	oxidation-reduction process	exon
19	chr2	22588275	22588278	3.24E+00	2.19E-02	5.36E-02	NA	ENSMUSG00000083563	Gm13340	sensory perception of sound	exon
20	chrM	3597	3600	3.24E+00	6.70E-08	2.07E-06	NA	ENSMUSG00000064341	mt-Nd1	oxidation-reduction process	exon
21	chrM	15336	15339	3.23E+00	1.86E-09	1.89E-07	NA	ENSMUSG00000064371	mt-Tt	NA	exon
22	chr18	44182952	44182955	3.21E+00	2.48E-02	5.66E-02	LINE/L1	NA	NA	NA	intergenic
23	chrM	14458	14461	3.21E+00	8.52E-06	1.41E-04	NA	ENSMUSG00000064370	mt-Cytb	respiratory electron transport chain	exon
24	chrM	5884	5887	3.14E+00	7.16E-08	2.12E-06	NA	ENSMUSG00000064351	mt-Co1	aerobic respiration	exon
25	chrM	2350	2353	3.12E+00	2.87E-08	1.00E-06	NA	ENSMUSG00000064339	mt-Rnr2	NA	exon
26	chrM	10735	10738	3.12E+00	1.78E-07	4.53E-06	NA	ENSMUSG00000064363	mt-Nd4	mitochondrial electron transport, NADH to ubiquinone	exon
27	chrM	16180	16183	3.11E+00	8.20E-07	1.57E-05	NA	NA	NA	NA	intergenic
28	chrM	14755	14758	3.09E+00	6.44E-05	8.65E-04	NA	ENSMUSG00000064370	mt-Cytb	respiratory electron transport chain	exon
29	chr18	44181674	44181677	3.09E+00	2.25E-04	2.58E-03	LINE/L1	NA	NA	NA	intergenic
30	chrM	11174	11177	3.01E+00	1.94E-05	2.93E-04	NA	ENSMUSG00000064363	mt-Nd4	mitochondrial electron transport, NADH to ubiquinone	exon
31	chrM	3103	3106	2.99E+00	1.89E-06	3.44E-05	NA	ENSMUSG00000064341	mt-Nd1	oxidation-reduction process	exon
32	chr1	24613190	24613193	2.96E+00	1.79E-06	3.35E-05	NA	NA	NA	NA	intergenic
33	chrM	12034	12037	2.90E+00	1.66E-07	4.37E-06	NA	ENSMUSG00000064367	mt-Nd5	ATP synthesis coupled electron transport	exon
34	chrM	13406	13409	2.87E+00	3.82E-07	8.48E-06	NA	ENSMUSG00000064367	mt-Nd5	ATP synthesis coupled electron transport	exon
35	chrM	2505	2508	2.81E+00	6.02E-05	8.23E-04	NA	ENSMUSG00000064339	mt-Rnr2	NA	exon
36	chr11	90538653	90538656	2.69E+00	3.16E-03	1.81E-02	NA	ENSMUSG00000081043	Gm11512	protein targeting	exon
37	chr1	24611839	24611842	2.62E+00	7.44E-03	2.89E-02	NA	NA	NA	NA	intergenic
38	chr1	24613337	24613340	2.49E+00	2.84E-03	1.68E-02	NA	NA	NA	NA	intergenic
39	chr9	3007432	3007435	1.47E+00	4.78E-02	8.55E-02	Satellite	ENSMUSG00000097312	Gm26870	NA	intron
40	chr9	3000001	3000004	1.36E+00	3.80E-02	7.31E-02	Satellite	NA	NA	NA	intergenic

Table 4.24: ***Ago2 enriched sites in mESC statistics for all aligned reads.***

Nr	chrom	start	end	Dam_F	Dam_30b	Dam_Ago2_4	Dam_Ago2_D	baseMean	baseMeanA	baseMeanB	foldChange	log2FoldChange	pval	padj
1	chrM	2438	2441	1224	1426	25856	27087	1.37E+04	1.35E+03	2.61E+04	1.93E+01	4.27E+00	7.18E-10	1.28E-07
2	chr10	85026105	85026108	130.6	873.2	8966.4	8391.4	4.54E+03	4.75E+02	8.60E+03	1.81E+01	4.18E+00	4.67E-03	2.31E-02
3	chr2	22588241	22588244	41.4	148	696	1856.8	6.55E+02	9.15E+01	1.22E+03	1.33E+01	3.74E+00	1.15E-03	9.32E-03
4	chr1	24615844	24615847	41.4	148.2	695.2	1857.4	6.55E+02	9.16E+01	1.22E+03	1.33E+01	3.73E+00	1.17E-03	9.32E-03
5	chrM	6731	6734	41.4	148.2	695.2	1857.4	6.55E+02	9.16E+01	1.22E+03	1.33E+01	3.73E+00	1.17E-03	9.32E-03
6	chrM	3223	3226	16771	17937	200587	244141	1.18E+05	1.78E+04	2.18E+05	1.22E+01	3.61E+00	3.37E-10	7.98E-08
7	chrM	6697	6700	47706	31835.4	612330.4	337025.4	2.60E+05	4.19E+04	4.79E+05	1.14E+01	3.52E+00	6.25E-07	1.27E-05
8	chrM	13820	13823	7672	9610	97689	105389	5.44E+04	8.79E+03	1.00E+05	1.14E+01	3.51E+00	6.81E-09	4.33E-07
9	chrM	13598	13601	8703	9028	113119.2	93244	5.59E+04	9.11E+03	1.03E+05	1.13E+01	3.49E+00	7.32E-09	4.33E-07
10	chrM	11329	11332	19884	17685	233367	206640	1.19E+05	1.95E+04	2.18E+05	1.12E+01	3.49E+00	1.09E-09	1.56E-07
11	chr2	22587887	22587890	22	10	200	187	1.04E+02	1.72E+01	1.92E+02	1.12E+01	3.48E+00	1.60E-03	1.18E-02
12	chrM	4276	4279	15710	16260	250679	108488	9.95E+04	1.64E+04	1.83E+05	1.11E+01	3.47E+00	3.60E-08	1.16E-06
13	chrM	3566	3569	14060	15772	162146	176710	9.11E+04	1.53E+04	1.67E+05	1.09E+01	3.45E+00	3.04E-09	2.33E-07
14	chrM	4065	4068	32108	25086	389632	251315	1.76E+05	2.99E+04	3.22E+05	1.08E+01	3.43E+00	3.04E-09	2.33E-07
15	chrM	6328	6331	95781	64474	1095874	713835	4.96E+05	8.43E+04	9.08E+05	1.08E+01	3.43E+00	3.09E-07	7.08E-06
16	chrM	1889	1892	35100	38234	436056	307907	2.05E+05	3.76E+04	3.72E+05	9.90E+00	3.31E+00	3.28E-09	2.33E-07
17	chrM	5988	5991	40701.6	31198.8	425182	311661.4	2.03E+05	3.76E+04	3.68E+05	9.80E+00	3.29E+00	1.70E-09	1.89E-07
18	chrM	13925	13928	11792.6	12541.8	109853.8	137568.2	6.69E+04	1.25E+04	1.21E+05	9.71E+00	3.28E+00	2.32E-08	8.70E-07
19	chr2	22588275	22588278	17	82.4	225.2	722.4	2.49E+02	4.75E+01	4.50E+02	9.48E+00	3.24E+00	2.19E-02	5.36E-02
20	chrM	3597	3600	12886	10070	134876	91129	6.26E+04	1.20E+04	1.13E+05	9.45E+00	3.24E+00	6.70E-08	2.07E-06
21	chrM	15336	15339	50235.6	45371.6	534882.6	394532.6	2.57E+05	4.95E+04	4.64E+05	9.38E+00	3.23E+00	1.86E-09	1.89E-07
22	chr18	44182952	44182955	17	84.4	223.4	722.4	2.49E+02	4.84E+01	4.49E+02	9.28E+00	3.21E+00	2.48E-02	5.66E-02
23	chrM	14458	14461	93136	53359	792782.2	649383.4	3.98E+05	7.77E+04	7.18E+05	9.24E+00	3.21E+00	8.52E-06	1.41E-04
24	chrM	5884	5887	13249.8	19237	153440.4	137624.8	8.04E+04	1.64E+04	1.44E+05	8.81E+00	3.14E+00	7.16E-08	2.12E-06
25	chrM	2350	2353	23043	22393	203863	209969	1.14E+05	2.34E+04	2.04E+05	8.71E+00	3.12E+00	2.87E-08	1.00E-06
26	chrM	10735	10738	76683.2	53306.8	673335.6	516423.4	3.31E+05	6.83E+04	5.93E+05	8.69E+00	3.12E+00	1.78E-07	4.53E-06
27	chrM	16180	16183	14777	13287	168132	80166	7.02E+04	1.45E+04	1.26E+05	8.66E+00	3.11E+00	8.20E-07	1.57E-05
28	chrM	14755	14758	132091.2	71064	1074543.8	768549	5.15E+05	1.08E+05	9.22E+05	8.53E+00	3.09E+00	6.44E-05	8.65E-04
29	chr18	44181674	44181677	128.2	143.8	1133.6	1268.4	6.61E+02	1.39E+02	1.18E+03	8.49E+00	3.09E+00	2.25E-04	2.58E-03
30	chrM	11174	11177	143310	86989	1152699	813508	5.53E+05	1.22E+05	9.84E+05	8.08E+00	3.01E+00	1.94E-05	2.93E-04
31	chrM	3103	3106	38963	25803	261224	289078	1.52E+05	3.41E+04	2.71E+05	7.94E+00	2.99E+00	1.89E-06	3.44E-05
32	chr1	24613190	24613193	4683	10237	58347	57429	3.23E+04	7.37E+03	5.72E+04	7.76E+00	2.96E+00	1.79E-06	3.35E-05
33	chrM	12034	12037	28236	29902	217302	233667	1.26E+05	2.99E+04	2.22E+05	7.44E+00	2.90E+00	1.66E-07	4.37E-06
34	chrM	13406	13409	20319	24318	168474.2	168816	9.47E+04	2.28E+04	1.67E+05	7.32E+00	2.87E+00	3.82E-07	8.48E-06
35	chrM	2505	2508	24140	14274	144739	143281	8.13E+04	2.03E+04	1.42E+05	7.00E+00	2.81E+00	6.02E-05	8.23E-04
36	chr11	90538653	90538656	2.6	2	7.2	25.6	8.97E+00	2.40E+00	1.55E+01	6.47E+00	2.69E+00	3.16E-03	1.81E-02
37	chr1	24611839	24611842	1.6	1.6	14.4	5.4	5.88E+00	1.65E+00	1.01E+01	6.13E+00	2.62E+00	7.44E-03	2.89E-02
38	chr1	24613337	24613340	141	213	835	1216	5.89E+02	1.78E+02	1.00E+03	5.61E+00	2.49E+00	2.84E-03	1.68E-02
39	chr9	3007432	3007435	669.6	448.8	1071.6	2314.2	1.11E+03	5.89E+02	1.63E+03	2.77E+00	1.47E+00	4.78E-02	8.55E-02
40	chr9	3000001	3000004	2613	1938	4073.4	8651.6	4.25E+03	2.38E+03	6.12E+03	2.57E+00	1.36E+00	3.80E-02	7.31E-02

Given that we observe some distortion of the read numbers due to over-amplification, the peaks identified based on aligned reads including duplicate reads are not reliable. Therefore, no further analysis is performed on the data set containing duplicate reads. A more conservative approach is to use only unique reads for peak calling, which was performed in the following part.

4.4.4.4 Analysis of Unique Reads with DESeq

A second analysis of only the unique reads acquired from the DAMID experiment in mESCs was performed. Applying the DESeq package on raw reads (Table 4.26, columns with sample names), corrected for library size, the average count values of all samples (baseMean), the average count value for replicates (baseMeanA, baseMeanB) of unique reads and enrichment levels (foldChange, log2FoldChange) could be estimated for the replicates of treated and control samples (Table 4.26). Here, it is very important to not use the default library size normalisation estimated from the reads in the count table but to estimate the library size from the NGS run, because the input to the DAMID experiment as well as for the library preparation were kept constant. Again, mean counts were calculated and compared. Also p-values (pval) and false discovery rates (padj), which correct the estimated p-values for testing of multiple features, were determined (Table 4.26). The loci were annotated with Biomart and repeatmasker as described for the previous tables. The first of the following tables shows the genomic annotation and classification (Table 4.25), while the second one holds information on the peak statistics (Table 4.26). However, only 24 peaks could be identified with a p-value ($p<0.05$) and a false discovery rate lower than 10 %.
The genomic sites Dam_Ago2 bound to comprise mitochondrial DNA, LINE L1, satellites and exonic sequences of several genes. The exonic sequences are part of four different predicted genes (Gm13340, GM13441, Gm11512, Gm13341), which are classified by GO term analysis into sensory perception of sound (GM13340 & GM13441), protein targeting (GM11512) and ATP catabolic processes (GM13341). The satellite peaks map all to a region on chromosome 9 (3000001-3013073), four of the five peaks overlap with Gm26870, a predicted lincRNA. The LINE L1 element, which Dam_Ago2 is associated with, is called L1Md_T. It has a length of 5191 bp and is localised on chr18:44177905-44183095. These regions were reported to our collaborator C. Ciaudo (ETH Zürich) for independent validation in mESCs, which is currently performed.

Table 4.25: ***Ago2 enriched sites in mESC based on unique reads.***

Nr	chrom	start	end	log2FoldChange	pval	padj	repclass	ensembl_gene_id	external_gene_id	GO_domain	source
1	chrM	2438	2438	3.3073458017	0.0054134897	0.0164865368	NA	ENSMUSG00000064339	mt-Rnr2	NA	exon
2	chrM	2505	2505	3.0613757471	0.0034402228	0.0115247463	NA	ENSMUSG00000064339	mt-Rnr2	NA	exon
3	chrM	3597	3597	3.0296885608	0.0006524988	0.0027323388	NA	ENSMUSG00000064341	mt-Nd1	oxidation-reduction process	exon
4	chrM	11329	11329	2.876310492	0.040324147	0.0656842451	NA	ENSMUSG00000064363	mt-Nd4	mitochondrial electron transport, NADH to ubiquinone	exon
5	chrM	13598	13598	2.8675794166	0.0100176135	0.0248585225	NA	ENSMUSG00000064368	mt-Nd6	oxidation-reduction process	exon
6	chrM	3566	3566	2.8560988746	0.0013735068	0.0054132328	NA	ENSMUSG00000064341	mt-Nd1	oxidation-reduction process	exon
7	chrM	13925	13925	2.8489813057	0.0215024087	0.0398866475	NA	ENSMUSG00000064368	mt-Nd6	oxidation-reduction process	exon
8	chrM	13820	13820	2.8230815581	0.0116974729	0.0273620746	NA	ENSMUSG00000064368	mt-Nd6	oxidation-reduction process	exon
9	chrM	5988	5988	2.6682318105	0.0239784234	0.0422777466	NA	ENSMUSG00000064351	mt-Co1	oxidative phosphorylation	exon
10	chr1	24613190	24613190	2.6515354342	0.0471520293	0.0692870188	NA	ENSMUSG00000101111	RP23-8J15.4	NA	exon
11	chrM	2350	2350	2.6510286691	0.0197644815	0.0389476547	NA	ENSMUSG00000064339	mt-Rnr2	NA	exon
12	chrM	13406	13406	2.5938495535	0.0213412317	0.0398866475	NA	ENSMUSG00000064367	mt-Nd5	ATP synthesis coupled electron transport	exon
13	chrM	4065	4065	2.5911534891	0.011843286	0.0273620746	NA	ENSMUSG00000064345	mt-Nd2	mitochondrial electron transport, NADH to ubiquinone	exon
14	chrM	6697	6697	2.4704771708	0.0160376408	0.0346619979	NA	ENSMUSG00000064351	mt-Co1	oxidative phosphorylation	exon
15	chrM	4276	4276	2.4651582423	0.0220269546	0.0398866475	NA	ENSMUSG00000064345	mt-Nd2	mitochondrial electron transport, NADH to ubiquinone	exon
16	chrM	5884	5884	2.4436637543	0.0070604973	0.0199585837	NA	ENSMUSG00000064351	mt-Co1	oxidative phosphorylation	exon
17	chr2	22587887	22587887	2.2918917263	0.0095930891	0.0247206527	NA	ENSMUSG00000083863	Gm13341	ATP catabolic process	exon
18	chrM	16180	16180	2.1981519453	0.0044326206	0.0141421704	NA	NA	NA	NA	intergenic
19	chrM	12034	12034	2.1794876486	0.0088819664	0.02380367	NA	ENSMUSG00000064367	mt-Nd5	ATP synthesis coupled electron transport	exon
20	chrM	15336	15336	2.1583300201	0.0141714839	0.0316496474	NA	ENSMUSG00000064371	mt-Tt	NA	exon
21	chr9	3007432	3007432	1.9503179249	0.0411751984	0.0656842451	Satellite	ENSMUSG00000097312	Gm26870	NA	intron
22	chrM	1889	1889	1.9497288625	0.0168438229	0.0352667543	NA	ENSMUSG00000064339	mt-Rnr2	NA	exon
23	chrM	10735	10735	1.6534823973	0.0297983612	0.0511920565	NA	ENSMUSG00000064363	mt-Nd4	mitochondrial electron transport, NADH to ubiquinone	exon
24	chr9	3000001	3000001	1.5636975779	0.0432109498	0.0673286891	NA	NA	NA	NA	intergenic

Table 4.26: ***Ago2 enriched sites in mESC statistics on unique reads.***

Nr	chrom	start	end	Dam_F	Dam_30b	Dam_Ago2_4	Dam_Ago2_D	baseMean	baseMeanA	baseMeanB	foldChange	log2FoldChange	pval	padj
1	chrM	2438	2438	263	329	1523	3285	1478.2427422131	271.2513265159	2685.2341579102	9.8994323545	3.3073458017	0.0054134897	0.0164865368
2	chrM	2505	2505	870	819	4584	7051	3617.4492496124	773.9777661209	6460.9207331039	8.3476826027	3.0613757471	0.0034402228	0.0115247463
3	chrM	3597	3597	436	646	3144	4166	2272.0334764767	495.734389704	4048.3325632494	8.1663339226	3.0296885608	0.0006524988	0.0027323388
4	chrM	11329	11329	940	996	3151	8473	3700.5004968818	887.1230823449	6513.8779114186	7.3426991599	2.876310492	0.040324147	0.0656842451
5	chrM	13598	13598	891	1296	4240	8861	4157.5528958472	1002.0136261663	7313.0921655281	7.2983959245	2.8675794166	0.0100176135	0.0248585225
6	chrM	3566	3566	465	791	3247	4277	2370.9097668273	575.4252730415	4166.3942606131	7.2405479144	2.8560988746	0.0013735068	0.0054132328
7	chrM	13925	13925	749	1096	3245	7645	3467.8817859897	845.318204226	6090.4453677534	7.2049144775	2.8489813057	0.0215024087	0.0398866475
8	chrM	13820	13820	797	1069	3440	7394	3452.6714050943	854.9683276423	6050.3744825462	7.0767235311	2.8230815581	0.0116974729	0.0273620746
9	chrM	5988	5988	1154	1639	5766	8882	4706.962942414	1279.6751713954	8134.2507134275	6.3564964729	2.6682318105	0.0239784234	0.0422777466
10	chr1	24613190	24613190	362	703	1516	3957	1776.7635621075	487.8969098691	3065.6302143459	6.2833564885	2.6515354342	0.0471520293	0.0692870188
11	chrM	2350	2350	947	1138	3328	7411	3478.0275786609	955.3511986846	6000.7039586371	6.2811497666	2.6510286691	0.0197644815	0.0389476547
12	chrM	13406	13406	800	1613	3727	8218	3889.4856236782	1105.4269049287	6673.5443424277	6.0370742857	2.5938495535	0.0213412317	0.0398866475
13	chrM	4065	4065	863	1286	3487	7145	3458.7822760067	984.5941638022	5932.9703882111	6.025802921	2.5911534891	0.011843286	0.0273620746
14	chrM	6697	6697	1001	1442	5197	6032	3661.4105465055	1119.3087933996	6203.5122996113	5.5422706729	2.4704771708	0.0160376408	0.0346619979
15	chrM	4276	4276	867	1825	4692	7561	4021.4598137739	1233.2219749308	6809.6976526169	5.5218750485	2.4651582423	0.0220269546	0.0398866475
16	chrM	5884	5884	691	1100	3432	4626	2642.2626600335	820.5510305267	4463.9742895403	5.440215323	2.4436637543	0.0070604973	0.0199585837
17	chr2	22587887	22587887	15	9	52	46	32.4328482271	10.9998199112	53.865876543	4.8969780394	2.2918917263	0.0095930891	0.0247206527
18	chrM	16180	16180	359	396	1653	1244	966.7536206209	345.9541779407	1587.553063301	4.5889113777	2.1981519453	0.0044326206	0.0141421704
19	chrM	12034	12034	726	832	2631	3217	1973.8852839497	713.8920478366	3233.8785200628	4.5299265202	2.1794876486	0.0088819664	0.02380367
20	chrM	15336	15336	959	1031	3561	3818	2491.1970574167	911.8619848251	4070.5321300083	4.4639783188	2.1583300201	0.0141714839	0.0316496474
21	chr9	3007432	3007432	122	92	220	459	238.5433144493	98.0732099306	379.013418968	3.8645968582	1.9503179249	0.0411751984	0.0656842451
22	chrM	1889	1889	786	914	2966	2513	1894.0328666526	778.9534746816	3009.1122586236	3.8630192385	1.9497288625	0.0168438229	0.0352667543
23	chrM	10735	10735	561	478	1430	1294	987.0120791476	476.1364736814	1497.8876846137	3.1459209017	1.6534823973	0.0297983612	0.0511920565
24	chr9	3000001	3000001	729	630	1288	2026	1231.8867148546	622.7775422663	1840.9958874428	2.9561051298	1.5636975779	0.0432109498	0.0673286891

The strong enrichment of Dam_Ago2 at mitochondrial regions suggests, that the band pattern found for Dam_Ago2 samples after DAMID, most likely results from mitochondrial DNA. However, based on these observations the question arises, how Ago proteins enter mitochondria. Therefore, the mitochondrial localisation and import was assessed for Ago proteins (see Section 4.5).

4.5 Analysis of Potential Mitochondrial Localisation of Ago Proteins

Mitochondria are organelles surrounded by a double membrane, which reside in the cytoplasm and potentially derived form an endosymbiontic event early in the evolution of eukaryotes. In HeLa cells the estimated number of mitochondria ranges from 380 to 880 per cell depending on the cell cycle stage [Posakony *et al.*, 1977]. They serve the cells by converting pyruvate, the product of the anaerobic glycolysis, by means of the citric acid cycle and oxidative phosphorylation into high amounts of ATP. In addition, mitochondria are involved in the metabolism of amino acids, lipids, haem and iron-sulphur clusters and in regulation of programmed cell death [Schmidt *et al.*, 2010].

4.5.1 Introduction to Mitochondria

The mitochondrial genome contains 16569 *nt* and encodes for 13 protein coding genes and 22 tRNAs [Alberts *et al.*, 2002]. In contrast to its small genome, mitochondria import about 1000 proteins from the cytoplasm by various import mechanisms [Schmidt *et al.*, 2010]. But also the destination of the imported proteins differ and comprise the outer membrane, the inter membrane space, the inner membrane and its inner lumen called the mitochondrial matrix.

Protein Import Mechanisms

The main import route into mitochondria is via the translocase of the outer membrane complex (TOM). In order to get transported across the double membrane proteins contain presequences, which are proteolytically cleaved of after import, or internal targeting signals [Schmidt *et al.*, 2010]. There are four main pathways and several minor routes into the organelle. The first requires a cleavable prese-

quence and transfers the protein from the TOM complex to the inner membrane TIM23 (translocase of the inner membrane, TIM) complex for lateral release into the inner membrane or matrix import. Many proteins of the inner membrane are imported through a combination of the TOM complex and TIM22, which comprises the second pathway. The third is called oxidative folding pathway and is mainly used by inter membrane space proteins. Following translocation by TOM, the proteins are delivered to the mitochondrial inter membrane space assembly complex. The last pathway is taken by outer membrane proteins which contain again internal import signals [Schmidt *et al.*, 2010]. For some proteins the import into mitochondria is also facilitated by localisation of their mRNA to the mitochondrial surface, mediated by a targeting sequence in the 3′-UTR. These proteins require a mitochondrial targeting signal peptide at the N-terminus and are co-translational imported into the organelle [Sylvestre *et al.*, 2003; Kaltimbacher *et al.*, 2006]. For mitochondrial import two sources of energy are described. The membrane potential gradient across the inner membrane (ψ) drives the translocation across the TIM complex and ATP which is utilised by Tom70 to unfold and import the preproteins across the outer membrane [Schmidt *et al.*, 2010]. Upon entering the mitochondrial matrix, the targeting sequence of most presequence containing proteins is proteolytically removed from the proteins by the matrix processing peptidase (MPP). In addition, also other peptidases with a smaller target range and different cleavage patterns were described [Schmidt *et al.*, 2010].

RNAi in Mitochondria

Recently, also the RNAi pathway was linked to the mitochondria. Several studies described miRNA or also Ago2 in mitochondria, which are summarized in Table 4.27.

MiRNA Detection in Lysate of Mitochondria The first report by Kren *et al.* [2009] identified 15 miRNAs associated with highly purified mitochondria from rat liver by microarray experiments. However, based only on sequence analysis the miRNAs were linked to the apoptosis pathway, however, lacking further independent validations of this proposed function [Kren *et al.*, 2009]. Subsequently, the hypothesis of miRNAs in mitochondria was supported by a row of publications which identified miRNAs by microarrays and NGS, but included also validation experiments like *in situ* hybridisation and qRT-qPCR in their analyses

miRNAs	Species	Methods	Ago localisation	Reference
miR-130a, miR-130b, miR-140, miR-320, miR-494, miR671	*Ratus norvegicus*, liver	Microarray		[Kren *et al.*, 2009]
miR-21, miR-130a, miR-130b, miR-140, miR-320, miR-494	*Ratus norvegicus*, liver	TaqMan-mRNA assay		[Kren *et al.*, 2009]
miR-365, pre-miR-302a, pre-let7b, miR-1974	*Homo sapiens*, myoblasts	*in situ* hybridisation		[Barrey *et al.*, 2011]
miR-720, miR-133b, miR-1974, miR-24, miR-133a, miR-125a-5p, miR-1979, miR-103, miR-125b, miR-103, miR-221, miR-23a (CT >33)	*Homo sapiens*, myoblasts	qRT-PCR		[Barrey *et al.*, 2011]
miR-705,miR-202-5p, miR134	*Mus musculus*, liver	Microarray, qRT-PCR, WB	yes	[Bian *et al.*, 2010]
miR-705, miR-122, miR-805, miR-690, miR-451,miR-7b, miR-26a	*Mus musculus*, liver	Microarray		[Bian *et al.*, 2010]
s miR-638, miR-1977, miR-1972,miR-1908, miR-1246, miR-513A-5P, miR-494, miR-1974 and miR-1275	*Homo sapiens*, HeLa	Mircoarray	yes	[Bandiera *et al.*, 2011]
miR-494, miR-1974 and miR-1275	*Homo sapiens*, HeLa	Mircoarray, qRT-PCR		[Bandiera *et al.*, 2011]
let7b, let7g, miR-107, miR-181a, miR-221 and miR-320a,	*Homo sapiens*, HEK293, HeLa	NGS, qRT-PCR	EGFP-Ago2	[Sripada *et al.*, 2012]
miR-320, miR-22, miR-883, several mouse and human miRNAs	*Ratus norvegicus*, heart myocytes	Microarray		[Das *et al.*, 2012]
miR-181-c (human)	*Ratus norvegicus*, heart myocytes	RIP, protein level (COX1), Microarray	yes	[Das *et al.*, 2012]
miR-1	*Homo sapiens*, C2C12; *Mus musculus*, heart	Protein level (ND1, COX1), Translation assay, CLIP-Seq, complementation assay, WB, mitoplast preparation	yes	[Zhang *et al.*, 2014]

Table 4.27: ***List of publications on miRNA and Ago localisation in mitochondria.*** *The table contains the detected miRNAs, the species or cell line the study was conducted in, the main methods used, if Ago2 localisation was tested and found and the reference.*

[Bian *et al.*, 2010; Barrey *et al.*, 2011; Sripada *et al.*, 2012; Das *et al.*, 2012].

Ago Localisation in Mitochondria The localisation of Ago2 was also examined by some of the previously cited studies. Thus, fractionation by differential centrifugation was used to successfully detect mitochondrial Ago2 [Bian *et al.*, 2010]. However, these preparations are often heavily contaminated by other organelles [Reinhart *et al.*, 1982]. In 2011 Bandiera *et al.* reported on Ago localisation in immunopurified mitochondria from HeLa cells. Also the mitochondrial localisation of EGFP-Ago2 was shown, however, only very limited overlap with their mitochondrial marker was detected by Sripada *et al.* [2012].

Functions of RNAi in Mitochondria Applying RNA isolated from mitochondria of rat hearts to Affymetrix transcript microarrays, Das *et al.* [2012] detected several miRNAs, among other they found enrichment of miR-181c. Using qRT-PCR and Ago2-RIP assays they could validate enrichment of miR-181c and association of miR181c and COX-1 mRNA with Ago2. COX-1 stands for Cytochrome c

oxidase I, which is a subunit of the respiratory complex IV and encoded by the mitochondrial DNA. Here, only protein levels of COX-1 but not mRNA levels were decreased by miR181c and resulted in an increase of O_2 consumption of transfected cells [Das *et al.*, 2012].

Another example of a functional miRNA-Ago2 complex in mitochondria was shown in mouse myoblasts (C2C12) cells. Analysing miR-1, which is specifically induced during myogenesis, Zhang *et al.* [2014] found mitochondrial enrichment of the miRNA in differentiated C2C12 (Myotubes) coinciding with the localisation of Ago2 to this compartment. Ago2 but not miR-1 was also present in undifferentiated C2C12. Unlike earlier studies, the localisation of Ago2 in mitochondria, was tightly controlled using appropriate antibodies (esp. for ER proteins) and preparation of so called mitoplasts, where the outer membrane of mitochondria is removed suggesting matrix localisation of the still detectable proteins. When they examined the effects on expression of mitochondrial encoded genes, they found an enhancement of the protein levels of ND1 (NADH dehydrogenase 1) and COX1 in radioactive translation assays upon miR-1 transfection in undifferentiated C2C12. In order to demonstrate target interaction, they used Ago2-CLIP assays, where they found enrichment of several mitochondrial encoded transcripts. They also tested their system in mouse embryonic fibroblasts (MEFs), where they observed rescue of the mitochondrial expression enhancement after transfection of miR-1 in combination with HA-Ago2 or a mitochondrial targeted versions of Ago2 in WT and Ago2 knockout MEFs. However, it still needs to be shown if Ago2 localisation to mitochondria is C2C12 specific and in other cells miR-1 dependent or a general mechanism observed also in other cell types.

In summary, all the above described studies suggest a mitochondrial localisation of miRNAs and Ago proteins based on mitochondria isolated by fractionation or immunoprecipitation. However, the import mechanism of Ago proteins or their effect on transcripts was not conclusively assessed.

In our own DAMID experiments, a ladder of certain highly enriched amplicons appeared in the Dam_Ago2 samples from mESCs (Figure 4.23) as well as for Dam-Ago proteins in MRC-5 cells (Figure 4.15). In contrast to this, control proteins showed a DNA smear for both cell lines, indicating a broad target range. Some of these bands resulted supposedly from mitochondrial DNA, which is supported by Sanger Sequencing of the respective bands and high read counts

for mitochondrial DNA, 10 to 20 fold above the level of control samples in the NGS data (Figure 4.26, 4.20). Given the lack of targets positive in Ago ChIP assays, we interpreted the methylation of mitochondrial DNA as a side effect of a mitochondrial-specific Dam-Ago localisation. Taking into account the set of publications reporting mitochondrial localised miRNAs and Ago proteins, we decided to closely examine a potential Ago localisation to mitochondria, which would be a requirement for a potential small RNA induced function in this compartment. As our lab had limited experience with the experimental methods specific for the field of mitochondrial research, the association of Ago with this organelle was further analysed in collaboration with the lab of Professor Dr. P. Rehling, namely Dr. S. Dennerlein (University of Göttingen, Germany).

4.5.2 Localisation of Argonaute in Different Compartments

In a first step, the localisation of Ago proteins in mitochondria was tested with a biochemical approach. HEK-293 cell lysates were fractionated by differential centrifugation in cytoplasm, nuclei and crude mitochondria.

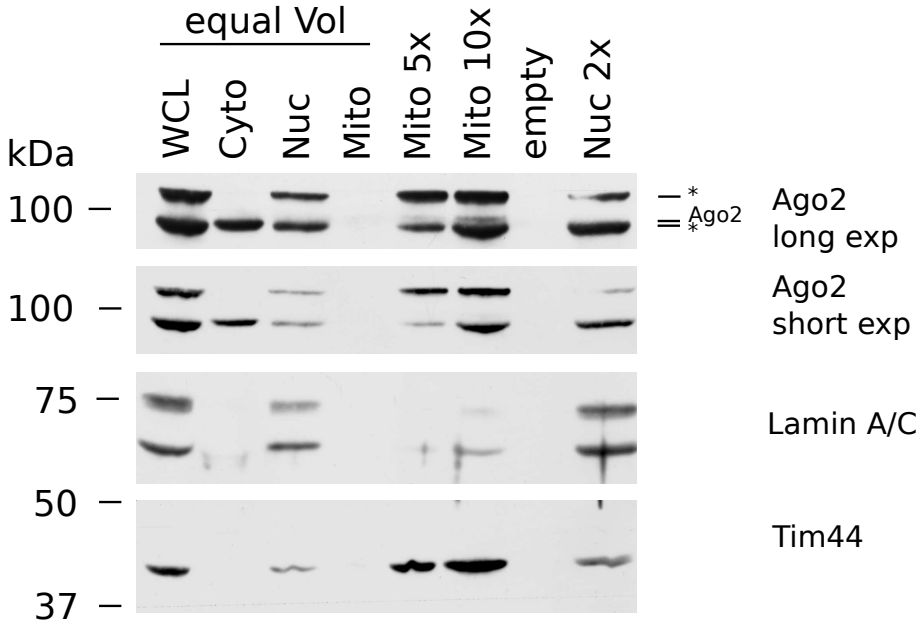


Figure 4.28: ***Localisation of Ago2 in different cellular compartments.***
Following lysis and differential centrifugation of HEK-293 cells, the different fractions enriched for cytoplasm, nuclei and crude mitochondria were analysed by WB. Ago2 yielded a distinct band in the cytoplasm, which split into several for nuclei and mitochondria*. The nuclear fraction was controlled by Lamin A/C and mitochondria by Tim44.*

The resulting lysates were analysed by Western Blot against different marker proteins (Figure 4.28). The nuclear fraction was controlled by LaminA/C, showing a clean cytoplasm and trace amounts of nuclei in the mitochondria containing fraction. The mitochondrial marker protein Tim44 resulted in a weak signal for nuclei and a high signal for mitochondrial fractions. The Ago2 protein was detected with the 11A9 antibody, which produces a distinct signal in the cytoplasm and several bands for the other compartments. Given that mitochondria acquired by differential centrifugation always co-purify with large amounts of the ER, the crude mitochondria purified by this assay were further analysed.

In summary, a strong signal for Ago2 was observed in the mitochondria containing fraction. This signal was clearly not derived from the nuclear contamination (see Figure 4.28, Lamin A/C control) but could originate from trace amounts of cytoplasm or contaminating ER.

4.5.3 Separation of Mitochondria from ER

In a next step, the contaminating ER was separated from mitochondria. Therefore, crude mitochondria from HEK-293 cells were applied to a Percoll gradient for density centrifugation (see Section 3.5.11). In the lower third of the gradient, mitochondria accumulated as distinct layer. The gradient was divided into three fractions. After several washes to completely remove the Percoll, which otherwise leads to aberrant migration behaviour on SDS-PAGE gels, the three fractions were analysed by WB (Figure 4.29, fractions F1-F3). This work was contributed by C. Richter-Dennerlein (collaboration, lab of Prof. Dr. Rehling, University of Göttingen, Germany).

The first fraction (Figure 4.29, F1) was highly enriched in ER vesicles, which can be appreciated by the strong signal of SEC61β, the central component of the ER protein translocation machinery. This protein is nearly completely absent from fractions F2 and F3. Mitochondria were analysed by the following marker proteins: MFN2, an integral membrane protein of the outer mitochondrial membrane, participating in fusion processes; MITRAC15, an integral membrane protein of the inner membrane, involved in complex I and IV biogenesis [Mick *et al.*, 2012]; and MRPS15, the mitochondrial ribosomal protein S15, a member of the 28S subunit. All of these proteins were detected already in F1, but were strongly enriched in F2. F3 seemed to be only enriched for MRPS15.

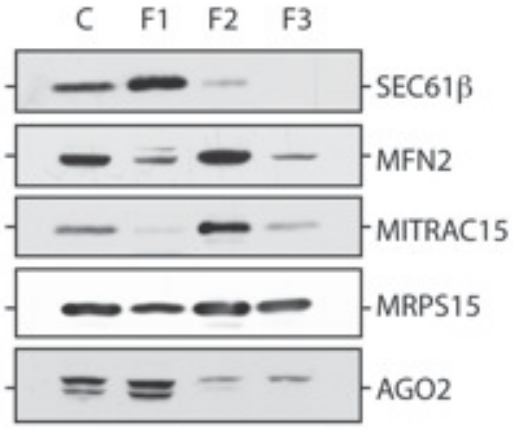


Figure 4.29: ***Purification of mitochondria from HEK-293 cells using Percoll.***
Crude mitochondria (C), isolated as above, were applied to density centrifugation on a Percoll gradient to separate mitochondria from contaminating ER. The gradient was divided in three fractions: F1: top layer to 12% Percoll transition; F2: interphase at 12% to 26 % Percoll; F3: interphase 26% to 40 % Percoll. Ago2 was found as double band in the fractions C and F1, which was enriched for ER. In the F2 and F3 fractions, which were highly enriched for mitochondria and nearly completely depleted for ER, the Ago2-specific 11A9 antibody detected a single slower migrating protein. The distribution of the different compartments was examined with the following antibodies: SEC61β: the main mediator of ER protein translocation; MFN2: an outer membrane component; MITRAC15: an integral inner membrane protein; MRPS15: the ribosomal protein S15, part of the S28 subunit of the mitochondrial ribosome.

Ago2 was found in all fractions, however, a double band was observed for crude mitochondria and F1. For F2 and F3 the double band detected by 11A9 was reduced to the protein with the higher molecular weight (Figure 4.29).
Therefore, it seems that the lower band results from an ER associated Ago2 species [Stalder *et al.*, 2013], whereas the slower migrating form is found also in mitochondria. As ER and mitochondria are surrounded by a membrane, the interaction of Ago proteins with membranes was further analysed.

4.5.4 Membrane Association of Ago

The membrane association of proteins can be probed by treating the respective membrane with carbonate buffers of ascending pH [Fujiki *et al.*, 1982]. At a pH value of 10.8 integral membrane proteins are retained as well as proteins associated via protein-protein interactions. Rising the pH to 11.5, however, destabilises protein-protein interactions and thus selects for integral membrane proteins.

In order to test the association of Ago with mitochondrial membranes, isolated mitochondria (Section 3.5.8) were exposed to different carbonate buffers (pH 10.8 and pH 11.5) and Trition X-100 containing buffer, which is a strong detergent solubilising membranes and proteins alike. Each sample was split in input (total), supernatant (SN), containing free proteins, and a 100.000 $\times g$ pellet (P), consisting of membrane-associated plus integral membrane proteins (pH 10.8) or integral membrane proteins only (pH 11.5). All fractions were analysed by SDS-PAGE and WB against Ago and various marker proteins (Figure 4.30).

Lysis by Triton X-100 was used as solubilisation control. Different mitochondrial proteins were used as controls for membrane association. Cox6A is part of the cytochrom c oxidase, which is the last complex in the mitochondrial electron transport chain residing in the inner membrane and catalysing the transfer of electrons to oxygen. As integral membrane protein, Cox6A was, as expected, found in the membrane fraction for both carbonate treatments (Figure 4.30, Pellet (P)). Tim44 is a part of the TIM complex, resides on the matrix side of the inner mitochondrial membrane and is bound to the membrane by protein-protein interactions with the integral membrane protein Tim23 [Rehling *et al.*, 2001]. In the carbonate extraction experiment Tim44 was bound to the membrane by protein-protein interactions and found in the pellet fraction at pH 10.8 (Figure 4.30, P at pH 10.8). However, the interaction of the protein with Tim23 was lost when the membranes were treated with carbonate buffer of pH 11.5 releasing Tim44 into the supernatant (Figure 4.30, SN at pH 11.5).

The marker protein for the ER is the lumenal protein ERp72 which is a protein disulfide isomerase-like protein and primarily functions as a chaperone [Prins and Michalak, 2011]. ERp72 seems to be membrane associated by protein-protein interactions, showing a pattern comparable to Tim44 (Figure 4.30, P at pH 10.8 and SN at pH 11.4). The Ago1 protein, detected by the 1C9 antibody, was not associated with the membrane fractions under all tested conditions and thus found consistently in SN fractions throughout the experiment (Figure 4.30). For Ago2 two bands were observed. The faster migrating form behaved similar to Ago1, can be regarded as soluble protein and is most likely Ago2 (Figure 4.30, SN at pH 10.5 and 11.8). The slower migrating protein detected by the Ago2-specific antibody (11A9) showed membrane association even under treatment with carbonate buffer pH 11.5 (Figure 4.30, P at pH 10.8 and pH 11.5).

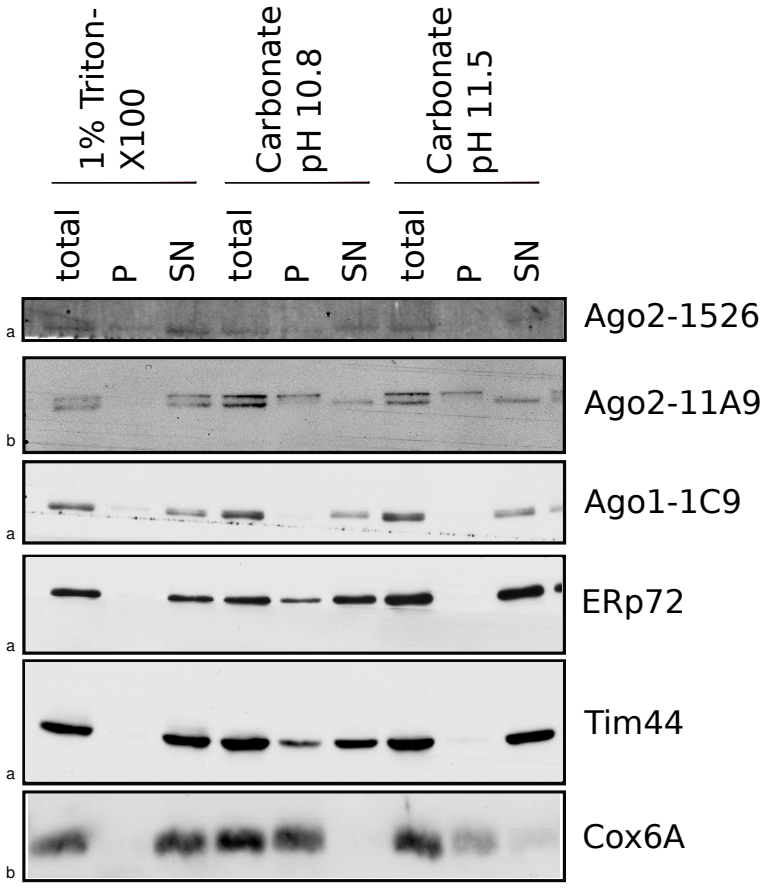


Figure 4.30: ***Analysis of mitochondrial membrane association of Ago proteins.***
Isolated mitochondria were lysed by different carbonate buffers (pH 10.8 and 11.5), which render membrane vesicles into sheets containing interacting proteins or only integral membrane proteins, respectively. Ago1 seems not to be bound to any membrane, which is also observed for the faster migrating form of Ago2. However, there is also a band with a slower migration rate detected by the Ago2 antibody (11A9). The respective isoform or protein behaves like an integral membrane protein, given its resistance to pH 11.5. When the WB membrane used for Ago1 detection was developed with a rabbit antibody specific for Ago2 (1526), only the faster migrating form was detected. ERp72: ER lumen; Tim44: Mitochondria inner membrane, bound by protein-protein interaction to integral membrane protein. Cox6A: member of the complex IV of the electron transport chain, integral membrane protein. One of two experiments is shown. Samples were loaded in equal amounts and order on a 10% (a) and 16% (b) SDS-PAGE gel. Further analysis by WB was performed in parallel.

Exposure of the Ago1 WB membrane (rat mAB) to an antibody specific for Ago2 (1526 rabbit pAB) (Figure 4.30, top blot) led only to detection of the faster migrating form of Ago2. Therefore, it is possible that the slower migrating band resembles an antibody cross-reactivity.
In conclusion, Ago1 and Ago2 (faster migrating band) were soluble proteins un-

der all tested conditions and were clearly not associated with any membrane. However, a slower migrating protein, detected by an Ago2-specific antibody (11A9), showed strong membrane interaction, which was comparable to Cox6A. This protein is therefore most likely an integral or covalently anchored membrane protein. If this band is a true isoform of Ago2 or a non-specific cross-reactivity still needs to be determined. Given the negative results from the membrane association assay regarding Ago association with the outer and inner mitochondrial membrane, the two lumenal compartments of mitochondria, the inter membrane space and the matrix, were tested next (Section 4.5.5).

4.5.5 Localisation of Ago Inside Mitochondria

Mitochondria consist of an outer membrane which surrounds a lumenal space called the inter membrane space. This is separated from the inner mitochondrial lumen, the matrix, by the inner mitochondrial membrane. In order to test, if Ago proteins reside in one of these compartments a mitochondrial swelling assay was applied. Here, isolated mitochondria are treated with either an isosmotic buffer (SEM) stabilising whole mitochondria or a hypotonic buffer (EM) which causes water to enter the inter membrane space. The rapid expansion of the outer membrane due to the osmotic pressure in the inter membrane space results in little ruptures of the outer membrane. Inter membrane space components will thus leak out through these tears and are removed by applying protease digestion and washing steps.

Mitochondria were freshly isolated from HEK-293 as described in Section 3.5.8 and split in three fractions: The first was resuspended in isosmotic buffer (SEM), the second was resuspended in hypotonic buffer (EM) and the third was resuspended in hypotonic buffer and served as control for protein degradation by Proteinase K combined with sonication. After swelling of the mitochondria treated with hypotonic buffer, three different amounts of Proteinase K were added to the mitochondria (SEM, EM: 0, 5, 10 μg; sonication: 0, 10 μg). Following digestion of the accessible proteins, the samples were analysed by SDS-PAGE and western blot against Ago and several other marker proteins (Figure 4.31).

Several mitochondrial proteins were used as controls. Tom70 is a receptor for hydrophobic preproteins at the translocon of the outer membrane and was, as expected, accessible to protease treatment under all conditions [Rehling *et al.*, 2001] (Figure 4.31, SEM, EM). Tim23 is a preprotein receptor and potential pro-

tein channel at the inner membrane [Rehling *et al.*, 2001]. Thus, the IMS exposed epitope detected by the Tim23 antibody, was digested by Proteinase K only under swelling conditions (Figure 4.31, EM). MITRAC12 is an integral inner membrane protein involved in cytochrome c oxidase biogenesis. Its C-terminus, recognized by the antibody, protrudes into the inter membrane space and was therefore also degraded by externally added protease under hypo-osmotic conditions [Mick *et al.*, 2012] (Figure 4.31, EM). Tim44 is bound to Tim23 on the matrix side and was so protected from any protease under swelling conditions (Figure 4.31, SEM, EM). The protease digest was controlled by a combination with sonication of isolated mitochondria. Here, all soluble proteins were successfully degraded (Figure 4.31, sonication: PK+). Ago1 was only partially protected from protease digestion under stabilising conditions (Figure 4.31, SEM).

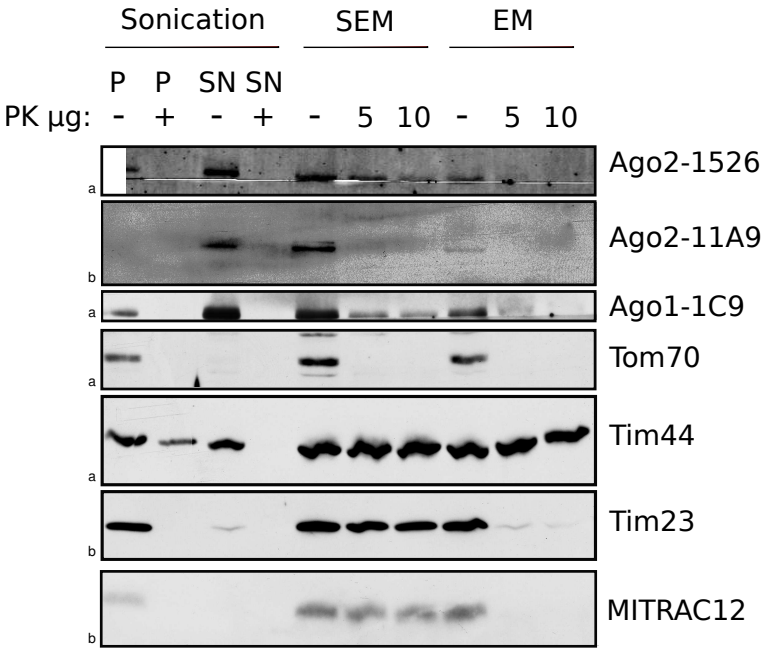


Figure 4.31: ***Localisation of Ago in mitochondria.***
Isolated Mitochondria were treated with isosmotic or hypotonic buffer, followed by Proteinase K digest of the accessible proteins. Ago1 and Ago2 seemed to be partially protected from protease, when the outer membrane was still intact. Under hypotonic conditions all signals were lost. The treatment was controlled by several mitochondrial proteins with well established topology: Tom70: outer membrane protein, accessible to protease under all conditions; Tim44: inner membrane-associated matrix protein, protected under all treatments; Tim23 and MITRAC12: integral inner membrane proteins with protease accessible regions, which are used for antibody detection. Samples were loaded in equal amounts and order on a 10% (a) and 16% (b) SDS-PAGE gel. Further analysis by WB was performed in parallel.

When the outer membrane was destructed, Ago1 was completely degraded by Proteinase K (Figure 4.31, EM). The same was observed for Ago2, which was to a certain extent resistant to Proteinase K addition, when the outer membrane was still intact (Figure 4.31, SEM). However, Ago2 signal was lost when the outer membrane was porous (Figure 4.31, SEM).
In summary, Ago1 and Ago2 proteins were partially protected from Proteinase K treatment under isosmotic conditions (SEM) but were lost from mitochondria under hypotonic conditions (EM). These findings suggest that a fraction of Ago proteins may be protected from protease digestion by the mitochondrial outer membrane, an interaction with contaminating ER, which is always present in these isolations, or another unknown complex which is sensitive to hypotonic conditions. Therefore, it might be interesting to also analyse the behaviour of a lumenal ER protein in this assay.

4.5.6 ^{35}S Labelling of Mitochondrial Proteins Under Ago Knockdown Conditions

So far four experiments were performed assessing mitochondrial localisation of Ago proteins (Figure 4.28, 4.29, 4.31) and membrane association (Figure 4.30), suggesting that Ago proteins potentially reside in the inter membrane space and do not associate with the membrane, except for the larger potential isoform detected by the Ago2-specific antibody 11A9. These findings are in contrast to our *in vivo* DAMID experiments, which showed Dam-Ago-dependent methylation of mitochondrial DNA and encouraged us to examine a possible function of Ago proteins in mitochondria. In order to further dissect these contradicting findings, we thought of another *in vivo* approach focussing on mitochondrial translation. In addition, this assay is independent of biochemical isolation of mitochondria which was required for all previous experiments (Figure 4.28, 4.29, 4.31, 4.30).
The proposed experiment utilises the difference in ribosomal subunits between the mitochondria, comprising a prokaryotic-related 30S small and 50S large subunit, and the human translation machinery, consisting of a 40S and 60S subunit. In the ^{35}S-labelling assay the chemical compound emetine is applied to human cells to irreversibly inhibit the 40S small subunit [Jimenez *et al.*, 1977]. Subsequently, radioactively labelled ^{35}S-methionine is added to the cells which is, due to the block of eukaryotic translation, only incorporated into freshly translated

mitochondrial proteins. In order to test, if Ago proteins have an influence on mitochondrial translation similar to their cytoplasmic function, HEK-293 cells were transfected with two different siRNAs against Ago1 and Ago2 (Si1 and Si2) 72 hours before the labelling reaction. As control, cells were transfected with a non-targeting siRNA (NT). Then, the cellular ribosomes were inhibited by emetine and ^{35}S-methionine was added to the cells for labelling of mitochondrial proteins (see Section 3.5.14). After a radioactive pulse for 1 hour, the cells were harvested and analysed by auto-radiography and WB (Figure 4.32).

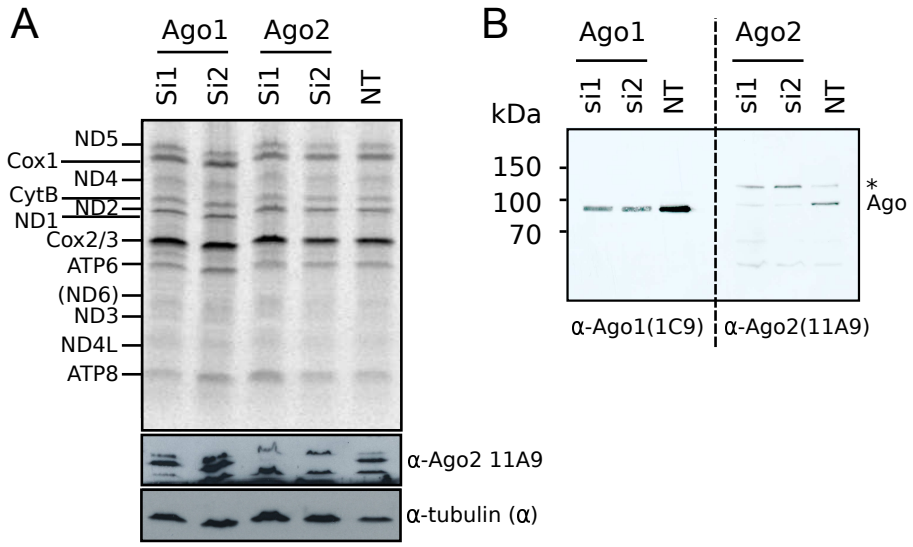


Figure 4.32: ***Analysis of mitochondrial translation under Ago knockdown conditions.*** *Ago proteins were knocked down by transfection of two siRNAs (Si1 and Si2, 33nM) and a control siRNA (NT) for 72 hours. Afterwards, the cellular translation machinery of HEK-293 cells was inhibited by exposure to emetine. Then ^{35}S-methionine was added to the medium and mitochondrial translation was monitored for one hour. Subsequently, the cells were lysed and analysed by WB.* ***A****) For autoradiography of mitochondrial proteins the samples were loaded on an 18% UREA-SDS-PAGE gel and transferred to a nylon membrane. Knockdown of Ago1 or Ago2 by two different siRNAs (Si1 and Si2) showed no change in translation rate when compared with the control transfection with non-targeting siRNA (NT). Additionally, WB analysis of the membrane transferred 18% UREA-SDS-PAGE was performed with antibodies specific for Ago2 and tubulin (loading control). The successfully knockdown of Ago2 with both siRNAs (middle band missing in lanes Ago2-Si1 and Ago2-Si2) was controlled for equal loading by tubulin.* ***B****) The knockdown of Ago proteins was also examined on 10% SDS-PAGE gels (same samples and volumes loaded as for UREA-PAGE, A). For both Ago1 and Ago2 successful siRNA knockdown (Si1 and Si2) compared to the control (NT) was demonstrated. One of two experiments is shown.*

The knockdown of either Ago1 or Ago2 did not seem to influence the translation of all 12 mitochondria encoded proteins (Figure 4.32, A). The knockdown of Ago2 and equal loading was confirmed by detection of Ago2 and tubulin by WB (18 % UREA-PAGE)(Figure 4.32, A, lower part). Due to the low resolution of the 18 % UREA gel the Ago2 band is surrounded by two cross-reacting bands. In order to get a better resolution of the knockdown, sample volumes equal to the UREA-PAGE were loaded on a 10% SDS-PAGE (Figure 4.32, B). Analysis by western blot revealed a good knockdown of Ago1 and Ago2, which was, however, not complete for Ago1. Interestingly, the WB probed for Ago2 (Figure 4.32, B) revealed two bands. The faster migrating protein (~95 *kDa*) was sensitive to siRNA treatment with Ago2-specific siRNAs and thus comprises Ago2. The slower migrating band, which was not altered by the siRNA treatment is most likely a cross-reactivity of the 11A9 antibody.
In conclusion, the above findings suggest that neither Ago1 nor Ago2 have an influence on mitochondrial translation as the successful knockdown of both proteins did not change the signal strength of the 12 mitochondria encoded proteins. Furthermore, the Ago2 knockdown experiment suggests that the siRNA-insensitive, slower migrating protein detected by the Ago2-specific 11A9 antibody, is most likely a cross-reactivity. This is also supported by the different localisation patterns of the faster (Ago2) and slower migrating proteins observed in the Percoll gradient (Figure 4.29) and membrane association assays (Figure 4.30). However, our results from swelling assays (Figure 4.31), suggest mitochondrial inter membrane space localisation for Ago proteins. In contrast, the DAMID experiments from mESC and MRC-5 cells (Figure 4.26, Figure 4.20) support Ago matrix localisation. Even though these findings contradict each other, both require a successful import of Ago proteins into mitochondria. Therefore, the translocation of Ago proteins into mitochondria was analysed by an *in vitro* import experiment (Section 3.5.13).

4.5.7 *In Vitro* Import of Ago Proteins into Mitochondria

For these assays the protein of interest is radioactively labelled with ^{35}S-methionine in a coupled transcription-translation reaction. Subsequently, the labelled protein is incubated with isolated mitochondria and energy. In the case of a successful import of a preprotein, which comprises a target peptide, a radioactive signal inside mitochondria as well as a size shift to a faster migrating form, due

to target peptide cleavage, can be observed. As explained in Section 4.5, there are four import pathways described, today. However, all of these require the proteins to contain a presequence, which is proteolytically removed after import, or an internal targeting signal, which is retained [Schmidt *et al.*, 2010]. Unfortunately, no presequence could be predicted for Ago proteins based on several online prediction programmes.

In the first step of the import experiment mitochondria were freshly isolated from HEK-293 cells and incubated with radioactively labelled preproteins (Section 3.5.12). After the import reaction, mitochondria were treated with protease (PK-/+) to degrade any residual protein on the outside (Figure 4.33). Following a washing step, mitochondria were analysed by SDS-PAGE and auto-radiography. For the import reaction the following proteins were chosen: First, His-Ago1 and His-Ago2 (pET-32), to analyse the influence of an N-terminal modification on the import. Second, Ago2-Δ C-DHFR, which comprises amino acids 1 to 222 of Ago2 fused to Dihydrofolate reductase, is much shorter than the His-Ago proteins and should, therefore, be more readily imported. Third, Cox6A which is commonly used as positive control for the import reaction.

On the autoradiograph of the import reaction His-Ago proteins are found associated with mitochondria before and after Proteinase K (PK-/+) treatment (Figure 4.33), indicating protection of a fraction of the proteins inside mitochondria. The Ago2-Δ-C-DHFR fusion protein is stable in the untreated sample (PK-), but degraded in the presence of the protease (PK+), suggesting a lack of import (Figure 4.33). This is further supported by the complete degradation of the input control His-Ago2 confirming the efficiency of the Proteinase K treatment (Figure 4.33). As import control Cox6A was used, which is commonly imported with high efficiency and N-terminally cleaved during translocation resulting in an observable size shift [Lazarou *et al.*, 2009]. In the autoradiograph, however, Cox6A (12 *kDa*) was not properly resolved due to the SDS-PAGE gel (10%), in which it migrated together with the dye front (Figure 4.33, last band in all lanes). Unfortunately, the signal of Cox6A can only be distinguished from the background signal in the non-protease treated samples (Figure 4.33).

The decrease in signal intensity of the His-Ago-specific bands upon protease treatments suggests that His-Ago import was only limited. This might be improved by adapting the import reactions to the needs of these large proteins. Excitingly, a size shift of His-Ago1 and His-Ago2 was observed when the bands

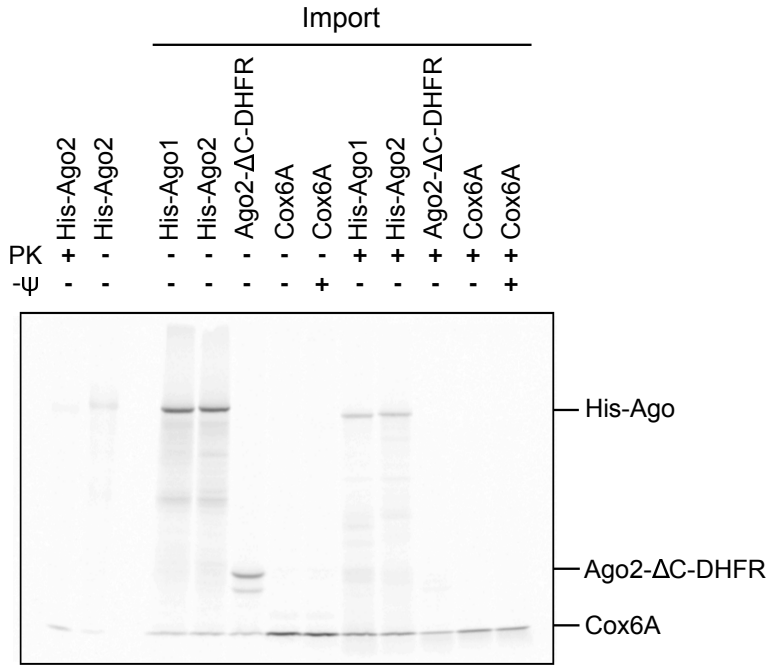


Figure 4.33: ***In vitro* Import of Ago proteins in Mitochondria.**
Mitochondria, freshly isolated form HEK-293 cells, were incubated in the presence of radioactively labelled preproteins in an import reaction. For the control protein Cox6A, also a control on membrane potential dependence was included (−ψ). After incubation for 30 minutes the import reaction was stopped and split into two tubes. One set of samples was treated with protease, the other was used as control. All proteins were successfully translated and stable in the import reaction, as shown in the non protease treated samples. After protease treatment a weak signal for full length Ago proteins and Cox6A were observed, while Ago2-ΔC-DHFR was completely degraded.

with and without Proteinase K treatment were compared, indicating a recognizable targeting sequence and a cleavage event following import. However, given that the Ago2-Δ-C-DHFR fusion protein, carrying the endogenous N-terminus is not imported, the question arises if the modified N-terminus of His-Ago2 carries the characteristics of a targeting sequence and by this triggers the mitochondrial import of the protein. Therefore, the sequence of the His-Ago fusion proteins was applied to the target prediction software targetP [Emanuelsson *et al.*, 2000]. Remarkably, the score for mitochondrial localisation was highly increased for His-Ago2 (0.737) compared to the endogenous N-terminus (0.204). This means, that mitochondrial import of endogenous Ago proteins is more likely represented by the Ago2-ΔC-DHFR fusion protein, while N-terminal modification with a His-tag may lead to false positive signals. The endogenous N-terminus is, therefore,

not sufficient for Ago import into mitochondria. In addition to independent validations of these findings, optimisation of the import efficiency and a control for dependence on mitochondrial membrane potential, driving the import across the inner mitochondrial membrane, is needed. A reliable signal for the positive control Cox6A is required, as well.

In summary, the above experiments suggest that Ago2 is not imported into the mitochondrial matrix, while N-terminal extensions (6xHis or Dam) may cause a limited degree of import. This is in agreement with the non-mitochondrial localisation of Ago1 and Ago2 from Percoll gradients (Figure 4.29) and the lack of membrane association (Figure 4.30). Further support comes from the absence of mitochondrial DNA in Ago-ChIP experiments and unchanged mitochondrial translation under Ago protein knockdown (Figure 4.32). Here, only the swelling assay (Figure 4.31) with its potential ER contamination, suggests a potential inter membrane space localisation, while supporting the absence of Ago proteins from the mitochondrial matrix. However, more experiments under Ago knockdown or knockout conditions also including the DAM-Ago fusion proteins are needed to fully explain above experiments.

Chapter 5

Discussion

Ago proteins mediate small RNA induced post-transcriptional gene silencing in the cytoplasm but also nuclear localisation and functions were described in several organisms (see Section 1.5.3, 1.5.4, 1.5.5, 1.5.6). The latter range from small RNA induced heterochromatin formation in *S. pombe* [Aravind *et al.*, 2000; Volpe *et al.*, 2002] to Ago4 dependent recruitment of the DNA and histone methylation machinery in plants [Zilberman *et al.*, 2003]. Inspired by the findings in other organisms, the search for indications on nuclear localisation and function of human Argonaute proteins began.

Applying cellular fractionation, the nuclear localisation of Ago2 was reproduced (see Section 4.1), as has been shown in similar experiments by [Robb *et al.*, 2005; Rüdel *et al.*, 2008; Ohrt *et al.*, 2008; Chu *et al.*, 2010; Gagnon *et al.*, 2014] and by detection of nuclear miRNA and cleavage competent complexes in Meister *et al.* [2004]. This was further confirmed by studies applying confocal microscopy on Ago1 and Ago2 [Rüdel *et al.*, 2008; Ameyar-Zazoua *et al.*, 2012; Benhamed *et al.*, 2012; Huang *et al.*, 2013a].

During the last 11 years many hints on the nuclear functions of human Ago1 and 2 were gathered (Section 1.5.7, 1.5.8). These include, the targeting of Ago proteins to chromatin by promoter complementary siRNAs. However, an Ago pathway guided by endogenous small RNA species that targets mammalian chromatin had not been convincingly reported. To elucidate such a pathway, potential chromatin association sites of Ago proteins had to be identified in somatic cells. In order to test if a fraction of nuclear Ago proteins are associated with chromatin even in the absence of promoter targeting siRNAs, a chromatin association assay was performed [Cernilogar *et al.*, 2011]. In this assay (Figure 4.2) a sub set of Ago

proteins were found in a chromatin containing pellet, which has been described for human Ago1 and Ago2 proteins also by other studies [Cernilogar *et al.*, 2011; Ameyar-Zazoua *et al.*, 2012; Huang *et al.*, 2013a; Gagnon *et al.*, 2014]. When the chromatin pellet was further digested with DNAse, Ago proteins and integral chromatin components (histone H3) were released, arguing for a strong interaction of Ago proteins with chromatin, which was only destroyed by removal of the DNA. This finding supports a potential targeting of Ago proteins to chromatin by endogenous triggers, which could result in the regulation of transcription and heterochromatin formation as observed in other organisms [Bühler *et al.*, 2006; Guang *et al.*, 2008; Naumann *et al.*, 2011; Kuhlmann *et al.*, 2014] but also in involvement in splicing or DNA DSB repair (compare Section 1.5.8). Based on the encouraging result from the chromatin association assay, we set out to elucidate potential sites for Ago dependent transcriptional regulation or epigenetic modifications by ChIP and DAMID experiments in HeLa, MRC-5 and mESC cells, respectively.

5.1 Results from the ChIP Experiment

For the study of Ago-chromatin interaction ChIP-sequencing experiments were applied, using three different antibodies each for Ago1 and Ago2 (Section 4.2). To select for regions with a high potential for being true Ago binding sites, only overlapping peaks from all three different antibody-libraries were regarded as functional binding sites. This was done to exclude non-specific binding sites from potentially cross-reacting proteins or other unspecific interactions of the antibodies. Unfortunately, only nine regions for Ago1 and ten regions for Ago2 were found to be enriched in all of the three different ChIP preparations.

5.1.1 Limitations of the ChIP-Sequencing Experiment

In general, the number of overlapping peaks is too low to draw any conclusions. The amount of overlapping peaks is, however, restricted by the number of peaks per antibody used for the overlap analysis. In the case of Ago1, the input set of peaks for the analysis across three different antibodies comprised only 578 (A1-4B8), 1106 (A1-Rabbit) and 3699 (A1-Millipore) regions while the set for Ago2 contained 2863 (A2-5D4), 871 (A2-Rabbit) and 715 (A2-Millipore) sites (see Ta-

ble 4.3) for a genome of $3 * 10^9 bp$. As the analysis used a stringent p-value of 0.00001 to identify the sites of Ago enrichment in the first place, the overlap between the heterogenic set of antibodies used, was rather limited. Therefore, it might be worthwhile to use the overlap across the different antibodies as stringent selection criterion and start on a rather broad basis of potential peaks.

Another reason for the low amount of detected peaks may be the read depth. When the NGS-data from the conducted Ago1/2 ChIP experiments with at least 10 million reads per sample, is compared with the first successful ChIP studies using two to five million reads and the ENCODE consortium guidelines of 20 million reads per TF [Sims *et al.*, 2014], the data set seemed large enough to get a first glimpse of the potential Ago1 and 2 chromatin interaction sites. Low strength interaction sites and low abundant components would have probably escaped our analysis. However, several studies implied that Ago1 and Ago2 may bind rather indirect to the DNA [Schwartz *et al.*, 2008; Chu *et al.*, 2010; Younger and Corey, 2011], thus aggravating our experiments because our measurements might be close to the background noise. Along these lines, low read numbers were found in our ChIP-sequencing experiments, which hint at an indirect binding mechanism but made it difficult to set a threshold to clearly separate the signal from the background noise. This problem might have also been severed by the purification state of the used antibodies. Here, the rat monoclonal antibodies (A1-4B8, A2-5D4) and the rabbit antibodies (A1-rabbit, A2-rabbit) were only purified on ProteinA-Sepharose beads used for IP and not added in concentrated form to the ChIP lysate. In addition, the lack of replicates made statistic analysis impossible and thus selection of potential peaks for validation by other techniques.

During the preparations for the ChIP-Seq experiment we have also tested several antibodies and cell lines, including siRNA mediated targeting of Ago proteins to the progesteron receptor promoter in MCF7 cells [Janowski *et al.*, 2007], miR-320 targeting to POLR3D in HEK293 [Kim *et al.*, 2008] and endogenous targeting to the CD44 locus [Ameyar-Zazoua *et al.*, 2012] to optimise our ChIP experiments. Unfortunately, these sites turned out to be non-reproducible in our hands. Therefore, we can not decide, if the negative ChIP results were caused by a lack of chromatin association sites or weak precipitation capability of our antibodies under ChIP conditions.

5.1.1.1 Results from the ChIP-Sequencing Experiment

Both Ago1 (6 peaks) and Ago2 (10 peaks) were enriched at satellites sequences annotated as Alpha repetitive DNA (ALR), which is a human satellite consensus sequence found at the centromeres of the human chromosomes [Vissel and Choo, 1987]. Our observed association of Ago1 and Ago2 with the pericentric chromatin could suggest a potential involvement of the RNAi machinery in heterochromatin formation similar to yeast (Section 1.5.3). According to this hypothesis, small RNAs deriving from centromeric satellites are needed to guide Ago proteins to complementary sequences on the chromatin. As small RNAs are commonly processed in a Dicer-dependent manner, Dicer could be also involved. Therefore, Ago or Dicer depletion should result in defects of centromer function and increased transcription from pericentric satellites as described for yeast [Volpe *et al.*, 2002].
This hypothesis is supported by studies form mESCs deprived of Dicer, which showed strong defects in differentiation, an increase in the expression of centromeric transcripts and reduced levels of small RNAs (25-30 *bp*) descending from minor satellite sequences [Kanellopoulou *et al.*, 2005; Murchison *et al.*, 2005]. However, a role of mouse Ago2 in silencing by DNA methylation of centromeric repeats, imprinted genes and methylation of the inactivated X chromosome was ruled out by studies of Ago2 knockout animals [Morita *et al.*, 2007]. Despite of this, the potential enrichment of Ago proteins at satellite sequences is highly interesting and needs experimental validation. Unfortunately, this data was supplied by the bioinformatician only after the end of the experimental part of my thesis.
Recently, a NGS study analysed 80-100 million reads form Ago1 and Ago2 ChIP samples isolated from PC3 cells [Huang *et al.*, 2013a]. As they found some overlap between peaks of the activating histone mark H3K4me3, Ago1 and RNA PolII ChIP-Seq experiments, the authors speculated about a transcription promoting function of Ago1 at the transcription start sites of most of their identified genes [Huang *et al.*, 2013a]. Since replicates were not performed [Huang *et al.*, 2013a], the significance of this study is not clear and needs further validation.
The suggestion of human Ago1 as being responsible for transcriptional gene regulation, while excluding Ago2, which was implied by others, stresses once more the controversial findings regarding Ago chromatin interaction (see Table 1.1). In our data we found overlap between Ago1 and Ago2, at least regarding the ge-

nomic element bound, which may support a potential role of Ago1 and Ago2 at pericentric satellites. As described above, the limited number of peaks, lack of validation and replicates, prohibits further speculations (Section 5.1.1).

5.2 Results from the DAMID-Seq Experiments

In addition, we used DAMID, an antibody independent *in vivo* approach, to examine Ago1 and Ago2 chromatin association [Steensel and Henikoff, 2000]. After successful confirmation of expression of all Dam-fusion proteins and slicer activity of Dam-Ago2 (Figure 4.14), DAMID experiments followed by NGS were performed in MRC-5 cells and mESCs.

5.2.1 High Rates of Duplicate Reads in DAM-ID Experiments

When DAMID-libraries form human and mouse cells were analysed with respective bioinformatic methods, a high rate of duplicate reads, which are reads with identical start and end coordinates deriving from the same DNA strand, were observed. While duplication levels of ~15 % for controls, Dam-Ago samples showed high rates of duplicate reads ranging form ~30 % for human cells to ~80 % for mESC (Table 4.13, 4.22). When single chromosomes were analysed, human mitochondria, several mouse chromosomes and mouse mitochondria were identified to generate most of the duplicate reads (Figure 4.19, 4.25).

Influence of PCR Amplification on Rates of Duplicate Reads

These high duplication rates are probably caused by the PCR amplification step in the NGS-library preparation. The observed over-amplification is most likely due to a low diversity of DNA sequences in the input sample, which was already implied by the band pattern observed for Dam-Ago2 fusion proteins (Figure 4.15b). In contrast to Dam-Ago, samples from Dam-Cbx1 and Dam control transduced cells comprised a highly diverse composition of DNA fragments, which resulted in a smear on the agarose gel (Figure 4.15b), and showed markedly reduced rates of duplicate reads (Figure 4.19). This trend can be improved by using more input material [Casbon *et al.*, 2011], but is in turn aggravated by reduced input amounts or a high number of PCR cycles during library preparation. Given that the original DAMID protocol is based on microarray analysis, these problems were not

addressed by Vogel *et al.* [2007]. Low complexity DAMID reactions were only described in the protocol for DNA samples from transiently transfected cells, which was in human cells prevented by transduction of cells using lentiviruses and in mESC by generating stable cell lines.

Additional Causes for Duplicate Reads

Another reason for the high duplication rate observed in above described experiments is inherent to the DAMID technique, as all amplified fragments start and end with GATC. However, during size fractionation by sonication, the respective 5′ and 3′-ends of the reads should be caused by sonication and thus at least reduce the fraction of reads with exactly matching start and end coordinates in the genome. The fragmentation of DAMID amplicons to a mean fragment length of ~250 bp was examined by Bioanalyzer analyses following sonication.
The especially high rate of duplicate reads deriving from mitochondria is explained by the nature of the mitochondrial genome. First, the mitochondrial DNA is quite short comprising only ~16 kB. Second, it contains 24 (32 for mouse) GATC sites. This is a mean distance of 679 bp (447 bp for mouse) and a maximum distance of 2334 bp (1605 bp for mouse) between two GATC sites. Therefore, all GATC flanking sequences from mitochondria can be easily amplified during the PCR amplification step during the DAMID experiment while some even longer fragments deriving from the nuclear genome are deselected [Vogel *et al.*, 2007]. Finally, human lung fibroblasts contain about 300 mitochondria, each of which contains 2-3 copies of DNA [Robin and Wong, 1988].
When comparing mitochondrial reads from the samples Dam-Cbx1, Dam and Dam-Ago2 (Figure 4.19), a specific problem of the experimental setting becomes apparent. The number of mitochondrial reads per sample increase steadily, starting lowest for the strictly nuclear Dam-Cbx1 (HP1-β), to Dam, which is equally distributed in the cell [Steensel and Henikoff, 2000], peaking for the Dam-Ago2 protein, which is assumed to distribute in a ratio of 4 to 1 between cytoplasm and nucleus [Ohrt *et al.*, 2008]. Even though N-terminally GFP-tagged Ago2 correctly localises to the cytoplasm and the nucleus [Ohrt *et al.*, 2008] and a note in the DAMID protocol on likely similar behaviour of GFP- and DAM- tagged proteins [Vogel *et al.*, 2007], the extent of nuclear localisation of Dam-Ago2 needs to be tested experimentally.
In summary, the Dam-Ago libraries from human cells contained many duplicate

reads, which was even more severe for the mouse samples. If the high number of reads deriving from mitochondrial DNA was caused by Ago localisation to mitochondria was experimentally assessed and will be discussed in Section 5.3. Future experiments could be significantly improved by directing the Dam-tagged target protein to the nucleus by fusing it to a nuclear localisation signal (NLS), which is unlikely to cause any defects in the cell due to the low expression level of the fusion-proteins. In addition, maximising the sample input in the library preparation while concomitantly reducing the PCR cycles in the library preparation to a minimum will considerably increase the quality of the NGS results.

Another improvement to the DAMID technique was suggested by Zhou [2012]. Her workflow follows the general approach including the DPN I digest and adaptor ligation [Vogel *et al.*, 2007], which selects for methylated fragments. Instead of the DPN II digest, used to deprive of non-methylated fragments, an adaptor specific PCR is performed with a biotinylated primer. This amplification step enriches all methylated fragments including those with non-methylated sites in between. After fragmentation to a size range amenable to NGS the DPN II digest is performed to cleave non-methylated sites. Then methylated sites are enriched by avidin-pull down and NGS-libraries are prepared from the enriched fraction [Zhou, 2012]. In the resulting library, methylation levels of GATCs can be distinguished by their accompanying sequence, which is either the DAM-ID PCR primer in the case of methylation, or no sequence, as the associated library adapter is automatically removed by the sequencing facility. The proposed workflow leads to selection of GATC-containing sequences for NGS-library preparations and thus reduces non-specific background DNA notably improving the signal-to-noise ratio of the original DAMID protocol.

5.2.2 Shared Sequence Between Mouse DNA and Mitochondrial Genome

A remarkable feature of the mouse genome, which became apparent during bioinformatic analysis, is the identification of a 4136 *bp* long sequence shared between the mitochondrial DNA (6600 to 10735) and chromosome 1 of the mouse genome (24611842 to 24615978) (Table 5.1). In the respective region a strong accumulation of reads could be seen, which could not be unambiguously assigned to any of the above mentioned regions.

SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
4137	1	4136	4136	100%	M	−	6600	10735	4136
4133	1	4136	4136	100%	1	+	24611842	24615978	4137

Table 5.1: ***Identical DNA fragment of mouse ChrM and Chr1***

Many nuclear DNA sequences of mitochondrial origin, called NUMTs, have been found in organisms ranging from yeast to humans [Ricchetti *et al.*, 2004]. There are several evolutionary conserved mitochondrial DNA integrations found in eukaryotes but also new integration events occur and are suggested to be caused by environmental conditions causing DNA DSBs. [Ricchetti *et al.*, 2004]. Unfortunately, no hints on a functional role of the mitochondrial sequence on chromosome 1 of the mouse genome, presented above, could be found in the current literature.

5.2.3 Identified Peaks and Validation of DAMID-MRC-5 Experiment

Based on the first bioinformatic analysis (Homer, Section 4.4.3.3), we could confirm the successful establishment and application of the DAMID technique by overlap of the CBX1-Dam sample with a reference data set [Vogel *et al.*, 2006]. The strong association of CBX1 with SINE elements, found in our genome wide study, was not supported by the reference study, which was, however, based on an exon-array of chromosome 19 with very limited representation of SINE elements. Even though the method was successfully applied with respect to Cbx1, none of the peaks selected for validation for Ago proteins by could be verified by ChIP validation experiments of 25 selected loci, (Figure 4.17, 4.22), comprising also peaks from the genomic element classes exon and SINE, which were found to be enriched for Dam-Ago1 and Dam-Ago2. Due to the lack of replicates for the human cell line the bioinformatic toolset to select for likely candidate regions was limited and false discovery rates could not be predicted. This might have hampered the identification of true enrichment sites in the data and their subsequent validation. In addition, low read counts for human genomic DNA loci (median count 17, second analysis) and high background noise in DAMID-Seq experiments negatively affected the data analysis. Therefore, a repetition of the experiment under improved experimental conditions as outlined in (Section 5.2.1) will help to gain further insights into Ago chromatin interaction.

Recently, Ago-chromatin interactions were analysed in pre-senescent versus senescent primary lung fibroblasts (WI-38) [Benhamed *et al.*, 2012], which are similar to the fetal primary lung fibroblasts, tested in our DAMID experiment. Applying ChIP-on-ChIP against Ago proteins using a pan-Ago antibody they identified 4516 bound promoters in senescent cells compared to 2619 for pre-senescent control cells [Benhamed *et al.*, 2012]. In addition, they could link the target promoters of the E2F transcription factor with 320 and 77 promoters bound by Ago proteins (described in Section 1.5.7). This is the only high throughput experiment probing Ago-chromatin interactions using replicates (compare Table 1.1). Therefore, they were able to apply statistic methods to remove random binding sites and enrich for true positive peaks. In addition, this study was conducted in a closely related cell line. Due to these reasons we used the reported chromatin interaction sites of Benhamed *et al.* [2012] to assess the different analyses methods applied to our DAMID experiment.

The first analysis was conduced using the software Homer. Here, 628 of 6289 predicted Dam-Ago1-chromatin interaction sites were found to overlap with the gene list from pre-senescent cells [Benhamed *et al.*, 2012] (Table 4.9). For Dam-Ago2, 89 of 866 binding sites were in common with the data set from pre-senescent cells (Table 4.10). The homer software supplies the user with two values to decide on the quality of the peaks: the peak score and the focus ratio. Unfortunately, both values which were used for selection of peaks for validation did not help to identify any of these overlapping, potentially true positive sites. Therefore, replicates would have helped to remove false positives from our list of predicted peaks and to decide for reproducible candidate regions.

The second approach based on the DESeq package and all mapped reads, identified 1583 peaks for Dam-Ago1 and 870 for Dam-Ago2. When the overlap with the data from Benhamed *et al.* [2012] was examined 101 and 53 peaks were found. However, only one overlapping peak was detected within the first 40 peaks sorted by their log2-fold change for Dam-Ago1 compared to Dam (Table 4.14). The first 40 peaks of Dam-Ago2, did not contain any of the peaks identified by Benhamed *et al.* [2012] (Table 4.15). Thus, the overlapping peaks were again not sorted to the top of the list and escaped our validation experiments. The DESeq package relies on replicates to estimate the false discovery rate, discard false positives and calculate proper p-values.

The last analysis which applied the DESeq package on unique reads was con-

ducted after my contract ended. Here, very stringent selection criteria were used which resulted in very low numbers of detected peaks yielding 8 peaks for Dam-Ago1 and 40 peaks for Ago2 (Table 4.16, 4.17). When compared with the genes identified in pre-senescent cells [Benhamed *et al.*, 2012], one peak of Dam-Ago1 and three peaks of Dam-Ago2 were found to overlap. However, the p-values, which are not corrected for false positives due to the lack of replicates, still not met the commonly accepted threshold of 0.05 (compare Table 5.2). If these overlapping peaks are true Ago-chromatin association sites, needs to be shown in independent experiments.

Dam-fusion	chrom	start	end	Ago	Dam	log2-Fold Change	pval	repclass	external-gene-id	GO-domain	source
Ago1 (1/8)	chr10	506102	506105	50	4	3.866	0.069	NA	DIP2C	metabolic process	intron
Ago2 (3/40)	chr1	92455156	92455159	44	4	3.781	0.057	SINE/Alu	BRDT	positive regulation of transcription during meiosis	intron
Ago2 (3/40)	chr17	30651373	30651376	34	4	3.409	0.065	NA	RHBDL3	proteolysis	exon
Ago2 (3/40)	chr1	43995062	43995065	32	4	3.322	0.067	NA	PTPRF	transmembrane receptor protein tyrosine phosphatase signaling pathway	intron

Table 5.2: ***Peaks of the DAMID experiment overlapping with potential chromatin association sites identified in pre-senescent cells Benhamed*** *et al.* ***[2012].***

In summary, we do see overlapping peaks between our analysis and predicted Ago-chromatin association sites in pre-senescent WI-38 cells [Benhamed *et al.*, 2012]. This suggests, that the DAMID experiment was also successfully applied for Dam-Ago proteins. The lack of detection of true peaks is probably caused by the absence of replicate experiments, the low median read count values for nuclear encoded genes and many duplicate reads (see Section 5.2.1). This is most likely caused by a low nuclear import rate of the Dam-Ago proteins, which could be increased by fusion of a NLS. Since we have not yet tested these overlapping peaks in validation experiments, we can still only speculate about Ago-chromatin interactions.

5.2.4 Identified Peaks and Validation of the DAMID-mESC experiment

In the case of mESC cells a very limited number of binding sites with high significance was reported by the DAMID experiment. Therefore, no binding pattern or class of transcripts could be inferred from the sites of enrichment. In addition to a set of mitochondrial genes which will be discussed in Section 5.3, also seven genomic sites could be identified (Table 4.25) (p-value <0.05). These comprise several predicted genes potentially associated with perception of sound, a telomeric satellite region, which partly overlaps with a predicted lincRNA, and a LINE L1 element. Since mESC are not cultured in our lab, the gene list was sent to the lab of our collaborator Prof. C. Ciaudo, for independent validation experiments, which are currently conducted.
Unfortunately, no information could be found for most of the genes or regions enriched in the DAMID experiment in mESC cells. For chr9 (3000001-3007435), which maps to a telomeric satellite sequence, the only link to the RNAi machinery is presented in *S. pombe* where RNAi plays a redundant role in silencing and heterochromatin formation at telomers Kanoh *et al.* [2005]. Therefore, it might be possible, that mouse RNAi is involved in heterochromatin formation at telomers. Given, that we identified only one locus this is rather speculative. Interestingly, the LINE L1 element (L1Md-T) has been already implied in a PTGS mechanism in mESC and will be discussed in the following.

The LINE L1Md_T

Ago2 seems to be recruited to a region in the second ORF of the LINE L1Md_T element in the DAMID experiment (p-value < 0.05). Line elements are transposable elements, which consist of a 5′-UTR, ORF1 and ORF2 and a 3′-UTR, translated by RNA Polymerase II. In germ cells, retrotranposons are silenced by Piwi proteins via TGS and PTGS [Aravin *et al.*, 2008]. In somatic tissues, transposable elements are suppressed by DNA promoter methylation, which is established at the stage of embryonic implantation [Smith *et al.*, 2012]. However, the various ways how methylation is restored, remain elusive. The LINE class L1Md_T is demethylated during fertilisation and activated, which has been shown by reduced representation in bisulfite sequencing experiments by Smith *et al.* [2012]. In bisulfite sequencing experiments the methylation of cytosine in CpG dinucleotides, an

epigenetic mark implicated in transcriptional repression, is probed [Smith *et al.*, 2012]. During the early blastocyste stage, modelled by mESCs, the transcripts of de-repressed LINEs are controlled by RNAi-dependent mechanisms [Ciaudo *et al.*, 2013]. Ago2 complexes seem to target especially the 5′-UTR, as these sequences could be precipitated in an Ago2-IP [Ciaudo *et al.*, 2013]. In contrast, Dam-Ago2 was possibly recruited to a sequence in ORF2 of L1Md_T, based on the genomic region enriched in the DAMID experiment. Excitingly, Ciaudo *et al.* [2013] identified a second small RNA peak in their Ago2-IPs and small RNA isolations from whole cells, likely mapping to the same position. Therefore, further analysis and experiments are needed to examine this link. This is currently performed by our collaborator Constance Ciaudo.

Another source of endo-siRNAs are antisense promoters in the 5′-UTR of LINE L1 analysed in HEK293 and HeLa cells [Yang and Kazazian, 2006]. These give rise to long dsRNAs involving an antisense transcript and the LINE element and are presumably processed by nuclear Dicer [Yang and Kazazian, 2006; Doyle *et al.*, 2013; Gagnon *et al.*, 2014]. However, these siRNA were implicated especially in targeting of transposable elements by PTGS mechanisms in cytoplasm. However, Yang and Kazazian [2006] can not exclude other RNA silencing mechanisms like TGS.

Further down this line, Chen *et al.* [2012] analysed retrotransposition of L1 elements in breast cells. In normal cells L1 retrotransposition is extremely rare, while L1 activity is highly increased in breast cancer cell lines accompanied by hypomethylation of the L1 promoter. Knockdown of Dicer or Ago2 in normal breast cells resulted, however, in increased retrotransposition activity of a L1 reporter construct. Comparing small RNA libraries from normal breast cells and T47D breast cancer cells, they could identify endo-siRNAs targeting the 5′-UTR of L1 [Chen *et al.*, 2012], which has also been observed in mESCs [Ciaudo *et al.*, 2013]. Applying bisulfit sequencing of the promoter of the L1 element in T47D and other breast cancer cells expressing sh-constructs of the endo-siRNAs, revealed increased DNA methylation of the promoter compared to hypomethylation in wild type cells. This silencing coincided with reduced expression of L1 in the respective cells. Therefore, targeting of reactivated L1 elements by endo-siRNA in somatic cells as suggested by Chen *et al.* [2012] seems a likely mechanism, but still needs independent validation. We do, however, not see an association of Dam-Ago2 with the 5′-UTR but with a sequence in the second ORF

of the L1Md_T. This could be due to a lack of GATC sites in 5′-UTR. In addition, the observation of only one incidence of LINE L1 binding may be caused by a reduced library sensitivity due to the over-amplification of mitochondrial sequences (see Section 5.2.1). Despite of this, more examples of Dam-Ago2 LINE association would be needed to support such a pathway. Furthermore, details on the proteins involved and the mechanism remain elusive.

5.3 Potential Ago2 Localisation and Function in Mitochondria

The notion of a potential mitochondrial localisation of Ago proteins was fostered by the high coverage of mitochondrial DNA fragments in the DAMID experiments for Dam-Ago compared to the Dam control (Figure 4.20, 4.26). As this enrichment of Ago at mitochondrial DNA could not be validated by ChIP (Figure 4.22), the idea emerged that methylation of mitochondrial DNA could just be a side effect of its localisation to this compartment. Therefore, a cellular fractionation experiment was conducted showing a clear enrichment in the mitochondria containing fraction (Figure 4.28).

Furthermore, several publications implied mitochondrial localisation of Ago proteins. Bandiera *et al.* and Barrey *et al.* [2011] rely on immunopurification of mitochondria via a Tom20-specific antibody from fractionated cell lysates. Here, special care must be taken. On the one hand, mitochondria were described to associate with P-bodies [Huang *et al.*, 2011], where also a big fraction of the cytoplasmic Ago is found. On the other hand, the ER was shown to be a nucleation site for the RISC complex [Cikaluk *et al.*, 1999; Stalder *et al.*, 2013]. Additionally, already early fractionation studies applying Percoll gradients implied the problems to separate mitochondria and the ER into pure fractions [Reinhart *et al.*, 1982]. Unfortunately, both of the studies lack this important control.

In several publications also small RNAs were isolated from mitochondria, which might suffer from contaminating Ago bound to co-purifying ER or associated P-bodies [Kren *et al.*, 2009; Bandiera *et al.*, 2011; Barrey *et al.*, 2011; Sripada *et al.*, 2012]. Most recently, Ago2 was shown to be localised inside mitochondria and mitoplasts from isolated mouse heart and C2C12 cells [Zhang *et al.*, 2014]. In contrast to previous studies, the potential contamination with ER was tightly

controlled and therefore this study presents a first solid link between increased translation of mitochondrial encoded proteins (ND1, COX1) and the presence of Ago2 associated with miR-1 in mitochondria [Zhang *et al.*, 2014].
In collaboration with the lab of Professor Dr. P. Rehling (Dr. S. Dennerlein), which is specialized on mitochondria, including protein import and functional analysis, the association of Ago proteins with this organelle was examined. In a first step of our experimental analysis, the separation of the ER and mitochondria was accomplished by Percoll density gradient centrifugation of mitochondria, following isolation by differential centrifugation. Here, Ago2 fractionated in a slower and faster migrating form with ER marker proteins. However, these fractions were also rich in mitochondrial marker proteins. Only the slower migrating form was associated with the mitochondria enriched fraction F2 and the pure mitochondrial fraction F3, controlled by the mitochondrial marker proteins MITRAC15, MFN2 and MRPS15 (Figure 4.29). The lower migrating form seems to solely associate with the ER, as it is absent from the latter fractions, which is in agreement with the results from Stalder *et al.* [2013]. The larger migrating form might either be a modified Ago protein, a larger isoform or an unspecific band of the Ago2 antibody (11A9) (compare Figure 4.29: F2, F3). This cross-reacting epitope might only become visible in cytoplasm-depleted fractions, from which the main fraction of Ago proteins had been removed. As mitochondria and the ER are surrounded by membranes, this structure was analysed in more detail.

5.3.1 Membrane Association of Ago1 and Ago2

The association of Ago1 and Ago2 with mitochondrial membranes was examined by carbonate extraction [Fujiki *et al.*, 1982]. In this assay, the membrane vesicles are broken into sheets. Integral membrane proteins and associated proteins still fractionate with the membrane sheets in the pellet, as Tim44 (see Figure 4.30). Under more basic, conditions protein-protein interaction is lost, as can be see by dissociation of Tim44 from the translocase subunit Tim23. In this assay, Ago1 and the similar sized Ago2 band showed no membrane association under both treatments. For Ago2, however, a second slower migrating protein or isoform was detected by the Ago2 antibody (11A9), which behaved like an integral membrane protein and resisted membrane extraction (see Figure 4.30, pH 11.5). The very last finding is in agreement with experiments by Bandiera *et al.* [2011], who demonstrate membrane association of Ago2 under similar conditions. In their

```
gi|119612614|gb|Iso_CRA_b             MAASLPLRLASRGIICFMFFCRDKGINRVTRVKNAEKAKEDAKSDERALAPPAPPPPIQG  60
gi|29171734|ref|NP_036286.2|Iso_1     ------------------------------------------MYSGAGPALAPPAPPPPIQG  20
gi|257467482|ref|NP_001158095.1|Iso_2 ------------------------------------------MYSGAGPALAPPAPPPPIQG  20
conservation                                                                    ..   *************
```

Figure 5.1: ***Clustal ω alignment of Ago2 isoforms.***
The first line of the alignment is shown, comprising the mitochondrial localisation signal and the difference in the amino acids at position 41 to 47, which might negatively influence the epitope recognition by the Ago2 antibody (11A9)

publication, the authors mention a longer Ago2 isoform, called CRA_b, which is extended by 40 amino acids at the N-terminus (Figure 5.1) and carries an insertion of 6 amino acids at the end of the PAZ domain relative to Ago2, isoform 1 [Schirle and MacRae, 2012]. Applying targetP on the CRA_b isoform a high prediction score (0.79) is gained for mitochondrial localisation in contrast to a score of 0.2 for isoform 1 and 2 [Emanuelsson *et al.*, 2000]. There are, however, further concerns regarding the CRA_b isoform. At least in our membrane association assay the detection of the potentially expressed longer CRA_b isoform with the Ago2 (11A9) antibody may be hampered given that 7 of the 15 amino acids of the peptide used as immunising antigen are not conserved in the longer isoform. In addition, the longer CRA_b isoform does not explain the strong methylation rate of mitochondrial DNA caused by Dam-Ago1 and Dam-Ago2. Both fusion proteins were based on isoform 1 of the respective protein. Furthermore, no mitochondrial import was observed for fusion proteins consisting of the N-terminus of the CRA_b isoform and GFP [Zhang *et al.*, 2014]. Therefore, further experiments are needed confirming the expression of the CRA_b isoform in different cell types at transcript level as well as the detection by available antibodies.

5.4 Localisation and Import of Ago Proteins into Mitochondria

The localisation of Ago proteins in mitochondria was further dissected by applying swelling assays. Here, mitochondria were treated either with a stabilising buffer or a swelling buffer leaving only the inner membrane intact, while large holes formed in the outer membrane. The accessibility of proteins under each condition was probed with Proteinase K treatment which lead to degradation of accessible proteins (i.e. Tom70). Under stabilising conditions, Ago1 and Ago2

were in part protected by the outer membrane (Figure 4.31). Under swelling conditions, when the outer membrane was broken, all Ago proteins were digested similar to the IMS protein Tim23 and in contrast to matrix control proteins (i.e Tim44). This supports a potential inter membrane space localisation of Ago proteins or can be interpreted as Ago proteins being protected by one membrane layer sensitive to hypo-osmotic conditions. In addition, also *in vitro* import into mitochondria of N-terminally His-tagged forms of Ago1/2 and the Ago2 N-terminus (1-74 aa) fused to di-hydrofolate reductase was tested. Successfully imported proteins are resistant to protease treatment, while stability of the proteins under experimental conditions is demonstrated in the absence of the enzyme. Here, the Ago2 N-terminus fused to DHFR was stable under standard conditions but completely degraded upon protease treatment, suggesting that no mitochondrial targeting sequence is present in residues 1 to 74 of the endogenous protein (Figure 4.33). For His-Ago1 and Ago2 only a weak slightly shorter protein band was observed, which implies a limited import and N-terminal processing of the proteins (Figure 4.33). When targetP analysis was performed with the His-Ago2 protein sequence a high mitochondrial targeting peptide score of 0.7 was predicted compared to 0.2 for the endogenous protein [Emanuelsson *et al.*, 2000]. Therefore, it seems that an N-terminal His-tag can lead to false positive signals in import assays at least in the case of Ago1/2.

Taken together, the failed import into mitochondria, the lack of a mitochondrial transport signal in the amino acid sequence and no matrix localisation of Ago2 in the swelling assay, a protein import into the mitochondrial matrix seems to be a rather unlikely explanation for the reads observed in the DAMID experiment. This stand in contrast to a recent publication working with C2C12 cells and mouse heart, showing mitochondrial localisation for a fraction of the cellular Ago2 pool [Zhang *et al.*, 2014]. Additionally, the transfection of MEF and HeLa cells with miR-1 leads to enhanced translation of the mitochondrial encoded proteins ND1 and Cox1, which was suggested by Zhang *et al.* [2014] to be dependent on the import of Ago2 into mitochondria. Here, the localisation of Ago2 in HeLa cells, however, still needs to be shown.

Similar to the study above, also a functional experiment inquiring a potential role of Ago in mitochondrial translation was performed (see Section 3.5.14). After knockdown of Ago2 and inhibition of the cellular translation machinery by emetine, mitochondrial proteins were radioactively labelled. When the pattern

of newly translated mitochondrial proteins were compared between Ago knockdown and control cells, no influence on the expression level of mitochondria encoded proteins could be found. This suggests no involvement of Ago proteins in the mitochondrial translation machinery. This is in agreement with data from MEF and MEF-Ago2-knockout cells, which showed no change in the translation level of mitochondrial proteins transfected with control siRNAs, whereas enhanced mitochondrial translation was observed upon miR-1 transfection [Zhang *et al.*, 2014].
In addition, the knockdown of Ago proteins was controlled by western blot. Here, the expression level of Ago1 and Ago2 were nicely reduced. Notably, a slower migrating protein detected by the Ago2 (11A9) antibody, was not influenced by the siRNA treatment. The visibility of the band was also reduced in the presence of the correct Ago2 epitope under control conditions. This supports the idea that there are cross-reacting epitopes expressed in the cell, which are recognised by the 11A9 antibody and might accumulate in mitochondrial preparations as observed in the membrane association or the Percoll assay (compare Figure 4.30, 4.29 F2, F3). An alternative explanation for the multiple bands detected by the 11A9 antibody may, however, lay in posttranslational modifications of Ago2 [Sahin *et al.*, 2014].

5.5 Conclusion on Mitochondrial Localisation

The above discussed experiments on the protein level do not yield conclusive support for a mitochondrial matrix localisation in HEK293 cells, which is further supported by observation of normal mitochondrial translation also under knockdown of Ago1 and Ago2. This is in line with the absence of Ago proteins from large mass spectrometric analyses of the mitochondrial proteome [Pagliarini *et al.*, 2008]. However, Ago import into mitochondria could also be a cell type specific event [Zhang *et al.*, 2014]. The association of Ago proteins with the ER is supported by the membrane association assay and the percoll gradient and stands in agreement with the findings of Cikaluk *et al.* [1999]; Stalder *et al.* [2013]. The strong methylation signature from Dam-Ago on mitochondrial DNA, compared to Dam or Dam-Cbx1 in MRC-5 or mESC cells might be explained by the finding of Huang *et al.* [2011], who showed that mitochondria associate with P-bodies, a main localisation site of Ago proteins [Sen and Blau, 2005; Rüdel *et al.*,

2008]. Here, the inactivation of mitochondria decreased the efficiency of miRNA-mediated gene silencing [Huang *et al.*, 2011]. Therefore, one can assume a closer spatial association of Dam-Ago with mitochondria compared to Dam or the nuclear Dam-Cbx1.

An alternative idea based on coinciding localisation is related to Ago and mitochondrial degradation. Recently, degradation of Ago by the autophagosome was described [Gibbings *et al.*, 2012]. SiRNA knockdown of autophagosome subunits (Atg5-8) or the autophagy receptor NDP52 resulted in elevated protein levels of Ago1 and Ago2. A similar behaviour was also seen by chemical inhibition of the lysosomal acidification, which blocks autophagy [Gibbings *et al.*, 2012]. The main mechanism for control of the mitochondria population in the cell is mitophagy. In the best studied pathway, mitochondria with aberrant membrane potential are recognized by the E3-ligase Parkin, which is dependent on PTEN-induced putative kinase 1 (PINK1) and polyubiquitilates mitochondrial outer membrane proteins. This leads to proteasomal digest of the outer membrane proteins and is followed by recruitment of the autophagosome [Mishra and Chan, 2014]. Therefore, Dam-Ago could associate with mitochondria inside autophagosomes, gaining access to mtDNA. This process, however, seems rather unlikely, as the inner mitochondrial membrane is still intact, despite its loss of membrane potential, when the autophagosome is recruited. Shortly after, the mitophagosome acidifies and fuses with lysosomes and hydrolyses its content within about 10 minutes [Lemasters, 2014]. It can be assumed, that Dam-Ago will denature already during acidification of the mitophagosome and is thus unable to methylate mitochondrial DNA in this compartment.

Finally, Dam-Ago proteins, which are predominantly cytoplasmic, could methylate the abundant mitochondrial DNA during cell lysis. This is dependent on the lysis buffers used and needs further testing.

The effect of mitochondrial DNA methylation seemed even more pronounced in the stably transfected mESC cells as compared to the lentiviral transfected MRC-5 cells, which express Dam-Ago only for 72 hours. However, this might be caused by the integration site or multiple insertions of the different transgenes. In the case of mESC, a clonal cell line was produced, where Dam-Ago was readily detected by RT-PCR. Thus, a region enhancing transcription of the transgene might have been selected, while in the case of MRC-5 a random pool of integration events was examined.

5.6 Conclusion and Outlook

This study aimed to identify chromatin association sites of human Argonaute proteins by applying the two high throughput techniques ChIP-Seq and DAMID-Seq. Even though, peaks could be identified, there was no agreement on Ago-chromatin associations sites between the different studies. Unfortunately, so far none of the peaks could be validated, inhibiting further studies on the potential nuclear functions and interactors of Ago proteins.

As outlined for ChIP-Seq (Section 5.1) and DAMID experiments (Section 5.2.1), all three experiments suffered from rather low genomic coverage, resulting in low signal-to-noise ratios, complicating peak detection. For the latter two, this was found to correlate with a high mitochondrial read count (Section 5.2.1). As discussed in Section 5.3, we have no evidence for mitochondrial matrix localisation of Ago proteins. Therefore, we conclude non-specific methylation of mitochondrial sequences by cytoplasmic Dam-Ago proteins leading to distortion of the library complexity as observed in the NGS experiments. Consequently, we ask the question, if the mitochondrial DNA sequences in the NGS-libraries are caused by the lack of chromatin binding sites of Ago proteins or if these sequences compromised the NGS-libraries in a way that prohibits identification of true Ago-chromatin interaction sites. As we found overlapping peaks of the DAM-ID experiment with a study conducted in WI-38 cells [Benhamed *et al.*, 2012], it seems rather likely that the mitochondrial DNA sequences negatively influenced the NGS-libraries and thus detection of potential nuclear Ago-chromatin interaction sites.

To successfully identify Ago-chromatin binding sites improved DAM-ID experiments as suggested in Section 5.2.1 or alternative approaches like DAM-IP could be used [Xiao *et al.*, 2010]. DAM-IP is based on expression of a modified Dam-fusion protein, which has a broader substrate specificity (recognition site: ATC) and leads to an increased methylation rate and lower distance between methylation sites. The combination of the DNA-methylation with an IP approach using a m6A-specific antibody additionally improves the signal-to-noise ratio.

In general, many functions have been proposed for nuclear Ago proteins, ranging form TGS to DNA DSB repair (Section 1.5.7), which are sometimes most controversially discussed. On the one hand, all these observations could be true suggesting a stunning versatility of Ago proteins in the nucleus, compared to its

exclusive function in the cytoplasm. On the other hand, one may argue that there are only one or two true nuclear functions of Ago proteins, for example DNA DSB repair, which, if perturbed or wrongly guided by i.e. promoter complementary siRNAs, may lead to the observed results. Therefore, more experimental work is needed involving critical assessment of the proposed pathways and search for unifying principles, which may help to refine the current picture to a concise nuclear function of Ago proteins.

Bibliography

B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell.* Garland Science, New York, 4th edition, 2002.

M. Allo, V. Buggiano, J. P. Fededa, E. Petrillo, I. Schor, M. de la Mata, E. Agirre, M. Plass, E. Eyras, S. A. Elela, R. Klinck, B. Chabot, and A. R. Kornblihtt. July 2009. Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat Struct Mol Biol*, 16(7):717–724.

S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403 – 410.

M. Ameyar-Zazoua, C. Rachez, M. Souidi, P. Robin, L. Fritsch, R. Young, N. Morozova, R. Fenouil, N. Descostes, J.-C. Andrau, J. Mathieu, A. Hamiche, S. Ait-Si-Ali, C. Muchardt, E. Batsch, and A. Harel-Bellan. 2012. Argonaute proteins couple chromatin silencing to alternative splicing. *Nat Struct Mol Biol*, 19(10): 998–1004.

S. Anders and W. Huber. 2010. Differential expression analysis for sequence count data. *Genome Biology*, 11(10):R106–.

S. Andrews. FastQC A Quality Control tool for High Throughput Sequence Data.

A. Aravin, M. Lagos-Quintana, A. Yalcin, M. Zavolan, D. Marks, B. Snyder, T. Gaasterland, J. Meyer, and T. Tuschl. 2003. The Small RNA Profile during Drosophila melanogaster Development. *Developmental Cell*, 5(2):337 – 350.

A. Aravin, R. Sachidanandam, D. Bourc'his, C. Schaefer, D. Pezic, K. F. Toth, T. Bestor, and G. J. Hannon. 2008. A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice. *Molecular Cell*, 31(6): 785–799.

A. A. Aravin, R. Sachidanandam, A. Girard, K. Fejes-Toth, and G. J. Hannon. 2007. Developmentally Regulated piRNA Clusters Implicate MILI in Transposon Control. *Science*, 316(5825):744–747.

L. Aravind, H. Watanabe, D. J. Lipman, and E. V. Koonin. 2000. Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proceedings of the National Academy of Sciences*, 97(21):11319–11324.

A. Ashe, A. Sapetschnig, E.-M. Weick, J. Mitchell, M. Bagijn, A. Cording, A.-L. Doebley, L. Goldstein, N. Lehrbach, J. LeÂ Pen, G. Pintacuda, A. Sakaguchi, P. Sarkies, S. Ahmed, and E. Miska. 2012. piRNAs Can Trigger a Multigenerational Epigenetic Memory in the Germline of C.Â elegans. *Cell*, 150(1):88 – 99.

D. C. Avgousti, S. Palani, Y. Sherman, and A. Grishok. 2012. CSR-1 RNAi pathway positively regulates histone expression in C. elegans. *The EMBO Journal*, 31(19):3821–3832.

M. Baer, T. W. Nilsen, C. Costigan, and S. Altman. Jan. 1990. Structure and transcription of a human gene for H1 RNA, the RNA component of human RNase P. *Nucleic Acids Research*, 18(1):97–103.

M. P. Bagijn, L. D. Goldstein, A. Sapetschnig, E.-M. Weick, S. Bouasker, N. J. Lehrbach, M. J. Simard, and E. A. Miska. 2012. Function, Targets, and Evolution of Caenorhabditis elegans piRNAs. *Science*, 337(6094):574–578.

S. Bandiera, S. RÃŒberg, M. Girard, N. Cagnard, S. Hanein, D. ChrÃ©tien, A. Munnich, S. Lyonnet, and A. Henrion-Caude. June 2011. Nuclear Outsourcing of RNA Interference Components to Human Mitochondria. *PLoS ONE*, 6 (6):e20746–.

A. J. Bannister, P. Zegerman, J. F. Partridge, E. A. Miska, J. O. Thomas, R. C. Allshire, and T. Kouzarides. Mar. 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*, 410(6824):120–124.

E. Barrey, G. Saint-Auret, B. Bonnamy, D. Damas, O. Boyer, and X. Gidrol. May 2011. Pre-microRNA and Mature microRNA in Human Mitochondria. *PLoS ONE*, 6(5):e20220–.

P. J. Batista, J. G. Ruby, J. M. Claycomb, R. Chiang, N. Fahlgren, K. D. Kasschau, D. A. Chaves, W. Gu, J. J. Vasale, S. Duan, D. C. Jr., S. Luo, G. P. Schroth, J. C. Carrington, D. P. Bartel, and C. C. Mello. 2008. PRG-1 and 21U-RNAs Interact to Form the piRNA Complex Required for Fertility in C. elegans. *Molecular Cell*, 31(1):67 – 78.

E. H. Bayne, S. A. White, A. Kagansky, D. A. Bijos, L. Sanchez-Pulido, K.-L. Hoe, D.-U. Kim, H.-O. Park, C. P. Ponting, J. Rappsilber, and R. C. Allshire. 2010. Stc1: A Critical Link between RNAi and Chromatin Modification Required for Heterochromatin Integrity. *Cell*, 140(5):666–677.

M. Beitzinger, L. Peters, J. Y. Zhu, E. Kremmer, and G. Meister. 2007. Identification of Human microRNA Targets From Isolated Argonaute Protein Complexes. *RNA Biology*, 4(2):76–84.

M. Benhamed, U. Herbig, T. Ye, A. Dejean, and O. Bischof. Mar. 2012. Senescence is an endogenous trigger for microRNA-directed transcriptional gene silencing in human cells. *Nat Cell Biol*, 14(3):266–275.

I. Berger, D. J. Fitzgerald, and T. J. Richmond. Dec. 2004. Baculovirus expression system for heterologous multiprotein complexes. *Nat Biotech*, 22(12):1583–1587.

E. Bernstein, A. A. Caudy, S. M. Hammond, and G. J. Hannon. Jan. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, 409 (6818):363–366.

Z. Bian, L.-M. Li, R. Tang, D.-X. Hou, X. Chen, C.-Y. Zhang, and K. Zen. Sept. 2010. Identification of mouse liver mitochondria-associated miRNAs and their potential biological functions. *Cell Res*, 20(9):1076–1078.

Biorad. 2005. Personal Molecular Imager™ System-Hardware Instruction Manual for Catalog Number 170-9400. pages 1–32.

K. Bohmert, I. Camus, C. Bellini, D. Bouchez, M. Caboche, and C. Benning. Jan. 1998. AGO1 defines a novel locus of Arabidopsis controlling leaf development. *The EMBO Journal*, 17(1):170–180.

A. Boland, F. Tritschler, S. Heimstädt, E. Izaurralde, and O. Weichenrieder. 2010. Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. *EMBO reports*, 11(7):522–527.

A. M. Bolger, M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, pages –.

N. G. Bologna and O. Voinnet. 2014. The Diversity, Biogenesis, and Activities of Endogenous Silencing Small RNAs in Arabidopsis. *Annual Review of Plant Biology*, 65(1):473–503. PMID: 24579988.

J. Braun, E. Huntzinger, M. Fauser, and E. Izaurralde. 2011. GW182 Proteins Directly Recruit Cytoplasmic Deadenylase Complexes to miRNA Targets. *Molecular Cell*, 44(1):120 – 133.

R. A. H. S. A. C. B. J. Braunstein, M. Apr 1993. Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. *Genes Dev.*, 7(4):592–604.

J. Brennecke, A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, and G. J. Hannon. 2007. Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. *Cell*, 128(6):1089 – 1103.

B. Brower-Toland, S. D. Findley, L. Jiang, L. Liu, H. Yin, M. Dus, P. Zhou, S. C. Elgin, and H. Lin. 2007. Drosophila PIWI associates with chromatin and interacts directly with HP1a. *Genes & Development*, 21(18):2300–2311.

B. A. Buckley, K. B. Burkhart, S. G. Gu, G. Spracklin, A. Kershner, H. Fritz, J. Kimble, A. Fire, and S. Kennedy. Sept. 2012. A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. *Nature*, 489 (7416):447–451.

M. Bühler, A. Verdel, and D. Moazed. 2006. Tethering RITS to a Nascent Transcript Initiates RNAi- and Heterochromatin-Dependent Gene Silencing. *Cell*, 125(5):873–886.

K. B. Burkhart, S. Guang, B. A. Buckley, L. Wong, A. F. Bochner, and S. Kennedy. 08 2011. A Pre-mRNAâAssociating Factor Links Endogenous siRNAs to Chromatin Regulation. *PLoS Genet*, 7(8):e1002249.

M. A. Carmell, A. Girard, H. J. van de Kant, D. Bourc'his, T. H. Bestor, D. G. de Rooij, and G. J. Hannon. 2007. {MIWI2} Is Essential for Spermatogenesis and Repression of Transposons in the Mouse Male Germline. *Developmental Cell*, 12(4):503 – 514.

M. A. Carmell, Z. Xuan, M. Q. Zhang, and G. J. Hannon. Nov. 2002. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes & Development*, 16(21):2733–2742.

J. A. Casbon, R. J. Osborne, S. Brenner, and C. P. Lichtenstein. 2011. A method for counting PCR template molecules with application to next-generation sequencing. *Nucleic Acids Research*, 39(12):e81.

G. Cecere, S. Hoersch, S. O'Keeffe, R. Sachidanandam, and A. Grishok. Apr. 2014. Global effects of the CSR-1 RNA interference pathway on the transcriptional landscape. *Nat Struct Mol Biol*, 21(4):358–365.

G. Cecere, G. Zheng, A. Mansisidor, K. Klymko, and A. Grishok. 2012. Promoters Recognized by Forkhead Proteins Exist for Individual 21U-RNAs. *Molecular Cell*, 47(5):734 – 745.

F. M. Cernilogar, M. C. Onorati, G. O. Kothe, A. M. Burroughs, K. M. Parsi, A. Breiling, F. L. Sardo, A. Saxena, K. Miyoshi, H. Siomi, M. C. Siomi, P. Carninci, D. S. Gilmour, D. F. V. Corona, and V. Orlando. Dec. 2011. Chromatin-associated RNA interference components contribute to transcriptional regulation in Drosophila. *Nature*, 480(7377):391–395.

S. W.-L. Chan, D. Zilberman, Z. Xie, L. K. Johansen, J. C. Carrington, and S. E. Jacobsen. 2004. RNA Silencing Genes Control de Novo DNA Methylation. *Science*, 303(5662):1336.

M. Chekulaeva, H. Mathys, J. T. Zipprich, J. Attig, M. Colic, R. Parker, and W. Filipowicz. Nov. 2011. miRNA repression involves GW182-mediated recruitment of CCR4Ã¢ÂÂNOT through conserved W-containing motifs. *Nat Struct Mol Biol*, 18(11):1218–1226.

S. Cheloufi, C. O. Dos Santos, M. M. W. Chong, and G. J. Hannon. June 2010. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature*, 465(7298):584–589.

C.-Y. A. Chen, D. Zheng, Z. Xia, and A.-B. Shyu. Nov. 2009. Ago-TNRC6 triggers microRNA-mediated decay by promoting two deadenylation steps. *Nat Struct Mol Biol*, 16(11):1160–1166.

L. Chen, J. E. Dahlstrom, S.-H. Lee, and D. Rangasamy. July 2012. Naturally occurring endo-siRNA silences LINE-1 retrotransposons in human cells through DNA methylation. *epigenetics*, 7(1559-2294):758–771.

T. Chen and S. Y. R. Dent. Feb. 2014. Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nat Rev Genet*, 15(2):93–106.

S. Cho, J. S. Park, and Y.-K. Kang. 2014. AGO2 and SETDB1 cooperate in promoter-targeted transcriptional silencing of the androgen receptor gene. *Nucleic Acids Research*.

Y. Chu, X. Yue, S. T. Younger, B. A. Janowski, and D. R. Corey. 2010. Involvement of argonaute proteins in gene silencing and activation by RNAs complementary to a non-coding transcript at the progesterone receptor promoter. *Nucleic Acids Research*, 38(21):7736–7748.

C. Ciaudo, F. Jay, I. Okamoto, C. Chen, A. Sarazin, N. Servant, E. Barillot, E. Heard, and O. Voinnet. 2013. RNAi-Dependent and Independent Control of LINE1 Accumulation and Mobility in Mouse Embryonic Stem Cells. *PLOS Genetics*, 9(11).

D. Cifuentes, H. Xue, D. W. Taylor, H. Patnode, Y. Mishima, S. Cheloufi, E. Ma, S. Mane, G. J. Hannon, N. D. Lawson, S. A. Wolfe, and A. J. Giraldez. 2010. A Novel miRNA Processing Pathway Independent of Dicer Requires Argonaute2 Catalytic Activity. *Science*, 328(5986):1694–1698.

D. E. Cikaluk, N. Tahbaz, L. C. Hendricks, G. E. DiMattia, D. Hansen, D. Pilgrim, and T. C. Hobman. Oct. 1999. GERp95, a Membrane-associated Protein that Belongs to a Family of Proteins Involved in Stem Cell Differentiation. *Molecular Biology of the Cell*, 10(10):3357–3372.

J. M. Claycomb, P. J. Batista, K. M. Pang, W. Gu, J. J. Vasale, J. C. van Wolfswinkel, D. A. Chaves, M. Shirayama, S. Mitani, R. F. Ketting, D. C. Jr., and C. C. Mello. 2009. The Argonaute CSR-1 and Its 22G-RNA Cofactors Are Required for Holocentric Chromosome Segregation. *Cell*, 139(1):123 – 134.

T. Dalmay, R. Horsefield, T. H. Braunstein, and D. C. Baulcombe. 2001. SDE3 encodes an RNA helicase required for post-transcriptional gene silencing in Arabidopsis. *The EMBO Journal*, 20(8):2069–2077.

P. Das, M. Bagijn, J. Goldstein, L andWoolford, N. Lehrbach, A. Sapetschnig, H. Buhecha, M. Gilchris, K. Howe, R. Stark, N. Matthews, E. Berezikov, R. Ketting, S. Tavaré, and E. Miska. 2008. Piwi and piRNAs Act Upstream of an Endogenous siRNA Pathway to Suppress Tc3 Transposon Mobility in the Caenorhabditis elegans Germline. *Molecular Cell*, 31(1):79 – 90.

P. Das, K. Ramachandran, J. van Wert, and R. Singal. 2004. Chromatin immunoprecipitation assay. *BioTechniques*, 37:961–969.

S. Das, M. Ferlito, O. A. Kent, K. Fox-Talbot, R. Wang, D. Liu, N. Raghavachari, Y. Yang, S. J. Wheelan, E. Murphy, and C. Steenbergen. 2012. Nuclear miRNA Regulates the Mitochondrial Genome in the Heart. *Circulation Research*, 110(12): 1596–1603.

W. Deng and H. Lin. 2002. Miwi, a Murine Homolog of Piwi, Encodes a Cytoplasmic Protein Essential for Spermatogenesis. *Developmental Cell*, 2(6):819 – 830.

H. Denis, M. N. Ndlovu, and F. Fuks. May 2011. Regulation of mammalian DNA methyltransferases: a route to new mechanisms. *EMBO Reports*, 12(7):647–656.

G. Deshpande, G. Calhoun, and P. Schedl. 2005. Drosophila argonaute-2 is required early in embryogenesis for the assembly of centric/centromeric heterochromatin, nuclear division, nuclear migration, and germ-cell formation. *Genes & Development*, 19(14):1680–1685.

R. M. R. R. Dignam, J.D. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Research*, 11(5):1476–1489.

M. Doyle, L. Badertscher, L. Jaskiewicz, S. GÃŒttinger, S. Jurado, T. Hugenschmidt, U. Kutay, and W. Filipowicz. 2013. The double-stranded RNA binding domain of human Dicer functions as a nuclear localization signal. *RNA*, 19(9): 1238–1252.

A. Dueck and G. Meister. 2014. Assembly and function of small RNA Argonaute protein complexes. *Biol. Chem*, 395(6):611–629.

M. Durand-Dubief and P. Bastin. 2003. TbAGO1, an Argonaute protein required for RNA interference, is involved in mitosis and chromosome segregation in Trypanosoma brucei. *BMC Biology*, 1(1):2.

M. El-Shami, D. Pontier, S. Lahmy, L. Braun, C. Picart, D. Vega, M.-A. Hakimi, S. E. Jacobsen, R. Cooke, and T. Lagrange. 2007. Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes & Development*, 21(20):2539–2544.

S. M. Elbashir, J. Harborth, W. Lendeckel, A. Yalcin, K. Weber, and T. Tuschl. May 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, 411(6836):494–498.

O. Emanuelsson, H. Nielsen, S. Brunak, and G. von Heijne. 2000. Predicting Subcellular Localization of Proteins Based on their N-terminal Amino Acid Sequence. *Journal of Molecular Biology*, 300(4):1005 – 1016.

C. Ender and G. Meister. 2010. Argonaute proteins at a glance. *Journal of Cell Science*, 123(11):1819–1823.

C. Faehnle and L. Joshua-Tor. 2010. Argonaute MID domain takes centre stage. *EMBO Reports*, 11(8):564–565.

A. A. Farooqi, Z. u. Rehman, and J. Muntane. Nov. 2014. Antisense therapeutics in oncology: current status. *OncoTargets and therapy*, 7:2035–2042.

P. S. Fey E G, Krochmalnic G. 1986. The Nonchromatin Substructures of the Nucleus: The Ribonucleoprotein (RNP)-containing and RNP-depleted Matrices Analyzed by Sequential Fractionation and Resinless Section Electron Microscopy. *The Journal of Cell Biology*, 102:1654–1665.

D. Finlay, D. Newmeyer, T. Price, and D. Forbes. 1987. Inhibition of in vitro nuclear transport by a lectin that binds to nuclear pores. *The Journal of Cell Biology*, 104:pp189–200.

A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, and C. C. Mello. Feb. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, 391(6669):806–811.

D. J. Fitzgerald, P. Berger, C. Schaffitzel, K. Yamada, T. J. Richmond, and I. Berger. Dec. 2006. Protein complex expression by using multigene baculoviral vectors. *Nat Meth*, 3(12):1021–1032.

S. Francia, F. Michelini, A. Saxena, D. Tang, M. de Hoon, V. Anelli, M. Mione, P. Carninci, and F. d'Adda di Fagagna. Aug. 2012. Site-specific DICER and DROSHA RNA products control the DNA-damage response. *Nature*, 488(7410): 231–235.

F. Frank, N. Sonenberg, and B. Nagar. June 2010. Structural basis for 5′-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature*, 465(7299): 818–822.

J. K. Friedrich, K. I. Panov, P. Cabart, J. Russell, and J. C. B. M. Zomerdijk. 2005. TBP-TAF Complex SL1 Directs RNA Polymerase I Pre-initiation Complex Formation and Stabilizes Upstream Binding Factor at the rDNA Promoter. *Journal of Biological Chemistry*, 280(33):29551–29558.

Y. Fujiki, A. L. Hubbard, S. Fowler, and P. B. Larzarow. 1982. Isolation of intracellular membranes by means of sodium crabonate treatment: appliaction to endoplasmatic reticulum. *Journal of Cell Biology*, 93(1):97–102.

K. Gagnon, L. Li, Y. Chu, B. A. Janowski, and D. Corey. 2014. RNAi Factors Are Present and Active in Human Cell Nuclei. *Cell Reports*, 6(1):211–221.

K. T. Gagnon and D. Corey. 2012. Argonaute and the Nuclear RNAs: New Pathways for RNA-Mediated Control of Gene Expression. *Nucleic acid Therapeutics*, 22(1):3–16.

M. Gao, W. Wei, M.-M. Li, Y.-S. Wu, Z. Ba, K.-X. Jin, M.-M. Li, Y.-Q. Liao, S. Adhikari, Z. Chong, T. Zhang, C.-X. Guo, T.-s. Tang, B.-T. Zhu, X.-Z. Xu, N. Mailand, Y.-G. Yang, Y. Qi, and J. M. R. Danielsen. May 2014. Ago2 facilitates Rad51 recruitment and DNA double-strand break repair by homologous recombination. *Cell Res*, 24(5):532–541.

M. Ghildiyal, H. Seitz, M. D. Horwich, C. Li, T. Du, S. Lee, J. Xu, E. L. Kittler, M. L. Zapp, Z. Weng, and P. D. Zamore. 2008. Endogenous siRNAs Derived from Transposons and mRNAs in Drosophila Somatic Cells. *Science*, 320(5879): 1077–1081.

D. Gibbings, S. Mostowy, F. Jay, Y. Schwab, P. Cossart, and O. Voinnet. Dec. 2012. Selective autophagy degrades DICER and AGO2 and regulates miRNA activity. *Nat Cell Biol*, 14(12):1314–1321.

L.-T. Gou, P. Dai, J.-H. Yang, Y. Xue, Y.-P. Hu, Y. Zhou, J.-Y. Kang, X. Wang, H. Li, M.-M. Hua, S. Zhao, S.-D. Hu, L.-G. Wu, H.-J. Shi, Y. Li, X.-D. Fu, L.-H. Qu, E.-D. Wang, and M.-F. Liu. June 2014. Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. *Cell Res*, 24(6):680–700.

M. R. Green and S. J. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, 4th edition, 2012.

R. I. Gregory, K.-p. Yan, G. Amuthan, T. Chendrimada, B. Doratotaj, N. Cooch, and R. Shiekhattar. Nov. 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature*, 432(7014):235–240.

F. Greil, C. Moorman, and B. van Steensel. DamID: Mapping of In Vivo Protein-Genome Interactions Using Tethered DNA Adenine Methyltransferase. In A. Kimmel and B. Oliver, editors, *DNA Microarrays, Part A: Array Platforms and Wet-Bench Protocols*, volume Volume 410, pages 342–359. Academic Press, 2006.

S. I. S. Grewal and S. C. R. Elgin. May 2007. Transcription and RNA interference in the formation of heterochromatin. *Nature*, 447(7143):399–406.

A. Grishok. Chapter One - Biology and Mechanisms of Short {RNAs} in Caenorhabditis elegans. volume 83 of *Advances in Genetics*, pages 1 – 69. Academic Press, 2013.

A. Grishok, A. E. Pasquinelli, D. Conte, N. Li, S. Parrish, I. Ha, D. L. Baillie, A. Fire, G. Ruvkun, and C. C. Mello. 2001. Genes and Mechanisms Related to RNA Interference Regulate Expression of the Small Temporal RNAs that Control C. elegans Developmental Timing. *Cell*, 106(1):23–34.

S. T. Grivna, B. Pyhtila, and H. Lin. 2006. MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis. *Proceedings of the National Academy of Sciences*, 103(36):13415–13420.

S. G. Gu, J. Pak, S. Guang, J. M. Maniar, S. Kennedy, and A. Fire. Feb. 2012. Amplification of siRNA in Caenorhabditis elegans generates a transgenerational

sequence-targeted histone H3 lysine 9 methylation footprint. *Nat Genet*, 44(2): 157–164.

W. Gu, M. Shirayama, D. C. Jr., J. Vasale, P. J. Batista, J. M. Claycomb, J. J. Moresco, E. M. Youngman, J. Keys, M. J. Stoltz, C.-C. G. Chen, D. A. Chaves, S. Duan, K. D. Kasschau, N. Fahlgren, J. R. Y. III, S. Mitani, J. C. Carrington, and C. C. Mello. 2009. Distinct Argonaute-Mediated 22G-RNA Pathways Direct Genome Surveillance in the C. elegans Germline. *Molecular Cell*, 36(2):231 – 244.

S. Guang, A. F. Bochner, K. B. Burkhart, N. Burton, D. M. Pavelec, and S. Kennedy. June 2010. Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. *Nature*, 465(7301):1097–1101.

S. Guang, A. F. Bochner, D. M. Pavelec, K. B. Burkhart, S. Harding, J. Lachowiec, and S. Kennedy. 2008. An Argonaute Transports siRNAs from the Cytoplasm to the Nucleus. *Science*, 321(5888):537–541.

L. S. Gunawardane, K. Saito, K. M. Nishida, K. Miyoshi, Y. Kawamura, T. Nagami, H. Siomi, and M. C. Siomi. 2007. A Slicer-Mediated Mechanism for Repeat-Associated siRNA 5′ End Formation in Drosophila. *Science*, 315(5818): 1587–1590.

M. Ha and V. N. Kim. Aug. 2014. Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol*, 15(8):509–524.

M. Halic and D. Moazed. 2010. Dicer-Independent Primal RNAs Trigger RNAi and Heterochromatin Formation. *Cell*, 140(4):504–516.

A. J. Hamilton and D. C. Baulcombe. 1999. A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants. *Science*, 286(5441):950–952.

S. M. Hammond, E. Bernstein, D. Beach, and G. J. Hannon. Mar. 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature*, 404(6775):293–296.

J. Han, Y. Lee, K.-H. Yeom, J.-W. Nam, I. Heo, J.-K. Rhee, S. Y. Sohn, Y. Cho, B.-T. Zhang, and V. N. Kim. 2006. Molecular Basis for the Recognition of Primary microRNAs by the Drosha-DGCR8 Complex. *Cell*, 125(5):887–901.

T. B. Hansen, E. D. Wiklund, J. B. Bramsen, S. B. Villadsen, A. L. Statham, S. J. Clark, and J. Kjems. Aug. 2011. miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *The EMBO Journal*, 30 (21):4414–4422.

J. Hauptmann, A. Dueck, S. Harlander, J. Pfaff, R. Merkl, and G. Meister. 2013. Turning catalytically inactive human Argonaute proteins into active slicer enzymes. *Nat Struct Mol Biol*, 20(7):814–817.

C. He, S. S. Pillai, F. Taglini, F. Li, K. Ruan, J. Zhang, J. Wu, Y. Shi, and E. H. Bayne. 2013. Structural analysis of Stc1 provides insights into the coupling of RNAi and chromatin modification. *Proceedings of the National Academy of Sciences*, 110 (21):E1879âE1888.

D. Holoch and D. Moazed. Feb. 2015. RNA-mediated epigenetic regulation of gene expression. *Nat Rev Genet*, 16(2):71–84.

M. Hooper, K. Hardy, A. Handyside, S. Hunter, and M. Monk. 1987. HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature*, 326(6110):292–295.

M. D. Horwich, C. Li, C. Matranga, V. Vagin, G. Farley, P. Wang, and P. D. Zamore. 2007. The Drosophila RNA Methyltransferase, DmHen1, Modifies Germline piRNAs and Single-Stranded siRNAs in RISC. *Current Biology*, 17(14):1265 – 1272.

J.-P. Hsin and J. L. Manley. 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes & Development*, 26(19):2119–2137.

L. Huang, S. Mollet, S. Souquere, F. Le Roy, M. Ernoult-Lange, G. Pierron, F. Dautry, and D. Weil. July 2011. Mitochondria Associate with P-bodies and Modulate MicroRNA-mediated RNA Interference. *Journal of Biological Chemistry*, 286(27):24219–24230.

V. Huang and L.-C. Li. 2014. Demystifying the nuclear function of Argonaute proteins. *RNA Biology*, 11(1):18–24.

V. Huang, J. Zheng, Z. Qi, J. Wang, R. Place, J. Yu, H. Li, and L.-C. Li. 2013a. Ago1 Interacts with RNA Polymerase II and Binds to the Promoters of Actively Transcribed Genes in Human Cancer Cells. *PLoS Genetics*, 9(9).

X. Huang, H. Yin, S. Sweeney, D. Raha, M. Snyder, and H. Lin. 2013b. A Major Epigenetic Programming Mechanism Guided by piRNAs. *Developmental Cell*, 24(5):502 – 516.

E. Huntzinger and E. Izaurralde. 2011. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nature Reviews Genetics*, 12 (2):99–110.

G. Hutvagner, J. McLachlan, A. E. Pasquinelli, E. Bálint, T. Tuschl, and P. D. Zamore. 2001. A Cellular Function for the RNA-Interference Enzyme Dicer in the Maturation of the let-7 Small Temporal RNA. *Science*, 293(5531):834–838.

G. Hutvagner and M. J. Simard. Jan. 2008. Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol*, 9(1):22–32.

L. T. Invitrogen. 2000. Bac-to-Bac Baculoviorus Expression Systems Manual. pages –.

L. T. Invitrogen. 2002. Guide to Baculovirus Expression Vector Systems (BEVS) and Insect Cell Culture Techniques.

Y. W. Iwasaki, M. C. Siomi, and H. Siomi. 2015. PIWI-Interacting RNA: Its Biogenesis and Functions. *Annual Review of Biochemistry*, 84(1):null. PMID: 25747396.

B. Janowski, K. Huffman, J. Schwartz, a. N. R. Ram, R., J. Shames, D.S.and Minna, and D. Corey. Sep 2006. Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nature Structural & Molecular Biology*, 13(9):787–792.

B. A. Janowski, K. E. Huffman, J. C. Schwartz, R. Ram, D. Hardy, D. S. Shames, J. D. Minna, and D. R. Corey. Sept. 2005. Inhibiting gene expression at transcription start sites in chromosomal DNA with antigene RNAs. *Nat Chem Biol*, 1(4):216–222.

B. A. Janowski, S. T. Younger, D. B. Hardy, R. Ram, K. E. Huffman, and D. R. Corey. Mar. 2007. Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nat Chem Biol*, 3(3):166–173.

A. Jimenez, L. Carrasco, and D. Vazquez. Oct. 1977. Enzymic and nonenzymic translocation by yeast polysomes. Site of action of a number of inhibitors. *Biochemistry*, 16(21):4727–4730.

B.-T. Juang, C. Gu, L. Starnes, F. Palladino, A. Goga, S. Kennedy, and N. D. L'Etoile. Aug. 2013. Endogenous Nuclear RNAi Mediates Behavioral Adaptation to Odor. *Cell*, 154(5):1010–1022.

P. Kafasla, I. Mickleburgh, M. Llorian, M. Coelho, C. Gooding, D. Cherny, A. Joshi, O. Kotik-Kogan, S. Curry, I. C. Eperon, R. J. Jackson, and C. W. Smith. 2012. Defining the roles and interactions of PT. *Biochemical Society Transactions*, pages 815–820.

V. Kaltimbacher, C. Bonnet, G. Lecoeuvre, V. Forster, J.-A. Sahel, and M. Corral-Debrinski. 2006. mRNA localization to the mitochondrial surface allows the efficient translocation inside the organelle of a nuclear recoded ATP6 protein. *RNA*, 12(7):1408–1417.

C. Kanellopoulou, S. A. Muljo, A. L. Kung, S. Ganesan, R. Drapkin, T. Jenuwein, D. M. Livingston, and K. Rajewsky. 2005. Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes & Development*, 19(4):489–501.

J. Kanoh, M. Sadaie, T. Urano, and F. Ishikawa. 2005. Telomere Binding Protein Taz1 Establishes Swi6 Heterochromatin Independently of RNAi at Telomeres. *Current Biology*, 15(20):1808–1819.

K. Karmodiya, A. Krebs, M. Oulad-Abdelghani, H. Kimura, and L. Tora. 2012. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC Genomics*, 13(1):424.

Y. Kawamura, K. Saito, T. Kin, Y. Ono, K. Asai, T. Sunohara, T. N. Okada, M. C. Siomi, and H. Siomi. June 2008. Drosophila endogenous small RNAs bind to Argonaute[thinsp]2 in somatic cells. *Nature*, 453(7196):793–797.

S. P. Keam, P. E. Young, A. L. McCorkindale, T. H. Dang, J. L. Clancy, D. T. Humphreys, T. Preiss, G. Hutvagner, D. I. Martin, J. E. Cropley, and C. M. Suter. 2014. The human Piwi protein Hiwi2 associates with tRNA-derived piRNAs in somatic cells. *Nucleic Acids Research*, 42(14):8984–8995.

D. H. Kim, P. Saetrom, O. Snove, and J. J. Rossi. 2008. MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proceedings of the National Academy of Sciences*, 105(42):16230–16235.

D. H. Kim, L. M. Villeneuve, K. V. Morris, and J. J. Rossi. Sept. 2006. Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nat Struct Mol Biol*, 13(9):793–797.

H. Kim, B. Erickson, W. Luo, D. Seward, J. H. Graber, D. D. Pollock, P. C. Megee, and D. L. Bentley. Oct. 2010. Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat Struct Mol Biol*, 17(10):1279–1286.

Y. Kirino and Z. Mourelatos. 2007. The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs. *RNA*, 13(9):1397–1401.

B. T. Kren, P. Y.-P. Wong, A. Sarver, X. Zhang, Y. Zeng, and C. J. Steer. Jan. 2009. MicroRNAs identified in highly purified liver-derived mitochondria may play a role in apoptosis. *rnabiology*, 6(1547-6286):65–72.

M. Kuhlmann, A. Finke, M. Mascher, and M. F. Mette. 2014. DNA methylation maintenance consolidates RNA-directed DNA methylation and transcriptional gene silencing over generations in Arabidopsis thaliana. *The Plant Journal*.

S. Kuramochi-Miyagawa, T. Kimura, T. W. Ijiri, T. Isobe, N. Asada, Y. Fujita, M. Ikawa, N. Iwai, M. Okabe, W. Deng, H. Lin, Y. Matsuda, and T. Nakano. 2004. Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development*, 131(4):839–849.

S. Kuramochi-Miyagawa, T. Watanabe, K. Gotoh, Y. Totoki, A. Toyoda, M. Ikawa, N. Asada, K. Kojima, Y. Yamaguchi, T. W. Ijiri, K. Hata, E. Li, Y. Matsuda, T. Kimura, M. Okabe, Y. Sakaki, H. Sasaki, and T. Nakano. 2008. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes & Development*, 22(7):908–917.

U. Laemmli. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227(5259):680–685.

B. Langmead, C. Trapnell, M. Pop, and S. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3):R25–.

J. A. Law, J. Du, C. J. Hale, S. Feng, K. Krajewski, A. M. S. Palanca, B. D. Strahl, D. J. Patel, and S. E. Jacobsen. June 2013. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature*, 498(7454):385–389.

M. Lazarou, S. M. Smith, D. R. Thorburn, M. T. Ryan, and M. McKenzie. Nov. 2009. Assembly of nuclear DNA-encoded subunits into mitochondrial complexÂ IV, and their preferential integration into supercomplex forms in patient mitochondria. *FEBS Journal*, 276(22):6701–6713.

A. Le Thomas, A. K. Rogers, A. Webster, G. K. Marinov, S. E. Liao, E. M. Perkins, J. K. Hur, A. A. Aravin, and K. F. Tóth. 2013. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes & Development*, 27(4):390–399.

Y. Lee, C. Ahn, J. Han, H. Choi, J. Kim, J. Yim, J. Lee, P. Provost, O. Radmark, S. Kim, and V. N. Kim. Sept. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature*, 425(6956):415–419.

Y. Lee, M. Kim, J. Han, K.-H. Yeom, S. Lee, S. H. Baek, and V. N. Kim. Sept. 2004a. MicroRNA genes are transcribed by RNA polymerase II. *The EMBO Journal*, 23 (20):4051–4060.

Y. Lee, K. Nakahara, J. W. Pham, K. Kim, Z. He, E. J. Sontheimer, and R. W. Carthew. 2004b. Distinct Roles for Drosophila Dicer–1 and Dicer–2 in the siRNA/miRNA Silencing Pathways. *Cell*, 117(1):69 – 81.

J. J. Lemasters. 2014. Variants of mitochondrial autophagy: Types 1 and 2 mitophagy and micromitophagy (Type 3). *Redox Biology*, 2(0):749 – 754.

P. J. F. Leuschner, S. L. Ameres, S. Kueng, and J. Martinez. 2006. Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO reports*, 7(3):314–320.

C. Li, V. V. Vagin, S. Lee, J. Xu, S. Ma, H. Xi, H. Seitz, M. D. Horwich, M. Syrzycka, B. M. Honda, E. L. Kittler, M. L. Zapp, C. Klattenhoff, N. Schulz, W. E. Theurkauf, Z. Weng, and P. D. Zamore. 2009a. Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies. *Cell*, 137(3):509 – 521.

H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and . G. P. D. P. Subgroup. 2009b. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079.

L.-C. Li, S. T. Okino, H. Zhao, D. Pookot, R. F. Place, S. Urakami, H. Enokida, and R. Dahiya. 2006. Small dsRNAs induce transcriptional activation in human cells. *Proceedings of the National Academy of Sciences*, 103(46):17337–17342.

A. Lingel, B. Simon, E. Izaurralde, and M. Sattler. Nov. 2003. Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. *Nature*, 426 (6965):465–469.

J. Liu, M. A. Carmell, F. V. Rivas, C. G. Marsden, J. M. Thomson, J.-J. Song, S. M. Hammond, L. Joshua-Tor, and G. J. Hannon. 2004. Argonaute2 Is the Catalytic Engine of Mammalian RNAi. *Science*, 305(5689):1437–1441.

Q. Liu, T. A. Rand, S. Kalidas, F. Du, H.-E. Kim, D. P. Smith, and X. Wang. 2003. R2D2, a Bridge Between the Initiation and Effector Steps of the Drosophila RNAi Pathway. *Science*, 301(5641):1921–1925.

T. Liu. Use Model-Based Analysis of ChIP-Seq (MACS) to Analyze Short Reads Generated by Sequencing ProteinâDNA Interactions in Embryonic Stem Cells. In B. L. Kidder, editor, *Stem Cell Transcriptional Networks*, volume 1150 of *Methods in Molecular Biology*, pages 81–95. Springer New York, 2014.

J. Long and J. Caceres. 2009. The SR protein family of splicing factors: master regulators of gene expression. *Biochemical Journal*, 417.

K. Luger, M. L. Dechassa, and D. J. Tremethick. July 2012. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat Rev Mol Cell Biol*, 13(7):436–447.

P. A. D. D. Luo, R. X. 1998. Rb Interacts with Histone Deacetylase to Repress Transcription. *Cell*, 92(4):463–473.

M. J. Luteijn, P. van Bergeijk, L. J. T. Kaaij, M. V. Almeida, E. F. Roovers, E. Berezikov, and R. F. Ketting. 2012. Extremely stable Piwi-induced gene silencing in Caenorhabditis elegans. *The EMBO Journal*, 31(16):3422–3430.

J.-B. Ma, K. Ye, and D. J. Patel. May 2004. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature*, 429(6989):318–322.

I. J. MacRae, K. Zhou, F. Li, A. Repic, A. N. Brooks, W. Z. Cande, P. D. Adams, and J. A. Doudna. Jan. 2006. Structural Basis for Double-Stranded RNA Processing by Dicer. *Science*, 311(5758):195–198.

A. Mallory and H. Vaucheret. 2010. Form, Function, and Regulation of ARGONAUTE Proteins. *The Plant Cell Online*, 22(12):3879–3889.

M. Marasovic, M. Zocco, and M. Halic. 2013. Argonaute and Triman Generate Dicer-Independent priRNAs and Mature siRNAs to Initiate Heterochromatin Formation. *Molecular Cell*, 52(2):173 – 183.

C. Matranga, Y. Tomari, C. Shin, D. P. Bartel, and P. D. Zamore. 2005. Passenger-Strand Cleavage Facilitates Assembly of siRNA into Ago2-Containing RNAi Enzyme Complexes. *Cell*, 123(4):607–620.

N. Matter, P. Herrlich, and H. Konig. Dec. 2002. Signal-dependent regulation of splicing via phosphorylation of Sam68. *Nature*, 420(6916):691–695.

M. A. Matzke and R. A. Mosher. June 2014. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet*, 15(6):394–408.

A. D. McCue, K. Panda, S. Nuthikattu, S. G. Choudury, E. N. Thomas, and R. K. Slotkin. 2014. ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation. *The EMBO Journal*, 34(1): 20–35.

G. Meister. July 2013. Argonaute proteins: functional insights and emerging roles. *Nat Rev Genet*, 14(7):447–459.

G. Meister, M. Landthaler, A. Patkaniowska, Y. Dorsett, G. Teng, and T. Tuschl. 2004. Human Argonaute2 Mediates RNA Cleavage Targeted by miRNAs and siRNAs. *Molecular Cell*, 15(2):185–197.

G. Meister and T. Tuschl. Sept. 2004. Mechanisms of gene silencing by double-stranded RNA. *Nature*, 431(7006):343–349.

D. Mick, S. Dennerlein, H. Wiese, R. Reinhold, D. Pacheu-Grau, I. Lorenzi, F. Sasarman, W. Weraarpachai, E. Shoubridge, B. Warscheid, and P. Rehling. 2012. MITRAC Links Mitochondrial Protein Translocation to Respiratory-Chain Assembly and Translational Regulation. *Cell*, 151(7):1528–1541.

P. Mishra and D. C. Chan. Oct. 2014. Mitochondrial dynamics and inheritance during cell division, development and disease. *Nat Rev Mol Cell Biol*, 15(10): 634–646.

S. Morita, T. Horii, M. Kimura, Y. Goto, T. Ochiya, and I. Hatada. 2007. One Argonaute family member, Eif2c2 (Ago2), is essential for development and appears not to be involved in {DNA} methylation. *Genomics*, 89(6):687 – 696.

K. V. Morris, S. W.-L. Chan, S. E. Jacobsen, and D. J. Looney. 2004. Small Interfering RNA-Induced Transcriptional Gene Silencing in Human Cells. *Science*, 305 (5688):1289–1292.

N. Moshkovich and E. P. Lei. 03 2010. HP1 Recruitment in the Absence of Argonaute Proteins in *Drosophila*. *PLoS Genet*, 6(3):e1000880.

M. R. Motamedi, A. Verdel, S. U. Colmenares, S. A. Gerber, S. P. Gygi, and D. Moazed. 2004. Two RNAi Complexes, RITS and RDRC, Physically Interact and Localize to Noncoding Centromeric RNAs. *Cell*, 119(6):789–802.

K. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn, and H. Erlich. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol.*, 51:263–273.

E. P. Murchison, J. F. Partridge, O. H. Tam, S. Cheloufi, and G. J. Hannon. 2005. Characterization of Dicer-deficient murine embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 102(34):12135–12140.

K. Nakanishi, D. E. Weinberg, D. P. Bartel, and D. J. Patel. June 2012. Structure of yeast Argonaute with guide RNA. *Nature*, 486(7403):368–374.

U. Naumann, L. Daxinger, T. Kanno, C. Eun, Q. Long, Z. J. Lorkovic, M. Matzke, and A. J. M. Matzke. 2011. Genetic Evidence That DNA Methyltransferase DRM2 Has a Direct Catalytic Role in RNA-Directed DNA Methylation in Arabidopsis thaliana. *Genetics*, 187(3):977–979.

P. T. Nelson, M. De Planell-Saguer, S. Lamprinaki, M. Kiriakidou, P. Zhang, U. O'Doherty, and Z. Mourelatos. July 2007. A novel monoclonal antibody against human Argonaute proteins reveals unexpected characteristics of miRNAs in human blood cells. *RNA*, 13(10):1787–1792.

K. M. Nishida, K. Saito, T. Mori, Y. Kawamura, T. Nagami-Okada, S. Inagaki, H. Siomi, and M. C. Siomi. 2007. Gene silencing mechanisms mediated by AubergineâpiRNA complexes in Drosophila male gonad. *RNA*, 13(11):1911–1922.

T. Ohrt, J. Mütze, W. Staroske, L. Weinmann, J. Höck, K. Crell, G. Meister, and P. Schwille. 2008. Fluorescence correlation spectroscopy and fluorescence cross-correlation spectroscopy reveal the cytoplasmic origination of loaded nuclear RISC in vivo in human cells. *Nucleic Acids Research*, 36(20):6439–6449.

K. Okamura, A. Ishizuka, H. Siomi, and M. C. Siomi. 2004. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes & Development*, 18(14):1655–1666.

K. Okano, A. L. Vanarsdall, V. S. Mikhailov, and G. F. Rohrmann. Jan. 2006. Conserved molecular systems of the Baculoviridae. *Virology*, 344(1):77–87.

V. Olmedo-Monfil, N. Duran-Figueroa, M. Arteaga-Vazquez, E. Demesa-Arevalo, D. Autran, D. Grimanelli, R. K. Slotkin, R. A. Martienssen, and J.-P. Vielle-Calzada. Mar. 2010. Control of female gamete formation by a small RNA pathway in Arabidopsis. *Nature*, 464(7288):628–632.

T. I. Orban and E. Izaurralde. Dec. 2004. Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *RNA*, 11(4):459–469.

D. J. Pagliarini, S. E. Calvo, B. Chang, S. A. Sheth, S. B. Vafai, S.-E. Ong, G. A. Walford, C. Sugiana, A. Boneh, W. K. Chen, D. E. Hill, M. Vidal, J. G. Evans, D. R. Thorburn, S. A. Carr, and V. K. Mootha. 2008. A Mitochondrial Protein Compendium Elucidates Complex I Disease Biology. *Cell*, 134(1):112–123.

G. S. Pall and A. J. Hamilton. June 2008. Improved northern blot method for enhanced detection of small RNA. *Nat. Protocols*, 3(6):1077–1084.

R. Parker and H. Song. Feb. 2004. The enzymes and control of eukaryotic mRNA turnover. *Nat Struct Mol Biol*, 11(2):121–127.

J. C. Peng and G. H. Karpen. Jan. 2007. H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nat Cell Biol*, 9(1): 25–35.

S. Petri, A. Dueck, G. Lehmann, N. Putz, S. Rüdel, E. Kremmer, and G. Meister. 2011. Increased siRNA duplex stability correlates with reduced off-target and elevated on-target effects. *RNA*, 17(4):737–749.

J. Pfaff, J. Hennig, F. Herzog, R. Aebersold, M. Sattler, D. Niessing, and G. Meister. Oct. 2013. Structural features of Argonaute-GW182 protein interactions. *Proceedings of the National Academy of Sciences*, 110(40):3770–3779.

R. D. Phair, P. Scaffidi, C. Elbi, J. VecerovÂ¡á, A. Dey, K. Ozato, D. T. Brown, G. Hager, M. Bustin, and T. Misteli. 2004. Global Nature of Dynamic Protein-Chromatin Interactions In Vivo: Three-Dimensional Genome Scanning and Dynamic Interaction Networks of Chromatin Proteins. *Molecular and Cellular Biology*, 24(14):6393–6402.

R. F. Place, L.-C. Li, D. Pookot, E. J. Noonan, and R. Dahiya. 2008. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proceedings of the National Academy of Sciences*, 105(5):1608–1613.

J. Posakony, J. England, and G. Attardi. 1977. Mitochondrial growth and division during the cell cycle in HeLa cells. *The Journal of Cell Biology*, 74(2):468–491.

J.-L. Prieto and B. McStay. 2007. Recruitment of factors linking transcription and processing of pre-rRNA to NOR chromatin is UBF-dependent and occurs independent of transcription in human cells. *Genes & Development*, 21(16):2041–2054.

D. Prins and M. Michalak. Mar. 2011. Organellar Calcium Buffers. *Cold Spring Harbor Perspectives in Biology*, 3(3):1–16.

Y. Qi, X. He, X.-J. Wang, O. Kohany, J. Jurka, and G. J. Hannon. Oct. 2006. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature*, 443(7114):1008–1012.

D. Qiao, A. Zeeman, W. Deng, L. L.H.J., and H. Lin. 2002. Molecular characterization of hiwi, a human member of the piwi gene family whose overexpression is correlated to seminomas. *Oncogene*, 21:3988–3999.

J. Quick-Cleveland, J. Jacob, S. Weitz, G. Shoffner, R. Senturia, and F. Guo. 2014. The DGCR8 RNA-Binding Heme Domain Recognizes Primary MicroRNAs by Clamping the Hairpin. *Cell Reports*, 7(6):1994–2005.

T. A. Rand, S. Petersen, F. Du, and X. Wang. 2005. Argonaute2 Cleaves the Anti-Guide Strand of siRNA during RISC Activation. *Cell*, 123(4):621–629.

P. Rehling, N. Wiedemann, N. Pfanner, and K. N. Truscott. 2001. The Mitochondrial Import Machinery for Preproteins. *Critical Reviews in Biochemistry and Molecular Biology*, 36(3):291–336.

B. J. Reinhart and D. P. Bartel. Sept. 2002. Small RNAs Correspond to Centromere Heterochromatic Repeats. *Science*, 297(5588):1831–1831.

P. Reinhart, W. Taylor, and F. Bygrave. 1982. A procedure for the rapid preparation of mitochondria from rat liver. *Biochemical Journal*, 204:731–735.

M. Reuter, P. Berninger, S. Chuma, H. Shah, M. Hosokawa, C. Funaya, C. Antony, R. Sachidanandam, and R. S. Pillai. Dec. 2011. Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature*, 480 (7376):264–267.

M. Ricchetti, F. Tekaia, and B. Dujon. June 2004. Continued Colonization of the Human Genome by Mitochondrial DNA. *PLoS Biology*, 2(9):e273–.

F. V. Rivas, N. H. Tolia, J.-J. Song, J. P. Aragon, J. Liu, G. J. Hannon, and L. Joshua-Tor. Apr. 2005. Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat Struct Mol Biol*, 12(4):340–349.

G. B. Robb, K. M. Brown, J. Khurana, and T. M. Rana. Feb. 2005. Specific and potent RNAi in the nucleus of human cells. *Nat Struct Mol Biol*, 12(2):133–137.

P. Robin and R. Wong. 1988. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J Cell Physiology*, 136:507–513.

J. G. Ruby, C. Jan, C. Player, M. J. Axtell, W. Lee, C. Nusbaum, H. Ge, and D. P. Bartel. 2006. Large-Scale Sequencing Reveals 21U-RNAs and Additional MicroRNAs and Endogenous siRNAs in C. elegans. *Cell*, 127(6):1193 – 1207.

S. Rüdel. Biochemische Studien zur Phosphorylierung humaner Argonaute-Proteine. February 2011.

S. Rüdel, A. Flatley, L. Weinmann, E. Kremmer, and G. Meister. 2008. A multifunctional human Argonaute2-specific monoclonal antibody. *RNA*, 14(6):1244–1253.

S. Rüdel, Y. Wang, R. Lenobel, R. Körner, H.-H. Hsiao, H. Urlaub, D. Patel, and G. Meister. 2011. Phosphorylation of human Argonaute proteins affects small RNA binding. *Nucleic Acids Research*, 39(6):2330–2343.

M. Sadaie, T. Iida, T. Urano, and J.-i. Nakayama. 2004. A chromodomain protein, Chp1, is required for the establishment of heterochromatin in fission yeast. *The EMBO Journal*, 23(19):3825–3835.

U. Sahin, P. Lapaquette, A. Andrieux, G. Faure, and A. Dejean. July 2014. Sumoylation of Human Argonaute 2 at Lysine-402 Regulates Its Stability. *PLoS ONE*, 9(7):e102957–.

K. Saito, K. M. Nishida, T. Mori, Y. Kawamura, K. Miyoshi, T. Nagami, H. Siomi, and M. C. Siomi. 2006. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. *Genes & Development*, 20(16):2214–2222.

K. Saito, Y. Sakaguchi, T. Suzuki, T. Suzuki, H. Siomi, and M. C. Siomi. 2007. Pimet, the Drosophila homolog of HEN1, mediates 2′-O-methylation of Piwi-interacting RNAs at their 3′ ends. *Genes & Development*, 21(13):1603–1608.

K. Saito and M. C. Siomi. 2010. Small RNA-Mediated Quiescence of Transposable Elements in Animals. *Developmental Cell*, 19(5):687 – 697.

P. Salmon and D. Trono. Production and Titration of Lentiviral Vectors. In *Current Protocols in Neuroscience*, pages –. John Wiley & Sons, Inc., 2001.

T. Schalch, G. Job, S. Shanker, J. F. Partridge, and L. Joshua-Tor. Dec. 2011. The Chp1Ã¢ÂÂTas3 core is a multifunctional platform critical for gene silencing by RITS. *Nat Struct Mol Biol*, 18(12):1351–1357.

N. T. Schirle and I. J. MacRae. 2012. The Crystal Structure of Human Argonaute2. *Science*, 336(6084):1037–1040.

O. Schmidt, N. Pfanner, and C. Meisinger. Sept. 2010. Mitochondrial protein import: from proteomics to functional mechanisms. *Nat Rev Mol Cell Biol*, 11 (9):655–667.

J. C. Schwartz, S. T. Younger, N.-B. Nguyen, D. B. Hardy, B. P. Monia, D. R. Corey, and B. A. Janowski. Aug. 2008. Antisense transcripts are targets for activating small RNAs. *Nat Struct Mol Biol*, 15(8):842–848.

G. L. Sen and H. M. Blau. June 2005. Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. *Nat Cell Biol*, 7(6):633–636.

M. Seth, M. Shirayama, W. Gu, T. Ishidate, D. C. Jr., and C. Mello. 2013. The C.Â elegans CSR-1 Argonaute Pathway Counteracts Epigenetic Silencing to Promote Germline Gene Expression. *Developmental Cell*, 27(6):656 – 663.

M. Shirayama, M. Seth, H.-C. Lee, W. Gu, T. Ishidate, D. C. Jr., and C. Mello. 2012. piRNAs Initiate an Epigenetic Memory of Nonself {RNA} in the C.Â elegans Germline. *Cell*, 150(1):65 – 77.

G. Sienski, D. Dönertas, and J. Brennecke. 2012. Transcriptional Silencing of Transposons by Piwi and Maelstrom and Its Impact on Chromatin State and Gene Expression. *Cell*, 151(5):964 – 980.

D. Sims, I. Sudbery, N. E. Ilott, A. Heger, and C. P. Ponting. Feb. 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet*, 15 (2):121–132.

M. C. Siomi, K. Sato, D. Pezic, and A. A. Aravin. Apr. 2011. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol*, 12(4): 246–258.

L. M. Smith, O. Pontes, I. Searle, N. Yelina, F. K. Yousafzai, A. J. Herr, C. S. Pikaard, and D. C. Baulcombe. 2007. An SNF2 Protein Associated with Nuclear RNA Silencing and the Spread of a Silencing Signal between Cells in Arabidopsis. *The Plant Cell Online*, 19(5):1507–1521.

Z. D. Smith, M. M. Chan, T. S. Mikkelsen, H. Gu, A. Gnirke, A. Regev, and A. Meissner. Apr. 2012. A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature*, 484(7394):339–344.

H. S. Soifer, A. Zaragoza, M. Peyvan, M. A. Behlke, and J. J. Rossi. 2005. A potential role for RNA interference in controlling the activity of the human LINE-1 retrotransposon. *Nucleic Acids Research*, 33(3):846–856.

J.-J. Song, J. Liu, N. H. Tolia, J. Schneiderman, S. K. Smith, R. A. Martienssen, G. J. Hannon, and L. Joshua-Tor. Dec. 2003. The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat Struct Mol Biol*, 10(12):1026–1032.

L. Sripada, D. Tomar, P. Prajapati, R. Singh, and A. Singh. 2012. Systematic Analysis of Small RNAs Associated with Human Mitochondria by Deep Sequencing: Detailed Analysis of Mitochondrial Associated miRNA. *PLOS ONE*, 7(9).

L. Stalder, W. Heusermann, L. Sokol, D. Trojer, J. Wirz, J. Hean, A. Fritzsche, F. Aeschimann, V. Pfanzagl, P. Basselet, J. Weiler, M. Hintersteiner, D. V. Morrissey, and N. C. Meisner-Kober. Mar. 2013. The rough endoplasmatic reticulum is a central nucleation site of siRNA-mediated RNA silencing. *The EMBO Journal*, 32(8):1115–1127.

B. v. Steensel and S. Henikoff. Apr. 2000. Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase. *Nat Biotech*, 18(4): 424–428.

H. Stroud, T. Do, J. Du, X. Zhong, S. Feng, L. Johnson, D. J. Patel, and S. E. Jacobsen. Jan. 2014. Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat Struct Mol Biol*, 21(1):64–72.

K. Struhl and E. Segal. Mar. 2013. Determinants of nucleosome positioning. *Nat Struct Mol Biol*, 20(3):267–273.

V. M. Studitsky, W. Walter, M. Kireeva, M. Kashlev, and G. Felsenfeld. 2004. Chromatin remodeling by RNA polymerases. *Trends in Biochemical Sciences*, 29(3):127 – 135.

D. C. Swarts, K. Makarova, Y. Wang, K. Nakanishi, R. F. Ketting, E. V. Koonin, D. J. Patel, and J. van der Oost. Sept. 2014. The evolutionary journey of Argonaute proteins. *Nat Struct Mol Biol*, 21(9):743–753.

J. Sylvestre, A. Margeot, C. Jacq, G. Dujardin, and M. Corral-Debrinski. 2003. The Role of the 3′ Untranslated Region in mRNA Sorting to the Vicinity of Mitochondria Is Conserved from Yeast to Human Cells. *Molecular Biology of the Cell*, 14(9):3848–3856.

H. Tabara, M. Sarkissian, W. G. Kelly, J. Fleenor, A. Grishok, L. Timmons, A. Fire, and C. C. Mello. 1999. The rde-1 Gene, RNA Interference, and Transposon Silencing in C. elegans. *Cell*, 99(2):123 – 132.

H. Tabara, E. Yigit, H. Siomi, and C. C. Mello. 2002. The dsRNA Binding Protein RDE-4 Interacts with RDE-1, DCR-1, and a DExH-Box Helicase to Direct RNAi in C. elegans. *Cell*, 109(7):861 – 871.

O. H. Tam, A. A. Aravin, P. Stein, A. Girard, E. P. Murchison, S. Cheloufi, E. Hodges, M. Anger, R. Sachidanandam, R. M. Schultz, and G. J. Hannon. May 2008. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature*, 453(7194):534–538.

S. S. Teves, C. M. Weber, and S. Henikoff. 2014. Transcribing through the nucleosome. *Trends in Biochemical Sciences*, 39(12):577 – 586.

T. Thomson and H. Lin. 2009. The Biogenesis and Function of PIWI Proteins and piRNAs: Progress and Prospect. *Annual Review of Cell and Developmental Biology*, 25(1):355–376. PMID: 19575643.

D. Tianfang Ge and P. Zamore. 2013. Small RNA-Directed Silencing: The Fly Finds Its Inner Fission Yeast? *Current Biology*, 23(8):R318 – R320.

A. H. Ting, K. E. Schuebel, J. G. Herman, and S. B. Baylin. Aug. 2005. Short double-stranded RNA induces transcriptional gene silencing in human cancer cells in the absence of DNA methylation. *Nat Genet*, 37(8):906–910.

T. Treiber, N. Treiber, and G. Meister. 2012. Regulation of microRNA biogenesis and function. *Thrombosis and Haemostasis*, 107(4):605–610.

J. Vaughn, R. Goodwin, G. Tompkins, and P. McCawley. April 1977. The establishment of two cell lines from the insect Spodoptera frugiperda (Lepidoptera; Noctuidae). *In Vitro*, 13(4):213–217.

A. Verdel, S. Jia, S. Gerber, T. Sugiyama, S. Gygi, S. I. S. Grewal, and D. Moazed. Jan. 2004. RNAi-Mediated Targeting of Heterochromatin by the RITS Complex. *Science*, 303(5658):672–676.

A. Vermeulen, L. Behelen, A. Reynolds, A. Wolfson, W. Marshall, J. Karpillow, and A. Khvorova. Feb. 2005. The contributions of dsRNA structure to Dicer specificity and efficiency. *RNA*, 11(5):674–682.

B. Vissel and K. Choo. 1987. Human alpha satellite DNA - consensus sequence and conserved regions. *Nucleic Acids Research*, 15(16):6751–6752.

M. J. Vogel, L. Guelen, E. de Wit, D. P. Hupkes, M. Lodn, W. Talhout, M. Feenstra, B. Abbas, A.-K. Classen, and B. van Steensel. Dec. 2006. Human heterochromatin proteins form large domains containing KRAB-ZNF genes. *Genome Research*, 16(12):1493–1504.

M. J. Vogel, D. Peric-Hupkes, and B. van Steensel. June 2007. Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nat. Protocols*, 2 (6):1467–1478.

T. A. Volpe, C. Kidner, I. M. Hall, G. Teng, S. I. S. Grewal, and R. A. Martienssen. Sept. 2002. Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. *Science*, 297(5588):1833–1837.

A. Vourekas, Q. Zheng, P. Alexiou, M. Maragkakis, Y. Kirino, B. D. Gregory, and Z. Mourelatos. Aug. 2012. Mili and Miwi target RNA repertoire reveals piRNA biogenesis and function of Miwi in spermiogenesis. *Nat Struct Mol Biol*, 19(8): 773–781.

T. M. Vu, M. Nakamura, J. P. Calarco, D. Susaki, P. Q. Lim, T. Kinoshita, T. Higashiyama, R. A. Martienssen, and F. Berger. 2013. RNA-directed DNA methylation regulates parental genomic imprinting at several loci in Arabidopsis. *Development*, 140(14):2953–2960.

T. Watanabe, S.-i. Tomizawa, K. Mitsuya, Y. Totoki, Y. Yamamoto, S. Kuramochi-Miyagawa, N. Iida, Y. Hoki, P. J. Murphy, A. Toyoda, K. Gotoh, H. Hiura, T. Arima, A. Fujiyama, T. Sado, T. Shibata, T. Nakano, H. Lin, K. Ichiyanagi, P. D. Soloway, and H. Sasaki. 2011. Role for piRNAs and Noncoding RNA in de Novo DNA Methylation of the Imprinted Mouse Rasgrf1 Locus. *Science*, 332 (6031):848–852.

T. Watanabe, Y. Totoki, A. Toyoda, M. Kaneda, S. Kuramochi-Miyagawa, Y. Obata, H. Chiba, Y. Kohara, T. Kono, T. Nakano, M. A. Surani, Y. Sakaki, and H. Sasaki. May 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature*, 453(7194):539–543.

W. Wei, Z. Ba, M. Gao, Y. Wu, Y. Ma, S. Amiard, C. White, J. RendtlewÂ Danielsen, Y.-G. Yang, and Y. Qi. 2012. A Role for Small RNAs in DNA Double-Strand Break Repair. *Cell*, 149(1):101 – 112.

L. Weinmann, J. Höck, T. Ivacevic, T. Ohrt, J. Mötze, P. Schwille, E. Kremmer, V. Benes, H. Urlaub, and G. Meister. 2009. Importin 8 Is a Gene Silencing Factor that Targets Argonaute Proteins to Distinct mRNAs. *Cell*, 136(3):496–507.

S. H. Weitz, M. Gong, I. Barr, S. Weiss, and F. Guo. Feb. 2014. Processing of microRNA primary transcripts requires heme in mammalian cells. *Proceedings of the National Academy of Sciences*, 111(5):1861–1866.

A. T. Wierzbicki, J. R. Haag, and C. S. Pikaard. 2008. Noncoding Transcription by {RNA} Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes. *Cell*, 135(4):635 – 648.

K. J. Woolcock and M. Bühler. June 2013. Nuclear organisation and RNAi in fission yeast. *Current Opinion in Cell Biology*, 25(3):372–377.

L. Wu, L. Mao, and Y. Qi. 2012. Roles of DICER-LIKE and ARGONAUTE Proteins in TAS-Derived Small Interfering RNA-Triggered DNA Methylation. *Plant Physiology*, 160(2):990–999.

R. Xiao, R. Roman-Sanchez, and D. D. Moore. Mar. 2010. DamIP: A novel method to identify DNA binding sites in vivo. *Nuclear Receptor Signaling*, 8:e003–.

M. Xie, G. Ren, P. Costa-Nunes, O. Pontes, and B. Yu. 2012. A subgroup of SGS3-like proteins act redundantly in RNA-directed DNA methylation. *Nucleic Acids Research*, 40(10):4422–4431.

S. Yamanaka and H. Siomi. May 2014. diRNA-Ago2-RAD51 complexes at double-strand break sites. *Cell Res*, 24(5):511–512.

N. Yang and H. H. Kazazian. Sept. 2006. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat Struct Mol Biol*, 13(9):763–771.

E. Yigit, P. J. Batista, Y. Bei, K. M. Pang, C.-C. G. Chen, N. H. Tolia, L. Joshua-Tor, S. Mitani, M. J. Simard, and C. C. Mello. 2006. Analysis of the C. elegans Argonaute Family Reveals that Distinct Argonautes Act Sequentially during {RNAi}. *Cell*, 127(4):747 – 757.

S. T. Younger and D. R. Corey. 2011. Transcriptional gene silencing in mammalian cells by miRNA mimics that target gene promoters. *Nucleic Acids Research*, 39 (13):5682–5691.

Y.-R. Yuan, Y. Pei, J.-B. Ma, V. Kuryavyi, M. Zhadina, G. Meister, H.-Y. Chen, Z. Dauter, T. Tuschl, and D. J. Patel. 2005. Crystal Structure of A. aeolicus Argonaute, a Site-Specific DNA-Guided Endoribonuclease, Provides Insights into RISC-Mediated mRNA Cleavage. *Molecular Cell*, 19(3):405–419.

X. Zhang, I. R. Henderson, C. Lu, P. J. Green, and S. E. Jacobsen. 2007. Role of RNA polymerase IV in plant small RNA metabolism. *Proceedings of the National Academy of Sciences*, 104(11):4536–4541.

X. Zhang, X. Zuo, B. Yang, Z. Li, Y. Xue, Y. Zhou, J. Huang, X. Zhao, J. Zhou, Y. Yan, H. Zhang, P. Guo, H. Sun, L. Guo, Y. Zhang, and X.-D. Fu. July 2014. MicroRNA Directly Enhances Mitochondrial Translation during Muscle Differentiation. *Cell*, 158(3):607–619.

V. Zhou. *Methods for Global Characterization of Chromatin Regulators in Human Cells.* PhD thesis, Harvard University, 2012.

V. W. Zhou, A. Goren, and B. E. Bernstein. Jan. 2011. Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet*, 12(1):7–18.

D. Zilberman, X. Cao, and S. E. Jacobsen. 2003. ARGONAUTE4 Control of Locus-Specific siRNA Accumulation and DNA and Histone Methylation. *Science*, 299 (5607):716–719.

List of Tables

List of Figures

Appendix

Python Programms

Pimer Removal

5P-primerrem.py

```
#!/usr/bin/python
# -*- coding: iso-8859-1 -*-
#Name 5p-primerrem-final.py
# Run programm as python programmname.py or find path and put it here #!/dir/dir/python
#then start with ./programmname.py

#Input name.fastq

#Remove Primer everywhere, 3'seq of read is kept


####################
#get values from comandline when programm was called

import sys
print sys.argv # returns: ['param.py']
letters = 'f:l:o:'# the : means an argument needs to be passed after the letter
#f:          call 5p-primerrem.py -f input-file.fastq
#l:          call 5p-primerrem.py -l primercontainingfile.txt or .fasta
#o:          call 5p-primerrem-py -o outputfile.fastq
import getopt

opts, extraparams = getopt.getopt(sys.argv[1:],letters)
# starts at the second element of argv since the first one is the script name
# extraparms are extra arguments passed after all option/keywords are assigned
# opts is a list containing the pair "option"/"value"
print 'Opts:',opts
print 'Extra parameters:',extraparams



for o,p in opts:
  if o in ['-f']:
     inputfile = p   #filehandle inputfile
  elif o in ['-l']:  #filehandle primerlist
     primer = p
  elif o in ['-o']:  #filehandle outputfile
     outputfile = p
```

```
print 'Inputfile: ',inputfile
print 'Primerlist: ',primer
print 'Outputfile: ',outputfile


from Bio import SeqIO



def trimm_all(records, adaptors):
   #"""Trims perfect adaptorsequences.
   #This is a generator function, the adaptors argument should
   #be a list or iterator returning SeqRecord objects."""
 for record in records:
   c=1
   index=-1
   for adaptor in adaptors:
        len_adaptor = len(adaptor) #cache this for later
        index = record.seq.find(str(adaptor.seq))
        if index == -1 and c<39:
          #adaptor not found, try next adaptor
          c=c+1
          pass
        elif -1< index <100 and c <= 39:
          #trim off the adaptor and break loop
          cleanrecord=record[index+len_adaptor-4:]
          c=c+1
          break
        else:
          #adaptor not found, so won't trim and yield record
          cleanrecord = record
   yield cleanrecord



original_reads = SeqIO.parse(inputfile, "fastq")
li = list(SeqIO.parse(primer,"fasta"))
trimmed_reads = trimm_all(original_reads,li)
count = SeqIO.write(trimmed_reads, outputfile, "fastq")
print "Saved %i reads" % count
```

3pRC-primerrem.py

```
#!/usr/bin/python
# -*- coding: iso-8859-1 -*-
#Name 3pRC-primerrem-final.py
# Run programm as python programmname.py or find path and put it here #!/dir/dir/python
#then start with ./programmname.py

#Input name.fastq
```

```
#Remove Primer everywhere, 5' seq of read is kept


####################
#get values from comandline when programm was called
import sys
print sys.argv # returns: ['param.py']
letters = 'f:l:'# the : means an argument needs to be passed after the letter
#f:          call 5p-primerrem.py -f input-file.fastq
#l:          call 5p-primerrem.py -l primercontainingfile.txt or .fasta
#o:          call 5p-primerrem-py -o outputfile.fastq
import getopt

opts, extraparams = getopt.getopt(sys.argv[1:],letters)
# starts at the second element of argv since the first one is the script name
# extraparms are extra arguments passed after all option/keywords are assigned
# opts is a list containing the pair "option"/"value"
print 'Opts:',opts
print 'Extra parameters:',extraparams



for o,p in opts:
  if o in ['-f']:
     inputfile = p   #filehandle inputfile
  elif o in ['-l']:  #filehandle primerlist
     primer = p
  elif o in ['-o']:  #filehandle outputfile
     outputfile = p

print 'Inputfile: ',inputfile
print 'Primerlist: ',primer
print 'Outputlist: ',outputfile



from Bio import SeqIO



def trimm_all(records, adaptors):
   #"""Trims perfect adaptorsequences.
   #This is a generator function, the adaptors argument should
   #be a list or iterator returning SeqRecord objects."""
 for record in records:
   cleanrecord=-1
   c=1
   index=-1
   for adaptor in adaptors:
        len_adaptor = len(adaptor) #cache this for later
        index = record.seq.find(str(adaptor.seq))
        if index == -1 and c<38:
```

```
            #adaptor not found, try next adaptor
            c=c+1
            pass
          elif -1< index <100 and c <= 38:
            #trim off the adaptor and break loop
            cleanrecord = record[:index+4]
            c=c+1
            break
          else:
            #adaptor not found, so won't trim and yield record
            cleanrecord = record
      yield cleanrecord



original_reads = SeqIO.parse(inputfile, "fastq")
li = list(SeqIO.parse(primer,"fasta"))
trimmed_reads = trimm_all(original_reads,li)
count = SeqIO.write(trimmed_reads, outputfile, "fastq")
print "Saved %i reads" % count
```

GATC genome

GATC.py

```
#!/usr/bin/python
#programm to find all GATC positions per chromosome
#input: chromosome.fa
##############################
#from Bio import SeqIO
from optparse import OptionParser
import subprocess #, shlex

def main():

    parser = OptionParser()
    parser.add_option('-f','--file', dest='inputfile')

    (options, args)= parser.parse_args()

    if not options.inputfile:
        parser.error("Sequence file.fa is not given [-f]")


    inputfile=options.inputfile
    fp = open(inputfile,"r")
    li=fp.read().splitlines()
```

```
    fp.close()
    li=li[1:len(li)]
    record=''.join(li)
    record=record.upper()
    print len(record)+len(li)+1+len(inputfile.rstrip('fa'))
    outputfile=''.join(["/loctmp/pun06483/mESC-DAMID/index/GATCgenome/",inputfile.rstrip
        ('.fa'),"_GATCnoncase_sites.txt"])
    fp=open(outputfile,"a")
    fp.write(" ".join(["chromosome","start","end","\n"]))
    i=0
    while i <len(record)-3:
       index= record[i:len(record)].find("GATC")
       if index > -1:
          #GATC found get position
          fp.write(" ".join([inputfile.rstrip('.fa'),str(i+index+1),str(i+index+4),"\n"
              ]))
          i+=index+3
       else:
          break
    fp.close()

if __name__=="__main__":
    main()
```

Sam File Processing

---

samtoolsprocessing-rm.py

```
#!/usr/bin/python
###############
####Samtools processing to indexed bam file
from Bio import SeqIO
from optparse import OptionParser
import subprocess, shlex

def main():

    parser = OptionParser()
    parser.add_option('-f','--file', dest='inputfile')

    (options, args)= parser.parse_args()

    if not options.inputfile:
        parser.error("list of .sam is not given [-f]")

    inputfile=options.inputfile
    fp = open(inputfile,"r")
```

```
li=fp.read().splitlines()
fp.close()
print li

for i in range (0, len(li)):
    #sort in .sam all reads which map exactly at least for one read, yield only read
        a as it contains mapping info start and end.
    cmdline1 = "".join(["awk '{if (NR<28) {print $0;} else {a=$0; getline; b=$0; if
        (a ~ \"XS:i:\" && b ~ \"XS:i:\"); else print a\"\\n\"b}}' ",li[i]," > ",li[
        i].rstrip(".sam"),"hf.sam"])
    print cmdline1
    p=subprocess.Popen(cmdline1,shell=True)
    p.wait()
    cmdline2 = "".join(["samtools view -bS ",li[i].rstrip(".sam"),"hf.sam > ",li[i].
        rstrip(".sam"),"hf.bam"])
    print cmdline2
    p=subprocess.Popen(cmdline2,shell=True)
    p.wait()
    cmdline3 = "".join(["samtools sort ", li[i].rstrip(".sam"),"hf.bam ",li[i].
        rstrip(".sam"),"hf-sorted"])
    print cmdline3
    p=subprocess.Popen(cmdline3,shell=True)
    p.wait()



cmdlinea = "rm -rf *hf.bam"
p=subprocess.Popen(cmdlinea,shell=True)
p.wait()
cmdlineb = "rm -rf *hf.sam"
p=subprocess.Popen(cmdlineb,shell=True)
p.wait()
cmdline4="ls *hf-sorted.bam"
print cmdline4
p=subprocess.Popen(cmdline4,shell=True,stdout=subprocess.PIPE)
list=p.stdout.read()
list=list.split()
#list=["file1hf-sorted.bam","file2hf-sorted.bam","file3hf-sorted.bam"]
if len(list)!=len(li):
    print "Not all files are processed"
    return;
else:
    cmdline5 = "".join(["samtools merge -f ",list[0].rstrip("01hf-sorted.bam"),"-
        merged.bam *hf-sorted.bam"])
    print cmdline5
    print list
    p=subprocess.Popen(cmdline5,shell=True)
    p.wait()


cmdlinec = "rm -rf *hf-sorted.bam"
p=subprocess.Popen(cmdlinec,shell=True)
p.wait()
```

```
    cmdline6 = "".join(["samtools index ",list[0].rstrip("01hf-sorted.bam"),"-merged.bam
        "])
    p=subprocess.Popen(cmdline6,shell=True)
    p.wait()

if __name__=="__main__":
    main()
```

mESC-Sam-sortk5.py

```
#!/usr/bin/python
#
# Name: Sam_sortk5.py
# Run programm as python programmname.py or find path and put it here #!/dir/dir/python
#then start with ./programmname.py

#Input: list of file.sam

#Select exactly matching and Mitochondiral reads, matching only;
#and move satellites to separate file, move Mitoreads multiple matching to speparate
     file.


####################
from optparse import OptionParser
from Bio import SeqIO
import HTSeq
import itertools
import numpy
from operator import itemgetter
from itertools import islice
import time

def main():

    parser = OptionParser()
    parser.add_option('-f','--file', dest='inputfile')

    (options, args)= parser.parse_args()

    if not options.inputfile:
        parser.error("list of .sam is not given [-f]")

    inputfile=options.inputfile
    fp = open(inputfile,"r")
    li=fp.read().splitlines()
    fp.close()
    print li
    countex=0
    countMass=0
```

```
countMput=0
countSatellite=0
countSatellitereadpairs=0


for r in range (0, len(li)):
    with open(li[r]) as myfile:
  header="".join(list(islice(myfile,68)))
  #print header
    alignment_file=HTSeq.SAM_Reader(li[r])
    #contains all reads which can be exactly matched or second match worse
    exsamOutFileHandle = open("".join([li[r].rstrip(".sam"),"-ex.sam"]), 'w')
    exsamOutFileHandle.write(header)
    #conatins all reads which can be assigned to Mitochondrial based on GATC-peaks
    MassignedsamOutFileHandle = open("".join([li[r].rstrip(".sam"),"-Mass.sam"]), 'w
        ')
    MassignedsamOutFileHandle.write(header)
    #conatins all reads which can be potentially assigned to Mitochondria based on
        comparison with human Seq exp
    MputsamOutFileHandle = open("".join([li[r].rstrip(".sam"),"-Mput.sam"]), 'w')
    MputsamOutFileHandle.write(header)
    #conatins all left over reads (muliple mapping file k=5)
    SatelliteOutFileHandle = open("".join([li[r].rstrip(".sam"),"-Satellite.sam"]),
        'w')
    SatelliteOutFileHandle.write(header)
  #         alignment_file=HTSeq.SAM_Reader(inputfile)
 # alignment_file=HTSeq.SAM_Reader("
      exSample_Dam_Ago2_4_ATCACG_L006_R1_allrem_paired_01-k5-trial.sam")
#counter for different files:
for bundle in HTSeq.bundle_multiple_alignments(alignment_file):
  #print len(bundle)
  if len(bundle)==2:
    countex=countex+1
    for i in range (0, len(bundle)):
      exsamOutFileHandle.write("".join([bundle[i].get_sam_line(),'\n']))

  else:
    #transform bundle in dictionary, compare the different values select
    keys=list(range(0,len(bundle)))
    #values=list(bundle.items.optional_field('AS'))
    keys
    #make list of all needed items from one bundle and merge this with keys to
        identifiy the final reads
    values=[]
    for j in range (0,len(bundle)):
      values.append((keys[j],bundle[j].iv.chrom,bundle[j].iv.start,bundle[j].
          mate_start.start,bundle[j].optional_field('AS')))
    values
    #sort reads into files:
    #reads which have the highest alignment score compared to others
    even=[values[k] for k in range(0,len(bundle),2)]
    uneven=[values[l] for l in range(1,len(bundle),2)]
    AS1=sorted(even,key=itemgetter(4),reverse=True)
```

```
    #AS1
    AS2=sorted(uneven,key=itemgetter(4),reverse=True)
    #AS2
    a=next((i for i,v in enumerate(AS1) if v[1]=='chrM'),None)
    b=next((i for i,v in enumerate(AS2) if v[1]=='chrM'),None)
    a
    b
    if AS1[0][4]>AS1[1][4] and AS2[0][4]>=AS2[1][4] and (AS1[0][0]+1)==AS2[0][0]:
            exsamOutFileHandle.write("".join([bundle[AS1[0][0]].get_sam_line(),'\n']))
            exsamOutFileHandle.write("".join([bundle[AS2[0][0]].get_sam_line(),'\n']))
            countex=countex+1
    elif AS1[0][4]>=AS1[1][4] and AS2[0][4]>AS2[1][4] and (AS1[0][0]+1)==AS2[0][0]:
            exsamOutFileHandle.write("".join([bundle[AS1[0][0]].get_sam_line(),'\n']))
            exsamOutFileHandle.write("".join([bundle[AS2[0][0]].get_sam_line(),'\n']))
            countex=countex+1
    elif a is not None and b is not None and AS1[a][2] in range(10734,11176) and AS2
          [b][2] in range(10734,11176):
            MassignedsamOutFileHandle.write("".join([bundle[AS1[a][0]].get_sam_line(),
                  '\n']))
            MassignedsamOutFileHandle.write("".join([bundle[AS2[b][0]].get_sam_line(),
                  '\n']))
            countMass=countMass+1
    elif a is not None and b is not None and AS1[a][2] in range(5987,6700) and AS2[b
          ][2] in range(5987,6700):
            MassignedsamOutFileHandle.write("".join([bundle[AS1[a][0]].get_sam_line(),
                  '\n']))
            MassignedsamOutFileHandle.write("".join([bundle[AS2[b][0]].get_sam_line(),
                  '\n']))
            countMass=countMass+1
    elif a is not None and b is not None and AS1[a][2] in range(6701,10733) and AS2[
          b][2] in range(6701,10733):
            MputsamOutFileHandle.write("".join([bundle[AS1[a][0]].get_sam_line(),'\n'
                  ]))
            MputsamOutFileHandle.write("".join([bundle[AS2[b][0]].get_sam_line(),'\n'
                  ]))
            countMput=countMput+1
    else:
            countSatellite=countSatellite+1
            for n in range(0,len(bundle)):
          SatelliteOutFileHandle.write("".join([bundle[n].get_sam_line(),'\n']))
          countSatellitereadpairs=countSatellitereadpairs+1



exsamOutFileHandle.close()
MassignedsamOutFileHandle.close()
MputsamOutFileHandle.close()
SatelliteOutFileHandle.close()
#print "Saved %i reads" % recordCounter

totalpairs=sum([countex,countMass,countMput,countSatellite])
outdata="""total pairs & exactly aligned & chrM assigned cc & putative chrM &
     Satellites & total Satellite pairs \\\\
```

```
%s & %s & %s & %s &  %s & %s\\\\
    """ % (totalpairs,countex,countMass,countMput,countSatellite,countSatellitereadpairs
        )

    reportFileHandle = open("".join(["sorting-report-",time.strftime("%d-%m-%Y"),".txt"
        ]),'w')
    reportFileHandle.write(outdata)

if __name__=="__main__":
    main()
```

R scripts

Read counting script for R

---

counttable-singlereads-quick-uniq.R

```
#counttable for GATCgenome
#supply with bam-file_list.txt containing bamfiles of samples
#supply with "GATC-site_list.txt" containing GATCgenome
library(Rsamtools)
library(parallel)
#library(DESeq)#maskes the left and right function from gennomic ranges!!!
#library(chipseq)
library(GenomicFeatures)
#library(lattice)
#library(

#counttable after rmdup
library(Rsamtools)
library(parallel)
#library(DESeq)#maskes the left and right function from gennomic ranges!!!
#library(chipseq)
library(GenomicFeatures)
#library(lattice)
#library(
filepathu<-("/loctmp/pun06483/DamIDSeq-ana2013/bam/")
filelist<-("/loctmp/pun06483/DamIDSeq-ana2013/bam/bamulist.txt")
fls<-readLines(filelist)
names(fls)<-strsplit(fls,"u.bam")
which<- RangesList(chr19=IRanges(1,61431566))
pl<- ScanBamParam(which=which)
aln<-readGappedAlignments(paste(filepathu,fls[1],sep=""), format="BAM", use.names=FALSE,
    param=pl)
whichname<-seqlengths(aln)
output.path<-("/loctmp/pun06483/DamIDSeq-ana2013/Ranalysis/analysis20140723_uniq/")
```

```
#########chromosomes
chrom<-c("1","2","3","4","5","6","7","8","9","10","11","12","13","14","15","16","17","18
    ","19","20","21","22","X","Y","M")
chromos<-c(paste("chr",chrom,sep=""))
names(chromos)<-chromos

############functions

##########loading the data

granger1<-
function(filepath,pl)
{
aln<-readGappedAlignments(filepath, format="BAM", use.names=FALSE, param=pl)
#alnfin<-GRanges(seqnames=seqnames(left(aln)),ranges=IRanges(start=start(left(aln)),end=
    end(right(aln))),strand=rep("*",times=length(left(aln))),seqinfo=seqinfo(aln))
}

###########loading the features table from GATC
GATC.path<-("/loctmp/pun06483/humanGenome")
filelist<-("GATClist.txt")
GATCfls<-readLines(paste(GATC.path,filelist,sep="/"))
for(o in 1:length(GATCfls)){names(GATCfls)[o]<-strsplit(GATCfls,"_")[[o]][1]}


################loadging function for features

featureload<-
function(GATC.path,featname){
featdata<-read.table(paste(GATC.path,featname,sep="/"),header=TRUE)
features<-GRanges(
  seqnames=featdata$chromosome,
  ranges=IRanges(start=featdata$start,end=featdata$end),
  strand=rep("*",times=length(featdata$chr)),
  ID=paste(featdata$chr,featdata$start,featdata$end,sep="-")
)
}




##########analysis function
#workchromosome wise on 4 3 cores
i<-1

overlapmacs<-
function(outfile,modus){
for (i in 1:length(chrom))
```

```
  {
  #a<-paste(c("analyse chromosome:",i," for ",names(filepath)),sep="")
  #print(a)
  which<-GRanges(seqnames=names(chromos[i]),ranges=IRanges(1,whichname[names(chromos[i])
      ]))
  print(which)
  pl<- ScanBamParam(which=which)
#load features
  featname<-GATCfls[names(GATCfls)==names(chromos[i])]
  print(featname)
  features<-featureload(GATC.path,featname)
  featuress<-features[seqnames(features)==names(chromos[i])]
  featuress<-split(featuress,mcols(featuress)[["ID"]])
  print(head(featuress))
  #print(filepath)
  #print(names(filepath))
  #print(pl)
  N=length(featuress)
  datafin<-data.frame(integer(N),integer(N),integer(N),integer(N),integer(N))
  names(datafin)<-names(fls)

  for(j in 1:length(fls)){
    res<-granger1(paste(filepathu,fls[j],sep=""),pl)
    print(head(res))
    solap<-summarizeOverlaps(featuress,res,mode=modus)
    rowsAsFeatures<-data.frame(assays(solap)$counts)
    if(j==1){
    row.names(datafin)<-row.names(rowsAsFeatures)
    datafin[,1]<-rowsAsFeatures[,1]
    #head(datafin)
    }else{datafin[,j]<-rowsAsFeatures[,1]}}
  datafin<-datafin[rowSums(datafin)!=0,]
  print(head(datafin))
  write.table(datafin,file=paste(output.path,outfile,names(featname),"_20140723",".txt",
      sep=""),col.names=TRUE,append=FALSE,sep="\t")
  }
}

overlapmacs("union","Union")
```

DESeq analysis with R

DESeq-analysis.R

```
#DESeq of unique
#awk 'FNR !=1' *.txt > unionavail.txt
#load libraries
```

```
library(DESeq)
library(genefilter)

file.path<-("/data/thesis/tables/mESC/DAMID-uniq/Data-Bioinfo/analysis20140331/GATC-chr"
    )

counttable1<-read.table(paste(file.path,"unionchrM_20140331.txt",sep="/"),header=TRUE,
    sep="\t",row.names=1)
names(counttable1)
countcoln<-character(length(counttable1))
for(i in 1:length(counttable1)){countcoln[i]<-substr(names(counttable1)[i],8,nchar(names
    (counttable1)[i]))}
counttable<-read.table(paste(file.path,"unionavail.txt",sep="/"),header=FALSE,sep="\t",
    row.names=1)
names(counttable)<-countcoln
DamDesign=data.frame(
row.names=countcoln,
condition=c("untreated","treated","untreated","treated"),
libType=c("paired-end","paired-end","paired-end","paired-end")
)

countfin_hi1<-counttable[counttable[,1]>3,]
countfin_hi1<-countfin_hi1[countfin_hi1[,2]>3,]
countfin_hi1<-countfin_hi1[countfin_hi1[,3]>3,]
countfin_hi1<-countfin_hi1[countfin_hi1[,4]>3,]


DamDesign
head(countfin_hi1)
condition=DamDesign$condition
condition
cds=newCountDataSet(countfin_hi1,condition)
#total Colsums
#> colSums(counttable)
#      Dam_F Dam_Ago2_4     Dam_30b Dam_Ago2_D
#   11963707      202191    14156753      440928
> colSums(countfin_hi1)
     Dam_F Dam_Ago2_4     Dam_30b Dam_Ago2_D
     30495      103469       39949      203634
#chrM Colsums
> colSums(counttable[sorta,])
     Dam_F Dam_Ago2_4     Dam_30b Dam_Ago2_D
     24107       99452       31850      194559
> colSums(countfin_hi1[sorta,])
     Dam_F Dam_Ago2_4     Dam_30b Dam_Ago2_D
     24106       99425       31843      194521
#all apart from chrM Colsums
> colSums(counttable[sortb,])
     Dam_F Dam_Ago2_4     Dam_30b Dam_Ago2_D
  11939600      102739    14124903      246369
> colSums(countfin_hi1[sortb,])
     Dam_F Dam_Ago2_4     Dam_30b Dam_Ago2_D
      6389        4044        8106        9113
```

```
anal<-"hi3_all_sizelib_"

libsize<-c(26226038,22682308,26267559,21041694)
a=mean(libsize)
sizeFactors(cds)<-libsize/a
head(counts(cds,normalized=TRUE))
cds=estimateDispersions(cds,fitType="local")


#########check data
cdsBlind = estimateDispersions( cds, method = "blind",fitType="local" )
vsd = varianceStabilizingTransformation( cdsBlind )


####################################################
### code chunk number 71: heatmap
####################################################
library("RColorBrewer")
library("gplots")
select = order(rowMeans(counts(cdsBlind)), decreasing=TRUE)[1:30]
hmcol = colorRampPalette(brewer.pal(9, "GnBu"))(100)


####################################################
### code chunk number 72: figHeatmap2a
####################################################
par(mfrow=c(1,1))
png(file=paste("Enriched-peaks-Heatmap",anal,".png",sep=""))
heatmap.2(exprs(vsd)[select,], col = hmcol, trace="none", margin=c(10, 6))
dev.off()
####################################################
### code chunk number 74: sampleClust
####################################################
dists = dist( t( exprs(vsd) ) )


####################################################
### code chunk number 75: figHeatmapSamples
####################################################
mat = as.matrix( dists )
rownames(mat) = colnames(mat) = with(pData(cdsBlind), paste(condition, sep=" : "))
png(file=paste("Correlation",anal,".png",sep=""))
heatmap.2(mat, trace="none", col = rev(hmcol), margin=c(13, 13))
dev.off()

####################################################
### code chunk number 76: figPCA
####################################################
plotPCAm<-
function (x, intgroup = "condition", ntop = 50)
{
    rv = rowVars(exprs(x))
```

```
    select = order(rv, decreasing = TRUE)[seq_len(ntop)]
    pca = prcomp(t(exprs(x)[select, ]))
    fac = factor(apply(pData(x)[, intgroup, drop = FALSE], 1,
        paste, collapse = " : "))
    colours = c("green", "blue")
    xyplot(PC2 ~ PC1, groups = fac, data = as.data.frame(pca$x),
        pch = 16, cex = 2, aspect = "iso", col = colours, ylim=c(-50,50), main = draw.
            key(key = list(rect = list(col = colours),
          text = list(levels(fac)), rep = FALSE)))
}
png(file=paste("PCA-top50-peaks",anal,".png",sep=""))
print(plotPCAm(vsd, intgroup=("condition")))
dev.off()

cds=estimateDispersions(cds,fitType="local")
res=nbinomTest(cds,"untreated","treated")



png(file=paste("chrall-Disp_20141108",anal,".png",sep=""))
plotDispEsts(cds)
dev.off()
png(file=paste("MAplot-chrall_20140228",anal,".png",sep=""))
plotMA(res)
dev.off()
png(file=paste("hist-pval-chrall_20140228",anal,".png",sep=""))
hist(res$pval,breaks=100,col="skyblue",border="slateblue",main="")
dev.off()

##############

N<-length(res[,1])
pos<-data.frame(character(N),integer(N),integer(N),stringsAsFactors=FALSE)
names(pos)<-c("chrom","start","end")
for (i in 1:N){
  IDsplit<-unlist(strsplit(res$id[i],"-"))
  pos[i,1]<-as.character(IDsplit[1])
  pos[i,2]<-as.integer(IDsplit[2])
  pos[i,3]<-as.integer(IDsplit[2])}
datafin<-cbind(ID=res[,1],pos,res[,2:length(res)])
write.table(datafin,paste("GATC-peaks-all-",anal,"2014-11-08.csv",sep=""),col.names=TRUE
    ,sep="\t",appen=FALSE,row.names=FALSE)
write.table(datafin,paste("GATC-peaks-all-",anal,"2014-11-08.txt",sep=""),col.names=TRUE
    ,sep="\t",appen=FALSE,row.names=FALSE)
```

Eidesstattliche Erklärung

1. Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe des Literaturzitats gekennzeichnet.

2. Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebene Weise unentgeltlich geholfen:

   (a) Bioinformatische Auswertung der ChIP-Seq Experiments: Eugene Berezikov

   (b) Bioinformatische Auswertung des DAMID-Seq Experiments mit der Software Homer: Qian Yang and Julia Engelmann

   (c) Stabile Zelllinen von mESCs und deren DNA Extraktion: Constance Ciaudo

   (d) Percoll gradient zur Trennung von ER und Mitochondria: C. Richter-Dennerlein

3. Weitere Personen waren an der inhaltlich-materiellen Herstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe eines Promotionsberaters oder anderer Personen in Anspruch genommen. Niemand hat von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

4. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Regensburg, 07.Mai.2015

Unterschrift:

Acknowledgements

A special thank goes to Professor Dr. Gunter Meister for giving me the opportunity to conduct my thesis in his lab. During this time he supported me by exciting scientific discussions and good supervision while granting at the same time an open and inspiring scientific atmosphere. In addition, he showed me great understanding and support for my young family.
I also want to thank Eugene Berezikov, Qian Wang and Julia Engelmann for bioinformatic analysis and interesting discussions. I am deeply grateful for my time at the lab of Professor Peter Rehling, enabling me a quick start in mitochondrial biology. Here, I want to thank especially Dr. Sven Dennerlein, Dr. Ricarda Richter-Dennerlein and Christin Helbig.
For a good working environment and constructive discussions, I want to give thanks to Dr. Treiber (Nora and Thomas), Dr. Inga Lödige, Dr. Anne Dück, Dr. Sebastian Petri, Daniele Hasler, Dr. Melina Musri, Dr. Mathias Stotz, Judith Hauptmann, Ludwig Wankerl, Gerhard Lehmann, Norbert Eichner, Sigrun Ammon, Corinna Friedrich and Dr. Stefan Wenzl, who had always helpful advise.
I want to thank all those people who supported me while writing my thesis, especially all my friends in Regensburg, who accompanied me through all these years, Georg, Wolfgang, Sabine, Malte, Uli, Michi, Dominika and Sandip.

My parents and my brothers and sisters I thank for supporting me and helping me to find my way in life. Especially, I also want to thank my father- and mother-in-law, who helped me on numberless occasions to find the time and inner-peace for writing up the thesis.
Thanks Moogie for your belief in me and being right there when I needed you.

At last I want to thank my wonderful family, for supporting me, cheering me up when needed or bringing me back to earth, when I was flying high.

Thank you Joshua and Lilly for all the laughter.
Thank you Katrin, for being the woman you are.