Charles University in Prague, Faculty of Science

Department of Cell Biology

Study program: Biology

Cell and Developmental Biology

Jan Petržílek

Creating a biosensor for miRNA effector complex formation using CRISPR nucleases

Příprava biosenzoru tvorby miRNA efektorového komplexu pomocí CRISPR nukleáz

Diploma thesis

Supervisor: Prof. Petr Svoboda, Ph.D.

Institute of Molecular Genetics of the ASCR

Laboratory of Epigenetic Regulations

Prague 2018

Prohlášení

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 11. 08. 2018

Acknowledgments

First and foremost, I thank to my supervisor Petr Svoboda for the guidance during my MSc studies, his valuable inputs during writing of this thesis and his general support of my research ambitions. I also thank Radek Malik for consultations regarding the technical side of the experiments and Shubhangini Kataruka for the discussions and sharing her relevant data with me. I'd like to express my gratitude to all members of Petr Svoboda's group for the pleasant working environment that they create. I am thankful to the core microscopy facility for the provided service and to Dónal O'Carroll for providing us with the anti-AGO2 specific antibody.

Last but not least, I am grateful to my parents for their moral and material support during my studies and for raising me to be a curious individual.

Abstrakt

miRNA jsou malé regulující RNA, které fungují jako posttranskipční regulátory mRNA. miRNA navigují ribonukleoproteinové komplexy na mRNA a umlčují je inhibicí translace a degradací. V savcích miRNA regulují tisíce různých mRNA a byly rozeznány jako regulující faktory ve většině buněčných a vývojových procesech. Poruchy v regulaci miRNA dráhy mouhou vézt k závažným defektům a nemocem. V myších oocytech existuje unikátní situace, kdy všechny komponenty miRNA dráhy jsou přítomny, ale přesto je ona dráha zbytná a nefunkční. Molekulární podstata tohoto fenoménu a jeho významu zůstává stále nejasná.

I přes rozsáhlý účinek miRNA dráhy na genovou regulaci v somatických buňkách, strategie jak tuto dráhu studovat jsou limitované. Současné metody pro studium miRNA dráhy používají korelativní studie (jako například sekvenování nevé generace) nebo využívají reporterových systémů, které studují pouze relativně malé množství molekul v daném čase a jsou náchylné k artefaktům. V této práci prezentuji návrh a vývoj nové strategie pro přímé monitorování aktivity a integrity miRNA dráhy v živých buňkách za podmínek blízkých fyziologickým podmínkám. Tato strategie by mohla být použita *in vivo* pro studie myších oocytů. Tato strategie je založena na endogeně fluorescenčně značených proteinech ribonukleového komplexu AGO2 a TNRC6C, které utváří biosenzor. Ten ukazuje formování a dynamiku efektorového komplexu miRNA dráhy. V této práci jsem navrhl knock-in strategii kódujících sekvencí fluorescenčních proteinů založenou na CRISP/Cas9 technologii a vyřešil mnohé problémy spojené s vývojem biosenzoru. Dokázal jsem vytvořit *Ago2* knock-in buněčné linie a připravil a optimalizoval strategii pro knock-in *Tnrc6c*.

Klíčová slova

CRISPR, RNA umlčování, miRNA, miRISC, knock-in, biosenzor, genové manipulace

Abstract

miRNAs are small regulatory RNAs, which function as post-transcriptional mRNA regulators. They direct ribonucleoprotein complexes to cognate mRNA to repress them by translational inhibition and degradation. miRNAs regulate thousands of mRNAs in mammals and have been recognized as regulatory factors in most cellular and developmental processes. Dysregulation of the miRNA pathway can lead to severe defects and diseases. Interestingly, a unique situation exists in mouse oocytes, where all the miRNA pathway components are present, yet the pathway is dispensable and nonfunctional, the molecular foundation of this phenomenon and its significance still remain unclear.

In spite of the pronounced effects of the miRNA pathway in gene regulation in somatic cells, study strategies of the pathway bare limitations. Current methods for studying the activity of the miRNA pathway employ corelative studies (such as NGS) or reporter assays, which have relatively low throughput and are prone to artifacts. Here, I present design and development of a new strategy for directly monitor global miRNA pathway activity and integrity in near physiological conditions in living cells, which could also be employed *in vivo* for studies of mouse oocytes. The strategy is based on fluorescently tagged endogenous proteins of the ribonucleoprotein complex AGO2 and TNRC6C, which would form a biosensor sensing formation and dynamics of the miRNA pathway effector complex. I designed CRISPR/Cas9-based knock-ins of fluorescent protein coding sequences and troubleshooted numerous obstacles. I managed to produce *Ago2* knock-in cell lines and developed an optimized strategy for knock-in of *Tnrc6c*.

Keywords:

CRISPR, RNA silencing, miRNA, miRISC, knock-in, biosensor, genome editing

Table of contents

List of Figures

Glossary

3' UTR	3' untranslated region
5' UTR	5' untranslated region
ADAR	adenosine deaminase acting on RNA
Ago	Argonaute
Amp	Ampicillin
cDNA	complementary DNA
CPM	counts per million
CRISPR	clustered regularly interspaced short palindromic repeats
Dcr-1	Dicer-1
Dcr-2	Dicer-2
ddH_2O	double-distilled water
dNTPs	deoxynucleoside triphosphates
DSB	double-stranded DNA break
dsRBD	double-stranded RNA binding domain
dsRNA	double-stranded RNA
EDTA	ethylenediaminetetraacetic acid
endo-siRNA	endogenous short interfering RNA
FACS	fluorescence-activated cell sorting
FCCS	fluorescence cross-correlation spectroscopy
FRET	Förster resonance energy transfer
GV	germinal vesicle
GW	glycin-tryptophan
HR	homologous recombination
Loqs-PB	Loquacious-PB
Loqs-PD	Loquacious-PD
LTR	long terminal repeat
mESC	mouse embryonic stem cell
miRISC	microRNA RNA induced silencing complex
miRNA	microRNA
mRNA	messenger RNA
NEB	New England biolabs
NGS	next generation sequencing
nt	nucleotide(s)
OET	oocyte-to-embryo transition
P-bodies	processing bodies
p.a.	*per analysis* = analytical purity
PABPC	polyA-binding protein C
PACT	protein activator of PKR
PAGE	poly-acryl amid gel electrophoresis
PAM	protospacer adjacent motif
PAM2	polyA-binding protein-interacting motif2
PAZ	PIWI-ARGONAUTE-ZWILLE
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
piRNA	PIWI-interacting RNA
PIWI	P-element induced wimpy
PKR	protein kinase R
Pol II	RNA polymerase II
Pol III	RNA polymerase III

pre-miRNA	precursor microRNA
pri-miRNA	primary microRNA
PVDF	polyvinylidene difluoride
RCF	relative centrifugal force
RIPA	radioimmunoprecipitation assay
RISC	RNA induced silencing complex
RLC	RISC loading complex
RNAi	RNA interference
RPM	rotations per minute
RRM	RNA recognition motif
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate poly-acryl amid gel electrophoresis
sgRNA	single-guide RNA
shRNA	short hairpin RNA
siRNA	short interfering RNA
snRNA	small nuclear RNA
ssDNA	single-stranded DNA
ssRNA	single-stranded RNA
TDMD	target RNA-directed microRNA degradation
TRBP2	trans-activation-responsive RNA-binding protein 2
TSS	transcription start site
TTBS	tris-buffered saline with Tween20
TUT	terminal-uridylyl-transferase
WB	western blotting
WP	well-plate
ZGA	zygotic genome activation

Introduction

Small regulatory RNAs are ~20-30 nucleotides (nt) long RNA molecules, which guide ribonucleoprotein complexes to regulate gene expression by RNA silencing or chromatin remodeling (Pal-Bhadra et al., 2004; Volpe et al., 2002). They also provide defence against mobile elements (Aravin et al., 2007) and viruses (reviewed in (Obbard et al., 2009). RNA silencing is done either by a direct cleavage of the targets or by degradation facilitated by corecruited RNA degradation machineries.

In vertebrates there are three main types of small regulatory RNAs: microRNAs (miRNAs), small interfering RNAs (siRNAs) and piwi-interacting RNAs (piRNAs). They differ in their biogenesis and function. piRNAs serve a role in the piRNA pathway as defensive molecules regulating mobile elements (Aravin et al., 2007). They are produced from single-stranded RNA (ssRNA) precursors (Aravin et al., 2006). siRNAs facilitate the targeting in the RNA interference (RNAi) pathway, which is an RNA silencing pathway involving siRNAs produced from long double-stranded RNA (dsRNA). The third class are miRNAs, which are produced from the genome in a form of hairpin structures, which get process into short regulatory RNAs and are loaded onto the same effector proteins as siRNAs. The following text shows the discovery of RNAi and miRNA pathway and continues with detailed review of the pathways.

The first observations of RNA silencing in animals were of the RNAi and miRNA pathway because of the shared effector proteins and similar manifestation. Although the phenotypic manifestation of RNA-mediated silencing was observed first in the Petunia plant (Napoli et al., 1990) this observation of an RNA dependent sequence specific silencing in plants was believed to be a specialty of the plant biology. However, few years later, developmental biologists studying the worm *Caenorhabditis elegans* reported a short *lin-4* gene, which does not code for a protein but its expression of short *lin-4* RNA negatively regulates levels of a protein coding RNA *lin-14* by an RNA-RNA interaction with the 3' untranslated region of *lin-14* (Lee et al., 1993). Another manifestation of RNA silencing was unknowingly reported by research, in which *par-1* antisense ssRNA was injected in *C. elegans* in antisense inhibition experiments and observed a *par-1* loss-of-function phenotype (Guo and Kemphues, 1995). Surprisingly they observed the phenotype in both antisense and sense strand injections. That was explained only when Andrew Fire and Craig Mello

reported dsRNA as the most effective RNA to silence an endogenous mRNA and named this mechanism the RNA interference. They observed that only few dsRNA molecules per cell are sufficient for the silencing effect and that it is preserved in the next generation of worms. Therefore they suggested a possibility of amplification element in the silencing pathway (Fire et al., 1998). Fire together with Mello's group also showed that the RNAi is acting directly on matured mRNA in the cytoplasm and not on the DNA or nascent RNA transcripts (Montgomery et al., 1998). It is probable that the *par-1* ssRNAs injected by Guo and Kemphues were contaminated by the other strands during the preparation of ssRNA which lead to accidental formation of dsRNA and hence caused RNAi in the controls of antisense inhibition experiments.

Soon after Fire's and Mello's observation, several groups reported that dsRNAs are cleaved in a target-independent manner into ~21-23 nt long RNAs (Hamilton and Baulcombe, 1999; Hammond et al., 2000; Zamore et al., 2000). It was showed that RNA-induced silencing complex (RISC) is the complex that facilitates the repression of the targets. However, it was not known at the time what are the protein components of this complex. The research fueled by the observations resulted in a discovery that the processing of long dsRNAs into short RNAs is independent of RISC (Bernstein et al., 2001). Bernstein *et al.* looked at genes that contained an RNase III motif – a motif with affinity to dsRNA. They discovered the evolutionary conserved bidentate endoribonuclease Dicer as the processing enzyme of dsRNA in the *Drosophila melanogaster* Schnieder 2 (S2) cell lysate. It was shown that Dicer cleaves long dsRNA in a sequence-independent manner into small ~21-23 nt long dsRNAs. Dicer captures the dsRNA in its binding pockets and cleaves it by two RNase III domains (Bernstein et al., 2001).

Based on genetic screens the catalytic activity of RISC was linked to several proteins of the Rde-1/Qde2/Argonaute family (Catalanotto et al., 2000; Tabara et al., 1999) however, the biochemical proof was yet to be provided. For that RISC was later purified from *D. melanogaster* S2 cell lysate and a subunit with nuclease activity was identified as a member of the Argonaute family and named Argonaute2 (Ago2) (Liu et al., 2004; Meister et al., 2004; Rand et al., 2004). It was also shown that Ago2 can facilitates the silencing role of RISC alone without other proteins (Rand et al., 2004).

It was established that the RNAi is a pathway that serves as a modulator of the genome expression and as a defense mechanism against viral infections and against the activity of mobile elements (Felix et al., 2011; Ketting et al., 1999; Li et al., 2013). The pathway is activated by dsRNA, which is chopped by Dicer to siRNA duplexes. One strand is subsequently loaded onto Ago2 and guides

the enzyme to their targets based on their sequence. Ago2 then targets by siRNA-target base pairing and cleaves the cognate target molecule and represses it (Meister et al., 2004). These discoveries started off the field of small regulatory RNAs biology and more than two decades of exciting research.

The third category of small RNAs are the longer 23-30 nt long piRNAs. They were found in testes of mouse (Aravin et al., 2006). Later, piRNAs were showed to be present in both male and female reproductive tissues of *D. melanogaster (Brennecke et al., 2007)*. They were then reported in many animal model systems including fish, mouse and human. piRNAs interact with the PIWI-clade Argonaute proteins and hold a defensive role against mobile elements. Their biogenesis is Dicer-independent. There are two types of piRNAs in regards of their origin – primary and secondary piRNAs. Primary piRNAs are produced in clusters from specific loci and guard the general activity of mobile elements. piRNA clusters are loci filled with incomplete sequences of mobile elements which are transcribed, cleaved and processed in a complex biogenesis pathway into matured piRNA-PIWI ribonucleoprotein complexes. Secondary piRNAs are produced from targeted cleaved mobile elements RNA in a mechanism called the ping-pong loop. It is a process, which amplifies particular piRNA to rapidly repress active mobile element which the piRNA originated from. piRNA pathway harness two mechanisms of repressing targets, transcriptional and post-transcriptional gene silencing. Which regulation is employed depends on the particular piRNA molecule and a PIWI Argonaute it is loaded on (reviewed in (Czech and Hannon, 2016; Klattenhoff and Theurkauf, 2008)). The piRNA pathway is a fascinating defense mechanism, however beyond the scope of this thesis.

Both piRNA and RNAi are predominantly defensive pathways against parasitic molecules such as viral or mobile elements RNAs. In mammals, they are limited to several specific cell types and are involved mostly in germlines and embryogenesis. This is partly because mammals developed more successful response to dsRNA in cells by employing the interferon response and other dsDNA binding proteins such as protein kinase R (PKR) in the viral defense (reviewed in (Wang and Carmichael, 2004)). In contrast the miRNA pathway is virtually omnipresent. There are at least hundreds of different miRNAs in mouse (Chiang et al., 2010) or human (Fromm et al., 2015) and because miRNAs do not require perfect complementarity of base-pairing each miRNA can potentially target hundreds (or even thousands) of RNA molecules (Grosswendt et al., 2014). This creates strong post-transcription regulation mechanism capable of shaping the whole transcriptome (reviewed in (Shenoy and Blelloch, 2014)).

In vertebrates, siRNAs and miRNAs share the cytoplasmic machinery of proteins and after DICER slicing these RNAs are virtually unrecognizable from each other. miRNAs and siRNA differ in the biogenesis of precursors which are cleaved by Dicer. Majority of miRNAs are produced from endogenous hairpin-containing RNA molecules and require processing in the nucleus. The effect of miRNAs and siRNAs depends on the loading onto Argonaute proteins and the binding of the targets. In mammals there are four AGO proteins present, but only AGO2 retained its catalytic activity (Liu et al., 2004; Meister et al., 2004; Nakanishi et al., 2013; Schurmann et al., 2013). Strikingly, Nakanishi's group showed recently that human AGO3 possesses catalytic activity and still can slice its bound targets in some cases (Park et al., 2017). The consensus however still remains that RNAi is facilitated by AGO2 and the slicing only occurs after perfect complementarity between the siRNA and the target, because perfect complementarity allows structural arrangement, where the target is available for the enzymatically active center (Schirle et al., 2014). The slicing facilitated by AGO3 seems to be limited to one miRNA family (Park et al., 2017).

On the other hand, if any of the RNAs loaded on AGO proteins (Ago1-4) interacts with a target with mismatches in the targeting sequence Argonautes are unable to cleave the target but instead they inhibit translation of the target in what is called the miRNA pathway (reviewed in (Bartel, 2018; Ha and Kim, 2014)). In this case AGO proteins do not execute the silencing directly, instead they interact with a GW182 protein (in mammals TNRC6). which acts as a platform for mRNA repressive complexes such as a translation repression protein DDX6 (Kamenska et al., 2016) and a deadenylation complex CCR4-NOT, which leads to degradation (Braun et al., 2011; Chekulaeva et al., 2011; Fabian et al., 2011).

A unique situation exists in mammalian oocytes, where all the components of the miRNA pathway are present (Tam et al., 2008; Tang et al., 2007; Watanabe et al., 2008) yet the miRNA pathway seems to be inactive (Flemr et al., 2010; Ma et al., 2010; Suh et al., 2010). Moreover, in mouse oocytes, siRNAs produced from endogenous loci (endo-siRNA) are essential and take part in regulating protein coding-genes (Flemr et al., 2013; Tam et al., 2008; Tang et al., 2007; Watanabe et al., 2008). It seems that the activity of the miRNA pathway is disrupted in the beginning of the oocyte growth and only returns after the third embryonic division (Flemr et al., 2010). It has been shown that the production of miRNAs is dispensable for the oocytes (Suh et al., 2010). The cytoplasmic granules called processing bodies (P-bodies) disappear gradually during the oocyte growth (Flemr et al., 2010). P-bodies are linked with active miRNA pathway and contain proteins

taking part in the miRNA pathway or the subsequent mRNA degradation (Eulalio et al., 2007; Hubstenberger et al., 2017). Reported data suggest that the essential feature of the active miRNA pathway – the interaction between AGO and GW182 proteins is also disappearing during the growth of the oocyte (Flemr et al., 2010).

Even though there is genetic and experimental evidence that the miRNA pathway is not functional we still know little what exactly is the reason behind it. Several hypotheses should be considered. There could be factors that could specifically inhibit the miRNA pathway and not the endo-siRNA driven Ago2 slicing. This could be brought about by disrupting of the AGO-GW182 interaction. Also, modification or alternative isoforms of the proteins involved might suppress the pathway. A former student in Petr Svoboda's group Radek Jankele reported a non-functional truncated isoform of Ago2 and Ago3 in his master's thesis (Jankele, 2015) and the short Ago2 isoform was also reported recently by another group (Freimer et al., 2018). However, the activity of the endo-siRNA directed RNAi clearly indicates that fully functional AGO2 is still present. The efficiency of the miRNA pathway could also be affected by limiting amounts of miRNAs and proteins in the oocyte.

The kinetics and whether the inactivation of the miRNA pathway is a rapid or a gradual process could help in understanding the process. This could be elucidated by studying the interaction of the endogenously produced AGO and GW182 proteins *in vivo*.

siRNAs biogenesis and Dicer function

As mentioned earlier siRNAs are generated from longer dsRNA molecules by RNase III DICER. The source of dsRNA can be both extrachromosomal or endogenous. Long dsRNA molecules in animal cells may signal presence of viral or mobile element activity (Ketting et al., 1999; Li et al., 2013). In animals such as *Drosophila* and *C. elegans* RNAi plays a key role in defense against viruses (Obbard et al., 2009)). Many of these cell pathogens utilize dsRNAs in their replication cycles and that can be exploited against them. Mobile elements can destabilize genome by copying and randomly integrating themselves. Although mobile element insertions in genomic loci can lead to evolutionary novelties and adaptations (Franke et al., 2017) the activity needs to be under surveillance, so it does not destabilase genome integrity. A part of the regulation of the mobile elements is mediated by RNAi (Ketting et al., 1999; Stein et al., 2015). Lastly, dsRNA can be

produced from annealed sense and antisense transcripts for example from a protein-coding gene transcript and an RNA product from an antisense transcription of a pseudogene (Tam et al., 2008).

dsRNA is recognized by DICER in a sequence independent manner. Higher eukaryotisc DICERs typically have an N-terminal helicase domain, DUF domain, PAZ domain, two RNase III domains (a and b) and a dsRNA binding domain (dsRBD). The PAZ domain interacts directly with the two RNase III domains and acts as a molecular ruler to capture a terminus of a dsRNA which is subsequently sliced. This mechanism guarantees production of siRNAs of the same size with a 3'-end 2 nt overhang on each strand (reviewed in (Svobodova et al., 2016; Wilson and Doudna, 2013)). The produced short RNA molecules are ready to be loaded onto Ago proteins.

DICER proteins have two major classes of binding partners. First group consist of the AGO proteins which interact with DICER during the loading of small RNAs onto them. Second group are proteins with multiple dsRBDs which help with substrate recognition, cleavage fidelity and Argonaute loading (reviewed in (Meister, 2013; Svobodova et al., 2016; Wilson and Doudna, 2013)).

In mammals there are two paralog protein binding single DICER – trans-activation-responsive RNA-binding protein 2 (TARBP2, also TRBP2) and protein activator of PKR (PACT). These proteins alter the enzymatic activity which effects for example the length of produced small RNAs. They also have a role in the AGO loading (reviewed in (Meister, 2013; Nejepinska, 2012; Svobodova et al., 2016; Wilson and Doudna, 2013)).

Interestingly, while most of higher Eukaryotes utilize a single DICER for the production of both siRNAs and miRNAs, in *D. melanogaster* the production split between two DICER proteins (Lee et al., 2004b). The more conserved DICER (Dcr-1) is in charge of generating miRNAs and a newly evolved DICER (Dcr-2) is responsible for siRNAs (Murphy et al., 2008). In *Drosophila* each DICER interacts with different dsRBD protein and these further specialize the function of the DICER proteins. Dcr-2 interact specifically with R2D2 and one isoform of LOQUACIOUS (Loqs-PD) which are not redundant proteins and differ in function. Loqs-PD mostly stimulates the processing of dsRNAs whiles R2D2 enhances loading of siRNAs on AGO2. Dcr-1 interacts with a second isoform of LOQUACIOUS (Loqs-PB) (reviewed in (Meister, 2013; Nejepinska, 2012; Svobodova et al., 2016; Wilson and Doudna, 2013)).

miRNAs biogenesis

miRNAs can be divided into canonical and noncanonical miRNAs based on the mechanism of their biogenesis (reviewed in (Bartel, 2018)).

Canonical miRNAs biogenesis

Canonical miRNAs in animals are transcribed from the genome by RNA polymerase II (Pol II) in the form of long RNA molecules called primary miRNAs (pri-miRNAs) (Cai et al., 2004; Lee et al., 2004a). These transcripts originate from miRNA genes. All canonical miRNAs come from a region which forms a hairpin. This structure is a substrate for a nuclear protein complex named Microprocessor. It is a heterotrimeric complex consisting of RNase III DROSHA and two molecules of DGCR8 (called Pasha in invertebrates) (Nguyen et al., 2015). DROSHA has two RNase III domains. Similarly to DICER, each RNase III domain cleaves one strand of a stem of the pri-miRNA resulting in a ~60 nt long hairpin RNA with a 2 nt overhang on the 3'-end (Lee et al., 2003). This hairpin is called precursor miRNA (pre-miRNA). As any other Pol II transcript most pri-miRNAs have a 5'-end cap but some pri-miRNA are not polyadenylated at the 3'-end. This is a result of co-transcriptional processing on nascent transcripts by Microprocessor prior to polyadenylation (Ballarino et al., 2009). pre-miRNAs are transported by Exportin 5 and Ran-GTP from the nucleus through the nuclear pore to the cytoplasm (Bohnsack et al., 2004; Yi et al., 2003) where they are recognized and processed by DICER.

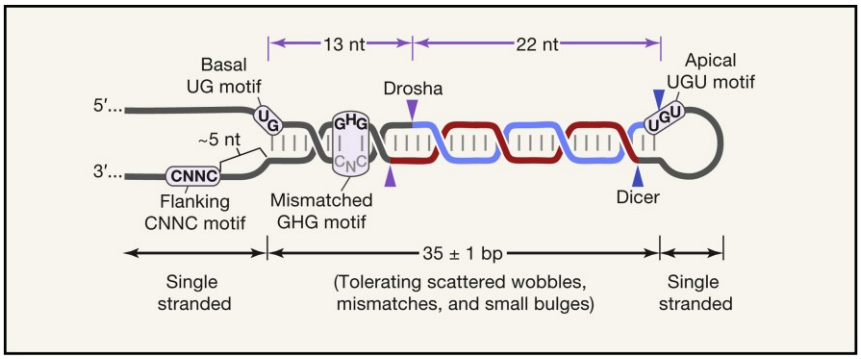


Figure 1 ***The structural and sequence features of a canonical pri-miRNA.***

Canonical pri-miRNA consist of a 35 nt ± 1 nt stem where base-pairing is preferred except of the GHG motif (where H is any nucleotide but N) on the further end from a single stranded loop. A few mismatches in the stem are tolerated and most naturally occurring miRNAs contain some. Additional sequence motifs (the basal UG motif, apical UGU motif and a flanking CNNC motif) contribute to the processing efficiency and enhance recognition of precise cleavage points by Drosha (purple arrows). Microprocessor recognizes the motifs and after binding functions as a molecular ruler to precisely cut the stem.

Drosha cleavage releases a pre-miRNA which can be processed by Dicer (blue arrows) into a short 22 nt dsRNA with 3'-end overhangs. Adopted from (Bartel, 2018).

Even though there are hundreds of substrates that Microprocessor successfully cleaves thousands of other hairpin RNA molecules are eliminated from entering the miRNA pathway. Microprocessor selects potential substrates based on their structure and sequence motifs. Mammalian Microprocessor prefers a stem of 35 nt ± 1 nt (Fang and Bartel, 2015) with an unstructured loop at the end of the stem (Zeng et al., 2005) and single strand regions on each of the RNA ends (Han et al., 2006; Zeng and Cullen, 2005). Perfect pairing though-out the stem loop is preferred although a few mismatches can be tolerated and a bulged conformation around the nucleotide 8 in the stem loop actually enhances the processing (Fang and Bartel, 2015). Sequence specific features are preferred at each end of the stem loop and one is located in the 3' single strand region. The Microprocessor recognizes the 35 nt stem loop and slices the pri-miRNA 13 and 11 nt from the base of the stem and 22 and 24 nt from the loop (Figure 1) resulting in a 22 nt stem with a loop, 5'-end phosphate group and a 3' 2 nt overhang (Fang and Bartel, 2015; Han et al., 2006; Ma et al., 2013).

The pri-miRNA preferences of Microprocessor are somewhat conserved across animals with the exception of nematodes (Auyeung et al., 2013; Fang and Bartel, 2015).

Noncanonical miRNAs

Not all miRNAs depend on the processing by Microprocessor or Dicer and can enter the miRNA pathway independently of DROSHA or DICER (Figure 2).

A group of small RNA molecules originates from introns of protein-coding genes which single strand tails were severed by the spliceosome. After a nuclear export they enter the miRNA pathway directly as pre-miRNA at the step of DICER. Because they are originally part of introns they were named mirtrons. Some mirtrons are left with a flanking 5' or 3' single strand tail and these are cut off by non-Microprocessor nuclease (Babiarz et al., 2008; Ruby et al., 2007a). Most mammalian mirtrons indeed possess a 5'-tail (Wen et al., 2015).

Another non-canonical class of miRNA are endogenous short-hairpin RNAs (shRNAs) (Babiarz et al., 2008). These can be transcribed by RNA polymerase III (Pol III); this is exemplified by two

hairpins originating from integrated adenoviral sequences. Pol III has a defined transcription start site (TSS) and termination, which in the case of the mentioned shRNAs define the 5' and 3' ends of the hairpins (Andersson et al., 2005; Bellutti et al., 2015). More complicated is the situation of shRNAs produced by Pol II as the transcription termination is not defined to a single nucleotide precision. The 5'-end of such shRNAs is determied by TSS. However, how is the 3' processed for DICER is yet to be described (Babiarz et al., 2008; Xie et al., 2013). Moreover, the cap on the 5' phosphate excludes this strand from loading on AGO proteins (Babiarz et al., 2008; Xie et al., 2013).

The third category of noncanonical miRNAs are chimeric shRNA. They are produced in tandem with other small RNA molecules and are recognized by DICER. These chimeric hairpins are for example part of a tRNA-like RNA (Pfeffer et al., 2005) or a small nuclear RNA (snRNA) (Cazalla et al., 2011).

A unique case of a noncanonical miRNA miR-451 that is essential for erythroblast maturation in mice has been intensively studied (Patrick et al., 2010). It is the most abundant miRNA in the erythrocytes of vertebrates. Unlike other noncanonical miRNAs, mir-451 is processed by Microprocessor but not DICER (Yang et al., 2010). The DROSHA produced hairpin is too short to be effectively cleaved by DICER and it directly enters AGO2, which slices the 3'-arm and the cleaved miRNA is turned into a mature miRNA by PARN exonuclease (reviewed in (Bartel, 2018)).

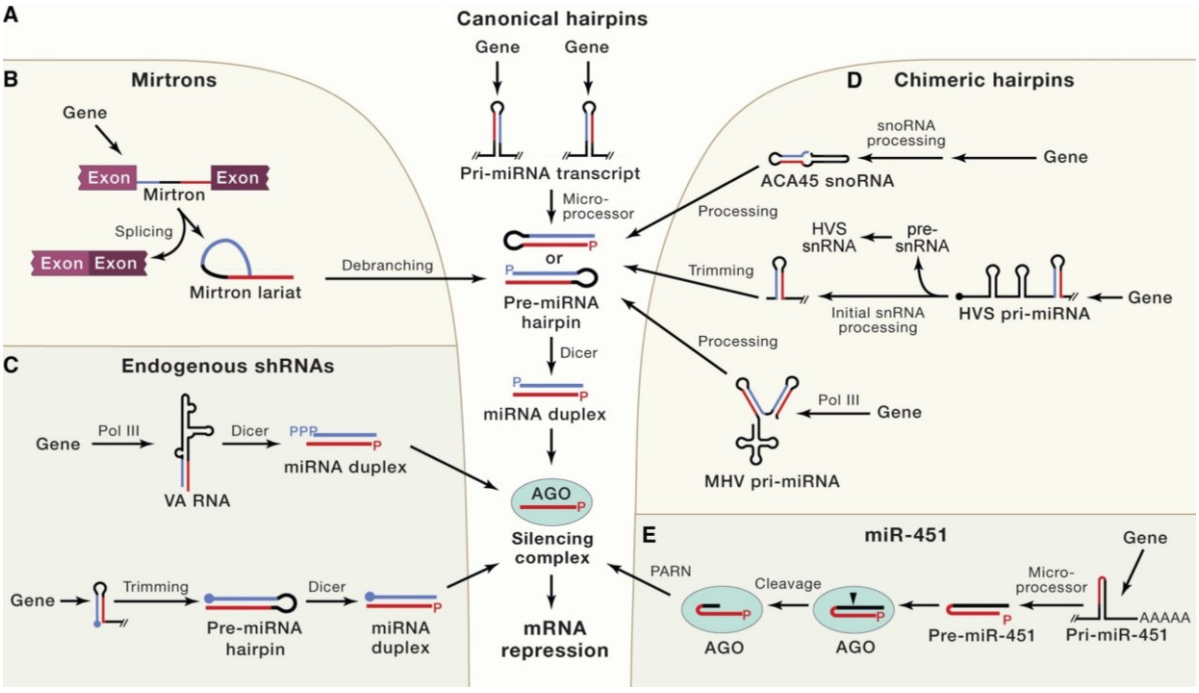


Figure 2 ***The biogenesis of canonical and noncanonical miRNA and their entry points to the canonical miRNA pathway.***

(A) *Canonical miRNAs are transcribed from independent miRNA genes by Pol II as pri-miRNA. They are processed by Microprocessor and Dicer.* ***(B-E)*** *Processing of different classes of noncanonical miRNAs. Mirtrons (B), endogenous shRNAs (C) and chimeric hairpins (D) enter the canonical pathway at the step of pre-miRNA and are processed by Dicer. miR-451 (E) is processed by Microprocessor but is Dicer-independent. Adopted from (Bartel, 2018).*

Loading of miRNAs onto Argonaute proteins

Argonaute proteins are specialized short-RNA-binding proteins. There are two subfamilies of AGOproteins – the AGO clade and the PIWI clade (reviewed in (Meister, 2013)). For the purpose of my thesis I only focus on the AGO clade proteins which are involved in RNAi and miRNA pathway and refer to them as AGO proteins in the text.

AGO proteins consist of four main domains (Figure 3) – an N-terminal domain, PIWI-ARGONAUTE-ZWILLE (PAZ) domain, middle (MID) and PIWI. The domains form two lobes connected by a hinge. One lobe is formed by the N-PAZ domains and MID-PIWI creates the second. During binding of a small RNA, the structure rearranges its conformation. The PAZ domain is common for DICER and AGO proteins and in it captures the 3'-end of the RNA in a specific binding pocket. Similarly, the MID domain anchors the first nucleotide of the 5'-end which interact with a conserved tyrosine residue (Jinek and Doudna, 2009). The PIWI domain of AGO2 possesses catalytic activity and can cleave targets (reviewed in (Meister, 2013)) leaving a 5' P and 3' OH termini (Elbashir et al., 2001). In AGO1 and AGO4 the catalytic center was lost during evolution. Special case is AGO3 which was thought catalytically inactive (Hauptmann et al., 2013) but lately slicing activity of human AGO3 was shown in specific cases (Park et al., 2017). The AGO3 activity is dependent on the guide RNA, which is loaded (Park et al., 2017). Interestingly, a recent report suggests that the C-terminal lobe (MID-PIWI) of a yeast *Kluyveromyces polysporus* is sufficient for loading, targeting and siRNA-dependent cleavage of its targets, however, this truncated protein is prone to off targeting and the N-terminal lobe is essential for mismatch sensing (Dayeh et al., 2018).

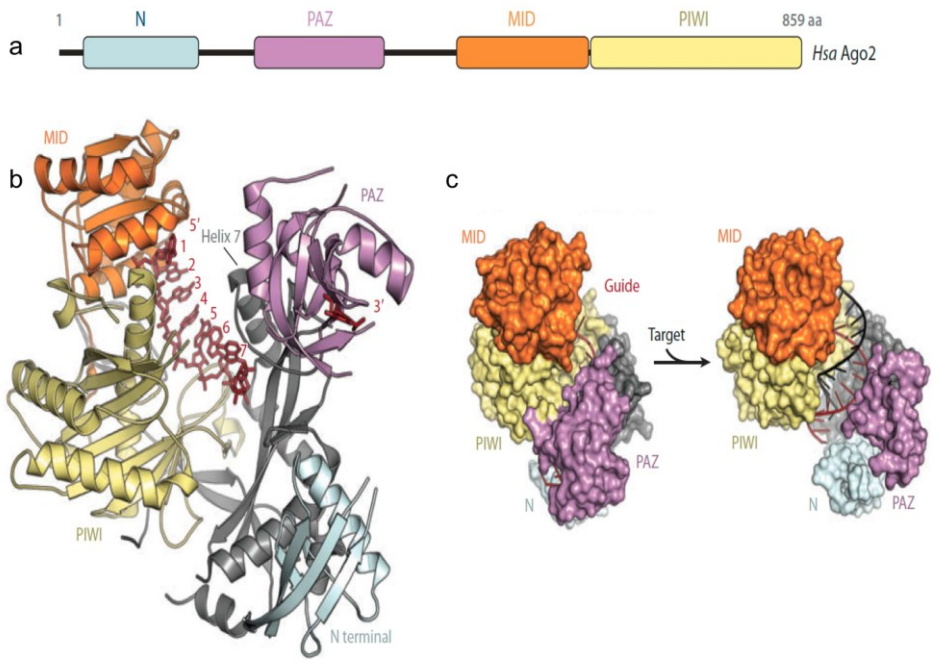


Figure 3 ***Visualization of the Ago2 structure***

(a) *Schematic sequence of the Ago2 domain composition in N- to C-terminal orientation. N domain and PAZ domain and PAZ and MID domains are connected by linkers visualized here as black lines.* ***(b)*** *Schematic ribbon visualization of the human Ago2 structure with a part of loaded guide RNA. The C-terminal lobe (left) consisting of MID (orange) and PIWI (yellow) domains and the N-terminal lobe (right) is formed by the N (teal) and PAZ (purple) domains. The MID domain binds the 5'-end of the RNA strand, the 3'-end (not shown) is bound by the PAZ domain. The seed region is exposed in the ridge however Helix 7 is creating a barrier to a potential target molecule and conformation change is required for successful targeting.* ***(c)*** *Illustration of Thermus thermophilus crystal structure of Ago2 and the conformation change that occurs during guide RNA-target base-pairing. Adopted and edited from (Wilson and Doudna, 2013).*

After cleavage by DICER, a miRNA or siRNA duplex molecule is ready to be loaded onto AGO proteins. However, only one strand is loaded in AGO (called guide strand) whiles the other strand, called passenger strand or miRNA* strand, is discarded. The selection which strand will be utilized, and which degraded is based mainly on the thermodynamic stability of the small RNA duplex. The 5'-end with less stable base pairing is preferred for loading (Khvorova et al., 2003; Schwarz et al., 2003). This is in no means an absolute and both strands of miRNAs and siRNAs can become the loaded guide strand.

AGO proteins are kept in an open conformation by chaperone HSP90 in humans (Johnston et al., 2010) and a complex HSP70-HSP90 in flies (Iwasaki et al., 2010) to enable the transfer of the RNA duplex. Recently it was shown that HSP70 is sufficient to open AGO2 and make it capable of accommodating the small RNA (Tsuboyama et al., 2018). HSP90 alone cannot open the AGO2. However, when paired with HSP70, it prolongs the time AGO2 is in the open conformation (Tsuboyama et al., 2018).

In mammals and other animals such as flies, DICER directly interacts with AGO proteins in a RISC loading complex (RLC) (Gregory et al., 2005; Maniataki and Mourelatos, 2005; Meister et al., 2005). In *D. melanogaster* R2D2 is key for the loading of siRNAs from Dcr-2 onto AGO2 and it has been shown that R2D2 serves as a sensor of thermodynamic asymmetry between the strands of siRNAs (Liu et al., 2003; Tomari et al., 2004). The same function has been confirmed for mammalian TRBP2 (Noland et al., 2011). Moreover, TAF11 which was previously annotated as a nuclear transcription factor was suggested as a new part of RLC and is proposed to have function in enhancing the loading activity of RLC (Liang et al., 2015). Recently, a paralog of R2D2 Loqs-PD was reported to form an alternative RLC with Dcr-2 by the binding by two dsRBDs of Loqs-PD and preferentially select more stable siRNA strand (Tants et al., 2017).

The strands of the duplex are separated from each other and the passenger strand is removed and subsequently degraded by the endonuclease C3PO (Liu et al., 2009; Ye et al., 2011). The N domain of AGO helps to separate the strands by wedging itself in between them (Kwak and Tomari, 2012). The duplex opens and unwinds leaving only the guide strand bound to the AGO protein. Whether additional factors are needed and what exactly is the nature of the unwinding is not completely clear (Kwak and Tomari, 2012).

miRNA targeting

After one small RNA strand has been successfully loaded onto an AGO protein, this complex is able to target complementary RNA molecules to repress their expression. If the complementarity is perfect and loaded AGO carries endonucleolytic activity the target RNA is severed. Such mechanism is common in plants or invertebrates such as *C. elegans* or *Drosophila*. In mammals, on the other hand, direct slicing by AGO is rare. Most guide RNAs are miRNAs and target RNAs with incomplete base-pairing, which is typical for miRNAs.

The most important region for target recognition of the miRNA is the sequence at miRNA nucleotide positions 2 to 7 (from the 5'-end). This short stretch of nucleotides is called the seed region. The base pairing of the seed to the seed matches is sufficient to trigger miRNA silencing. However, base-pairing downstream of the seed region is preferred as it decreases the dissociation constant and stabilizes the binding. The miRNA seed matching sequences can be found anywhere in the mRNA sequences, but the chance of repression is increased if the region is located in the 3' untranslated region (3' UTR) (reviewed in (Bartel, 2009)).

It is worth noting that because the seed region is only six or seven nucleotides long, one miRNA can recognize many different RNA molecules and therefore miRNAs have a massive effect on the transcriptome with possibly half of the protein-coding genes being regulated by miRNAs (Friedman et al., 2009; Lewis et al., 2005). miRNAs have a strong effect on the evolution of 3' UTRs of mRNAs and many sequences of mRNAs apparently underwent selection to avoid targeting by common miRNAs (Farh et al., 2005; Stark et al., 2005).

miRNA isoforms

Slight changes in the biogenesis of a miRNA can yield miRNA isoforms (isomirs) with widely different targets. Isomirs are miRNAs originating from the same hairpin, which target different set of targets. Incorporation of the miRNA* strand during loading onto AGO leads to a completely different set of targeted RNAs. Even subtle changes, such as one nucleotide shift in cleavage by either DROSHA or DICER, will have a profound consequence in targeting. Not only the heterogeneity of the 5'-end alters the seed region but also might influence the AGO loading and miRNA might be rejected by AGO because of a 5'-end nucleotide bias (Czech et al., 2009; Ghildiyal et al., 2010; Okamura et al., 2009; Ruby et al., 2007b). Isomirs are common however represent only ~10% of all miRNA reads in human, worms and flies. There are however examples of miRNAs that exist as two or more abundant isomirs with separate functions. The *D. melanogaster* miR-210 is processed into two almost equally represented 5' isomirs with different seed regions controlling different targets (reviewed in (Ameres and Zamore, 2013)).

Variability of miRNA 3'-end generally does not change targets but can influence repression kinetics or the half-life of the miRNA RISC (miRISC). Apart from the described variable cleavage during biogenesis there are two other known mechanisms of generating 3'-end heterogeneity. There is 3' trimming of miRNAs in *D. melanogaster,* which can be detected in ~40% of AGO1-bound miRNAs (Westholm et al., 2012). This trimming requires 3'-5' exonuclease NIBBLER (Han et al., 2011). Homologs of Nibbler were annotated in *C. elegans* and humans, however no role in the miRNA pathway was reported. The second mechanism is tailing of the 3'-end of miRNAs by the terminal nucleotidyl transferases. One of them is ZCCHC11 in mouse embryonic stem cells (ESCs), which produces uridylation of the pre-*let-7* (Hagan et al., 2009; Heo et al., 2009), which prevents it from DICER processing. Interestingly ZCCHC11 can also monouridylate class of pre-

miRNAs (including members of the *let-7* family), which in contrast enhances their processing by DICER (Heo et al., 2012).

In human and mouse, isomirs can also be generated by adenosine-to-inosine RNA editing by adenosine deaminase acting on RNA (ADAR) (Hundley and Bass, 2010). This is a tissue specific phenomenon primarily observed in brain. Inosine forms a Watson-Crick pair with cytosine and therefore the edited RNA can, although rarely, change target specificity. Intriguingly, the edited miRNA with altered seeds and hence targeting originate from an imprinted, locus (Chiang et al., 2010; Kawahara et al., 2008).

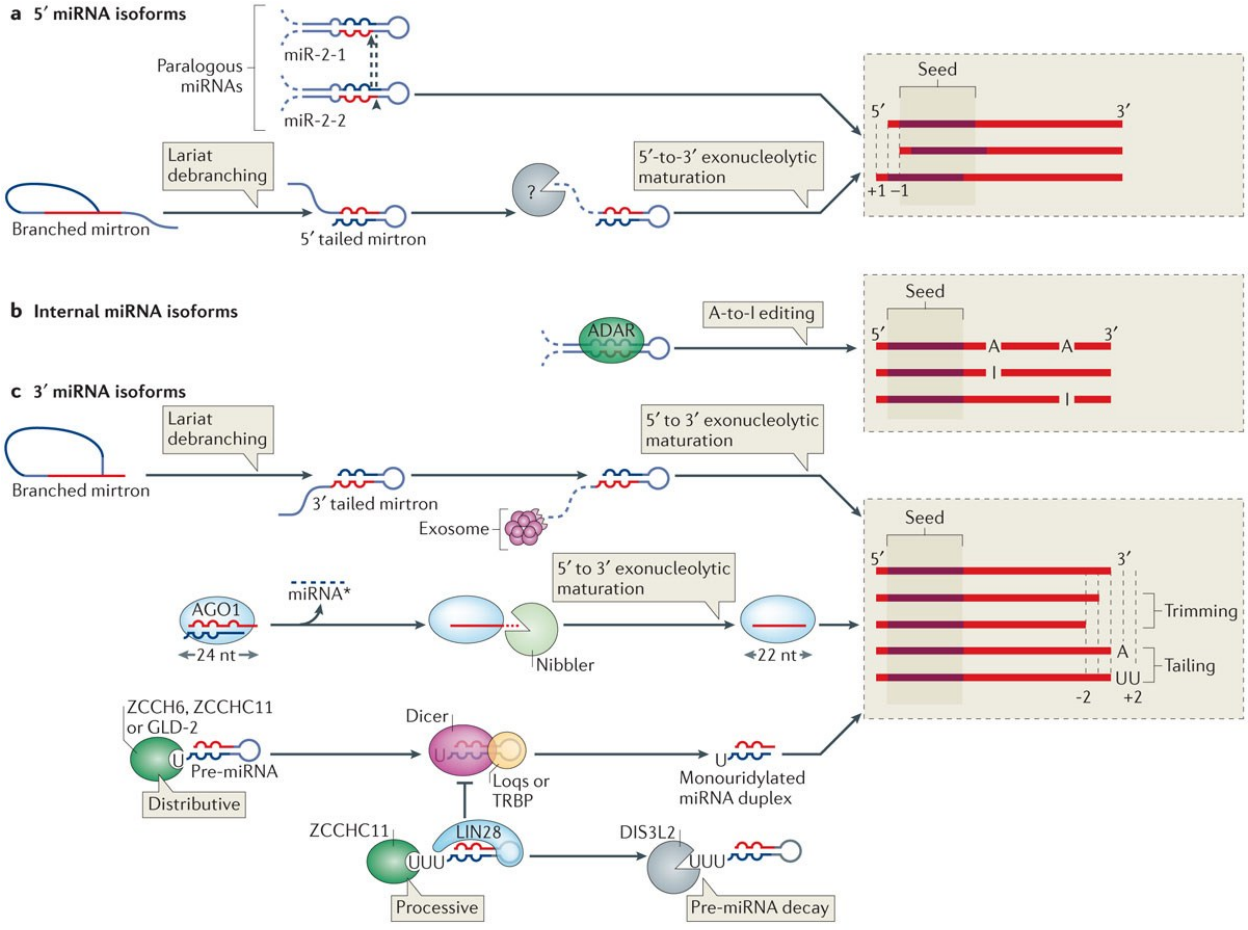


Figure 4 ***Mechanisms of miRNA isoform biogenesis.***
(a) *5'-end miRNA isoforms often result in altered repertoire of targets due to shift in the sequence of the seed region. The different 5'-ends can be brought on by Drosha or Dicer processing or as a result of 5'-3' trimming by an unknown exonuclease.* ***(b)*** *ADAR directed A-to-I conversion can produce miRNAs with different sets of targets. The effect is stronger if the seed region is edited but that happens relatively sporadically.* ***(c)*** *The 3'-end of miRNAs can be modified by trimming of tailed mirtrons, in D. melanogaster by the exonuclease Nibbler or by tailing by uridylation. Changes to the 3'-end rarely change the targeting of isomirs but can alter the half-life of the miRNA or prevents processing by Dicer. Adopted from (Ameres and Zamore, 2013).*

Mechanism of miRNA-mediated silencing

In the case of less imperfect base-pairing, none of Ago protein can slice the target. Instead repression in mediated through AGO-binding partner GW182 (called TNRC6 in mammals and AIN-1/2 in nematodes) to form miRISC (also called full RISC) (Figure 5). This complex comprises of an AGO-bound miRNA strand, AGO protein and a GW182 protein (reviewed in (Meister, 2013)). The miRISC inhibits translation of bound mRNAs and facilitates their degradation.

GW182 does not possess catalytic activity and serves as a docking platform for RNA degradation machineries. The N-terminal part of the protein consist of multiple glycine-tryptophan (GW) repeats. Studies elucidated the interaction of AGO and GW182 showing that the PIWI domain of AGO contains two tryptophan binding pockets and GW182 is bound here by the tryptophan residues (Chekulaeva et al., 2011; Schirle and MacRae, 2012). Argonaute can only bind one GW182 protein at the time however due to the excess of GW repeats human proteins from the GW182 family can interact with up to three Agos at once (Elkayam et al., 2017).

The C-terminal part of Gw182 proteins carries an RNA recognition motif (RRM) and a polyA-binding protein-interacting motif2 (PAM2). These motifs facilitate an interaction with polyA-binding protein C (PABPC) bound to the polyA-tail of a targeted mRNA (Figure 5). There are two proposed function of this interaction. First the binding to PABPC prevents its interaction with the cap-binding complex and blocks the circularization of mRNA, which is involved in translation initiation (Fabian et al., 2009; Zekri et al., 2009). Secondly the PABPC enhances the binding of the targeted mRNA and enhances the silencing in this manner (Moretti et al., 2012).

The C-terminal domain of GW182 also holds several GW repeats and via these repeats, the protein binds with NOT1. NOT1 is part of the CCR4-NOT deadenylase complex, which removes a polyA tail from an mRNA (Braun et al., 2011; Chekulaeva et al., 2011; Fabian et al., 2011) leading to decapping and degradation by 5'-3' exonuclease XRN1 (Huntzinger and Izaurralde, 2011). CCR4-NOT also interacts with the translation inhibition protein DDX6 (Mathys et al., 2014).

miRNA-mediated silencing is linked with the appearance of cytoplasmic foci called processing bodies (P-bodies) (Liu et al., 2005; Yang et al., 2004). P-bodies harbor miRISC proteins, proteins facilitating translation repression and mRNA degradation proteins along with mRNA molecules (Hubstenberger et al., 2017). They are, however, not essential for mRNA degradation (Eulalio et

al., 2007). The function of P-bodies is not completely known, but they play a role in regulating the amount of accessible mRNA regulatory proteins (Luo et al., 2018).

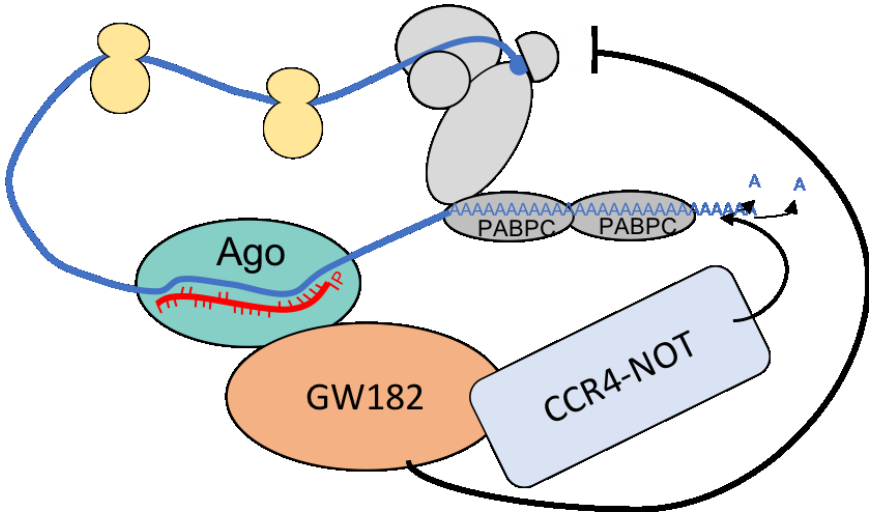


Figure 5 ***Schematic visualization of the miRISC and its binding partners acting on an mRNA.*** *AGO protein is guided by a ssRNA (red) onto a target RNA molecule (blue) by base-pairing. Extensive pairing of the nucleotides 2 to 7 (seed region) of the guide RNA with the target is required for efficient binding. Mismatches downstream of the seed region are tolerated. After successful binding AGO interacts with GW182 protein and forms miRISC which can inhibit translation of the target. Additionally, GW182 interacts with the CCR4-NOT deadenylation complex and this leads to 3'-5' trimming of the polyA tail and consequently decapping of the mRNA target and complete degradation by the exosome (not shown). Graphics inspired by (Bartel, 2018).*

Regulation of the miRNA pathway

Transcriptional and post-transcriptional regulation of miRNAs

The transcription of miRNA genes is dependent on by which RNA polymerase the genes is transcribed. miRNA originating from introns of other Pol II genes are under the regulation of those genes. Most of other miRNA genes are also products of Pol II as mentioned earlier and their regulation is similar to the regulation of protein coding genes. A few genes are Pol III transcribed and are regulated differently (Andersson et al., 2005; Bogerd et al., 2010).

Post-transcription regulation of pri- and pre-miRNAs is often specific for individual miRNAs or specific families of miRNAs. One of the best described post-transcriptional regulations is the earlier mentioned uridylation of *let-7* (Hagan et al., 2009; Heo et al., 2009). Let-7 promotes cellular differentiation and LIN28 (its target and negative regulator) is a factor that can reverse induction of differentiation of stem cells (Melton et al., 2010; Yu et al., 2007).

miRNA decay

Once miRNAs are loaded onto AGO proteins they are generally stable with half-lives of days (Guo et al., 2015). Exceptions are, for example, miRNAs in retina, which have variable stability (Krol et al., 2010), or other miRNAs with rapid turnover, which allows quicker response if needed (Rissland et al., 2011).

Another reported mechanism of miRNA degradation shows that highly complementary artificial targets of miRNAs (called antagomirs) designed to deplete certain miRISC-loaded miRNA by constitutive binding cause degradation of bound miRNAs (Ameres et al., 2010). Targets with high complementarity to 3' ends of miRNAs cause decay of bound miRNAs in a process referred to as target RNA-directed miRNA degradation (TDMD). The mechanism involves addition of adenosine or uracil to the 3'-end of the miRNA (tailing) and a consecutive 3'-5' exonucleolytic degradation (trimming) (Ameres et al., 2010). There are two proposed hypotheses why a strong 3'-end complementarity could trigger the tailing and trimming. The 3'-terminal nucleotide of miRNAs is normally bound to the PAZ domain of AGO however, upon 3'-end base-pairing the terminal nucleotide might become exposed to a nucleotidyltransferase (Wang et al., 2008; Yan et al., 2003). Alternatively, all miRNAs can be potentially a target of tailing and the probability is increased by slower dissociation from highly complementary matches (reviewed in(Ameres and Zamore, 2013)). A terminal-uridylyl-transferase (TUT) 1 has been identified as part of the pathway however, a direct role couldn't be assigned most likely due to redundancy in function (Haas et al., 2016). The exonuclease has been identified as DIS3L2 (Haas et al., 2016). Loss of DIS3L2 activity leads to reduction of TDMD in human and mouse cells (Haas et al., 2016). Uridylated pre-miRNAs can also be targeted by the DIS3L2 exonuclease, which is exemplified by the specific degradation of the uridylated pre-*let-7* miRNA in mouse embryonic stem cells (mESCs) (Ustianenko et al., 2013).

Regulation of small RNAi and the miRNA pathway in mouse oocytes

Mouse oocyte are highly specialized cells and have a unique regulation of RNAi and the miRNA pathway.

Mouse oocytes are one of the most specialized cells in the mouse, yet they need to maintain the ability to give rise to the whole animal after fertilization. Oocyte is also the biggest single-nuclear cell of the animal in volume. The cytoplasmic volume increases approximately 8 times during the oocyte growth and fully-grown oocytes (called germinal vesicles – GV) are transcriptionally silent (reviewed in (De La Fuente, 2006)) thus, growing oocytes have to store all the RNA and proteins needed for later development. Another consequence of transcriptionally quiescent cell is that its transcriptome is regulated only on the post-transcriptional level.

During the growth phase oocytes create unique environment that is preparing for fertilization and dedifferentiation into a totipotent zygote. After fertilization one of the earliest developmental traits is the oocyte-to-embryo transition (OET), which is marked by two processes – a zygotic genome activation (ZGA) and a degradation of maternal transcripts. Maternal transcripts are transcripts deposited in the oocyte and provide the program for the post-fertilization mechanisms and contribute to the reprogramming to the totipotent state. In mammals OET happens considerably faster than in other vertebrates which is given by the environment in which embryos are developing. In mice OET is unusually early even for mammals. The major phase of ZGA is carried out in a two-cell stage and degradation of some maternal transcripts start right after fertilization, which includes a plethora of siRNAs and miRNAs (reviewed in (Svoboda, 2017)).

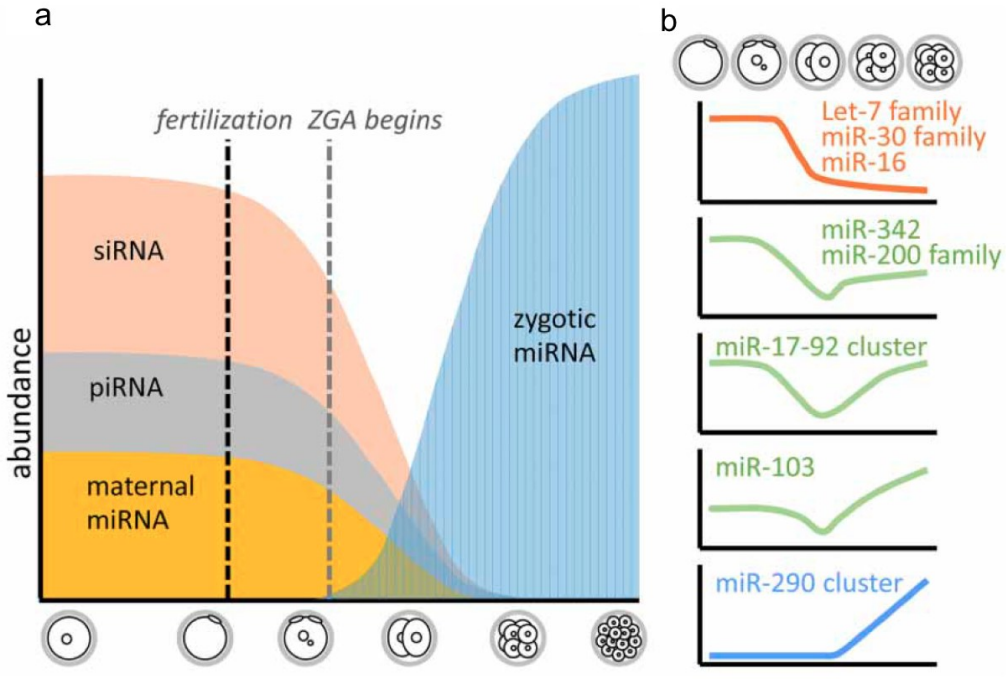


Figure 6 ***Degradation of maternal transcripts in mouse oocytes.***

(a) *Maternal transcripts are degraded soon after fertilization and zygotic transcripts are produced in a process called oocyte-to-embryo transition. Oocyte/embryo stages from left to right: germinal vesicle oocyte, ovulated MII oocyte, one-cell embryo, two-cell, morula, blastocyst.* ***(b)*** *Expression profiles of oocyte abundant miRNAs. Adopted from (Svoboda, 2017).*

RNAi in mouse oocytes

To this day, mouse oocytes are the best model for endo-siRNAs in vertebrates.
DICER is essential for mouse development and its loss results in embryonic lethality (Bernstein et al., 2003). Similarly, oocyte specific knock-outs of DICER cause female sterility and oogenesis defects (Murchison et al., 2007; Tang et al., 2007). The transcriptome of oocytes from the oocyte specific knock-out exhibit differential expression of ~20% genes and higher levels of mobile element transcripts (Suh et al., 2010; Tang et al., 2007). Similar phenotype also showed AGO2 knock-out or a catalytically dead version of AGO2 (Kaneda et al., 2009; Lykke-Andersen et al., 2008; Stein et al., 2015). These phenotypes were originally believed to be caused by the disruption of the miRNA pathway. Shockingly, it was reported that mouse oocytes produce high levels of endo-siRNAs and their targets correspond to the genes upregulated in the aforementioned knock-out models (Stein et al., 2015; Watanabe et al., 2008) pointing towards RNAi.

Moreover, Petr Svoboda's group found a truncated oocyte specific DICER (DICER^{O}) isoform, which processes dsRNA more effectively than full-length somatic DICER. *Dicer*^{O} has an alternative promoter which originated from an MT family long terminal repeat insertion (LTR). The MT LTR is silenced in cells other than the oocyte, which explains the tissue specificity of Dicer^{O}. Knock-out of the alternative MT LTR promoter results in female sterility and meiotic defects in oocytes, here phenocopying the Dicer phenotype (Flemr et al., 2013). A phenotype which is similar to the previous DICER knock-outs phenotype of an Ago2 knock-out or a catalytically dead version of AGO2, which is an evidence of the working RNAi pathway (Kaneda et al., 2009; Lykke-Andersen et al., 2008; Stein et al., 2015).

miRNA pathway in mouse oocytes

The composition of miRNA population in mouse oocytes differs from early embryo, which implies that miRNAs undergo similar clearance as maternal mRNAs in the early embryo. The most represented miRNA are members of the Let-7 and miR-30 families (Tam et al., 2008; Watanabe et al., 2008; Yang et al., 2008). Presence of Let-7 is puzzling because as mentioned early this conserved family is inhibitor of differentiation and reprogramming (Bussing et al., 2008). Early embryos

miRNAs correspond to what we observe in pluripotent embryonic stem cells (cells derived from blastomeres from the inner cell mass of blastocysts).

The phenotypes in oocytes lacking DICER or AGO2 were initially assumed as a result of non-functional miRNA pathway and not RNAi as mentioned earlier (Kaneda et al., 2009; Murchison et al., 2007; Tang et al., 2007). miRNAs were assumed to contribute to the degradation of the maternal transcripts, but studies have shown that oocytes lacking *DGCR8* gene are capable of developing beyond OET and ultimately giving rise to viable animals when fertilized by wildtype sperm (Suh et al., 2010). The developmental potential and virtually no changes in the transcriptome (Figure 7a) of the DGCR8 knockout oocytes suggests that the miRNA pathway is not essential for the development and fertilization of the oocyte.

Active miRNA pathway is linked with the formation of P-bodies in the cytoplasm. During the oocyte growth phase P-bodies are disappearing (Figure 7c-d) and reappear after OET. The loss of P-bodies corelates with the inability to detect the interaction of AGO and GW182 as the hallmark of active miRNA pathway (Flemr et al., 2010) in the oocyte, suggesting that the miRNA pathway is not active. Moreover, luciferase-based screenings of miRNA activity showed reduced repression potential of Let-7 and miR-30 members in oocytes (Figure 7 e-f) also pointing to reduced or non-existent activity of the pathway in oocytes (Ma et al., 2010).

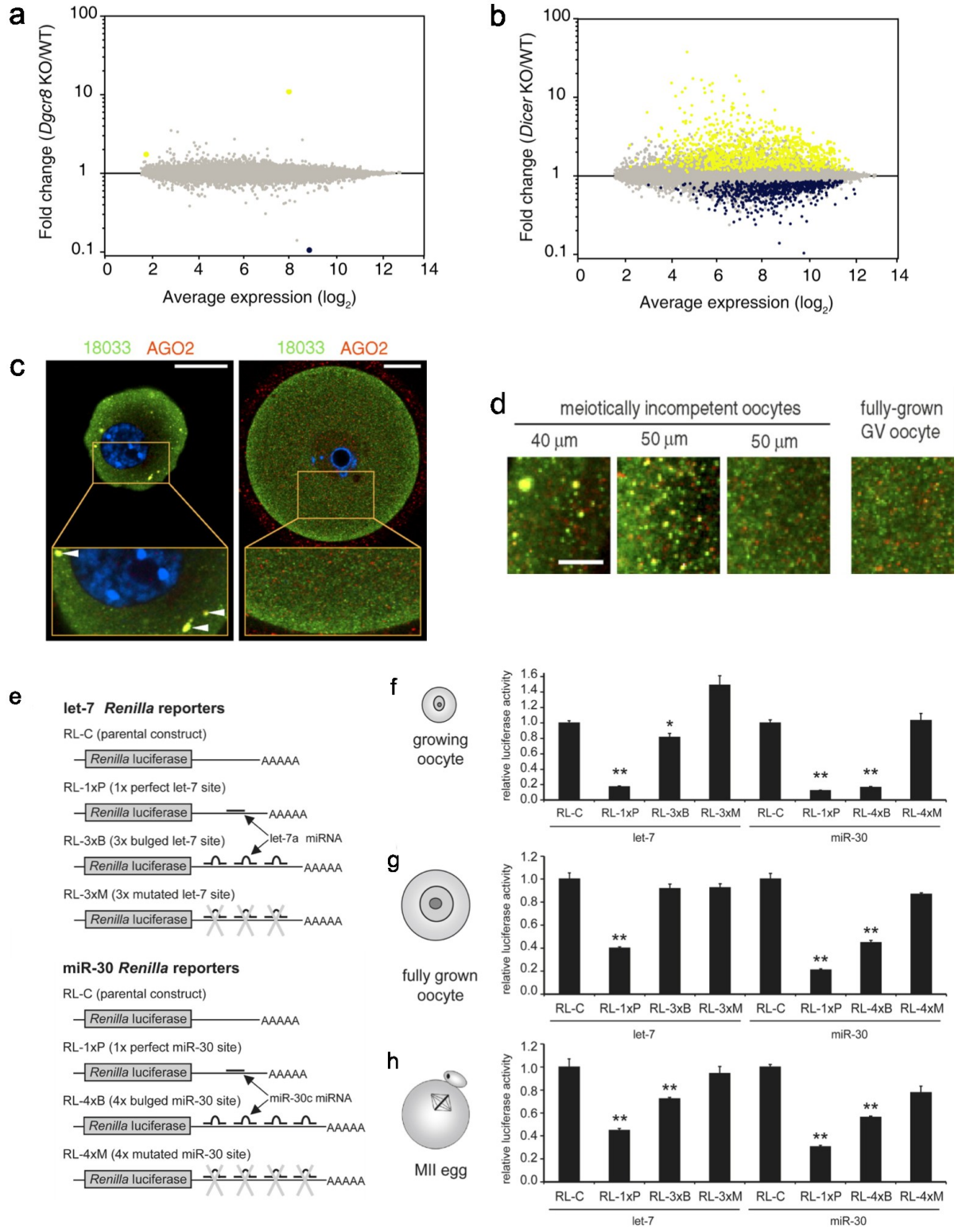


Figure 7 ***miRNA pathway inactivity in mouse fully-grown oocyte.***
(a) *Transcriptome of DGCR8 knock-out oocytes show virtually no changes in respect to wild type oocytes. Adopted from (Suh et al., 2010).* ***(b)*** *Differential expression in DICER knock-out oocytes points out key role of DICER. Adopted from (Suh et al., 2010).* ***(c-d)*** *The loss of P-bodies (yellow, arrows) during the growth of the oocyte indirectly indicates the loss of miRNA activity (bars – left is 20 µm, right is 5 µm). Stained with anti-AGO2 (red) 18033 (green). 18033 is an antibody that nonspecifically stains GW182. Adopted from (Flemr et al., 2010).* ***(e)*** *Reporters used for luciferase assays to test the repressive capabilities of Let-7 or miR-30 miRNAs. Adopted from (Ma et al., 2010).* ***(f-h)*** *Luciferase assays in growing oocytes (13*

*days old mice) (f), fully grown GV oocyte (g) and ovulated MII oocytes. Shown data are mean ± SEM, *p < 0.05, **p < 0.01. Adopted from (Ma et al., 2010).*

It was speculated that the miRNA pathway is not active because the oocyte needs a "buffering period" to effectively exchange maternal miRNAs, and produce new miRNAs, which will regulate the new embryonic developmental program (Svoboda, 2010). Activity of these zygotic miRNAs in the oocyte could cause dysregulations of the maternal program, which would have to be prevented. The purpose of highly abundant Let-7 and miR-30 in the oocyte is still not clear. The true mechanistical nature of the non-functionality of the miRNA pathways is also not understood. It could be brought about by qualitative or quantitative changes in the components of the pathway. As mentioned in the beginning there are studies suggesting possible ways why the miRNA pathway is not active however we think that there must be more to it and that the presented results do not explain the whole story.

The oocyte model system brings many constrains and makes many experimental biochemical approaches unusable. To study the miRNA pathway low sensitivity biological assays were employed and histochemistry was done with problematic antibodies on fixed samples. Because of these limitations some key aspects of the pathway are still not described. We are trying to address questions of miRNA regulation in the oocytes. In this thesis, I present a method which will allow us to study the kinetics of the formation of miRISC in the least intrusive way possible.

Aims and outlook

Aims

The main aim of my thesis was to fluorescently tag the key components of the miRISC – *Ago* and *Tnrc6* – in their endogenous loci producing a double knock-in cell line of mouse embryonic stem cells (mESCs). The endogenous tagging should neither alter the expression of the AGO and TNRC6 proteins nor it should influence function of the proteins. Interaction between AGO and TNRC6 proteins is an essential feature of the effector complex of the miRNA pathway; the fluorescent tagging would allow us to study the AGO-TNRC6 interaction in near physiological conditions as a proxy for the global miRNA pathway activity, hence acting as a biosensor. The AGO-TNRC6 interaction could be subsequently studied by live-cell microscopy methods including fluorescent cross-correlation spectroscopy (FCCS) and a Förster resonance energy transfer (FRET).

Strategy

For the tagging of AGO2 and TNRC6 proteins in their endogenous loci CRISPR/Cas9 technology would be used to generate targeted dsDNA breaks (DSB) to employ homologous recombination (HR) (Wang et al., 2015). The cells would be supplied with template DNA containing the desired knock-in sequence flanked by segments of DNA homologous to the gDNA of the endogenous locus for precise integration.

The CRISPR/Cas9 is a relatively recent technology, which was developed for a target-specific introduction of DSBs into genomes (Jinek et al., 2012; Ran et al., 2013). The technology is based on short single-guide RNA (sgRNA) molecules directing Cas9 nuclease originally found in *Streptococcus pyogenes (Sp*Cas9 here referred as Cas9) to DNA molecules through sequence complementarity. sgRNAs are composed fof a 18-21 nt guide sequence used for targeting and an invariant anchoring part, which facilitates the binding with Cas9 (Jinek et al., 2012). The nuclease scans a DNA molecule for the presence of a so-called protospacer-adjacent motif (PAM) – a trinucleotide sequence motif NGG. After the recognition of PAM, the guide sequence of the sgRNA is tested directly upstream of the PAM (Anders et al., 2014) and if the complementarity is extended the nuclease cleaves both DNA strands resulting in a DSB. The DSB is recognized as

DNA damage and activates the DNA damage response, which can lead to reparation by the homology directed repair (HDR) of the sequence if a homologous template is available (Wang et al., 2015). The cells can also integrate an extrachromosomal DNA, which has sequence homology at both sides of the damaged sequence utilizing homologous recombination. If the extrachromosomal DNA contains a non-homologous part, which is flanked by the homologous arms, it also becomes integrated.

Significance

As mentioned earlier. the developed fluorescent reporter system will function as a sensor of global miRNA pathway activity/integrity. It would offer a unique opportunity to monitor miRNA activity through monitoring of an interaction of two key protein components of the miRNA effector complex. The available methods up to date have been typically based on monitoring of miRNA targets abundancy through targeted fluorescent or luciferase reporters or high-throughput sequencing to determine the levels of the miRNA targets (Brustikova et al., 2018; Mullokandov et al., 2012). However, the target abundancy can be influenced by numerous other mechanisms, which may yield artefacts (Brustikova et al., 2018). The biosensor developed in this work would sense miRNA-specific interaction of AGO2 and GW182 proteins, hence it would offer a different assessment of miRNA pathway integrity and functionality and would be presumably insensitive to non-specific effects, which were troubling previous studies employing miRNA-targeted reporters (Brustikova et al., 2018).

In the laboratory, we have generated a set of compounds, which could modify miRNA pathway activity. The biosensor would be a perfect method how to further characterize their effect in living cells. Moreover, the mESC lines could be used to produce a mouse model, which could provide a miRNA pathway biosensor, which could be used to study miRISC biogenesis and dynamics as well as P-body decay during oocyte growth.

Materials and methods

Isolation of genomic DNA and RNA

Genomic DNA (gDNA) isolation

Mouse embryonic stem cells were trypsinized and collected, centrifuged at 10 000 RCF for 5 min, washed with PBS, centrifuged again and resuspended and lysed in a cell lysis buffer (100mM TRIS pH 8.5, 200 mM NaCl, 5 mM EDTA, 0.2% SDS in deionized water) containing Proteinase K (1 µl/ 100µl, Thermo Fisher Scientific). After 5 minutes at room temperature p.a. grade isopropanol was added to the cell lysate in 1:2 ration (isopropanol: cell lysate). DNA was extracted from the solution by spooling using a glass rod, resuspended in double-distilled water (ddH_2O) and heated at 95°C for 10 min for inactivation of proteinase K. Concentration was measured using NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific).

Total RNA isolation

R1 mESCs were centrifuged at maximum speed for 5 min at 4°C and total RNA was extracted using the RNeasy Mini Kit (Qiagen) following the manufacturer's manual. Concentration of isolated RNA was determined by NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific).

Reverse transcription

Complementary DNA (cDNA) was prepared from total RNA using RevertAid Reverse Transcriptase (Thermo Fisher Scientific) according to manufacturer's manual. Random hexamers (Thermo Fisher Scientific) were used as primers. Approximately 500 ng of total RNA was added per reaction. Incubation was done at 25°C for 10 min followed by 60 min at 42°C and terminated by 10 min step at 70°C.

Molecular cloning

DNA amplification

DNA for homologous recombination templates was amplified by nested polymerase chain reaction (PCR) using the proofreading Q5 DNA polymerase (New England BioLabs (NEB)). For the first reactions (20 µl (24 cycles), 2 µl were diluted 100x and 1 µl was used as template for the consecutive 30 µl reaction (30 cycles). For amplifying cDNA for expression vectors standard PCR was done in 30 µl reaction volume and 34 cycles using the Q5 polymerase. Standard reaction setup and program as follows:

Reagent	Final concentration/ amount	20 µl reaction	30 µl reaction
5X Q5 Reaction buffer	1X	4.0 µl	6.0 µl
(5X Q5 High GC enhancer)	(1X)	(4.0 µl)	(6.0 µl)
12.5 mM dNTPs	250 µM	0.4 µl	0.6 µl
100 µM Forward primer	500 nM	0.1 µl	0.15 µl
100 µM Reverse primer	500 nM	0.1 µl	0.15 µl
Q5 polymerase	0.02 U/µl	0.2 µl	0.3 µl
DNA template	30 - 1000 ng	1.0 µl	1.0 µl
ddH2O	to required volume	to 20.0 µl	to 30 µl

Step #	Step	Time	Temperature
1	Denaturation	3 min	95°C
2	Denaturation	30 s	95°C
3	Annealing	30 s	58-64°C
4	Elongation	45 s – 2 min	72°C
5	Elongation	1.5 – 4 min	72°C
6	Cooling	until removed	12°C

Annealing temperatures for PCR reactions were calculated for primer sequences complementary to the template using NEB TM calculator (https://tmcalculator.neb.com). Elongation times and annealing temperatures were adjusted for optimal conditions depending on the length of desired amplicons or primer properties, respectively. Q5 high GC enhancer was used for transcripts which failed to yield amplicons in the standard PCR setup.

Reactions were subsequently loaded on a 1.0% agarose gel (LE agarose – Lonza) in approximately 33 mM lithium bromide buffer (LB – prepared as a 20X solution – 8.93 g LiOH monohydrate,

36 g H_3BO_4 in 1.0 l of deionized H_2O) and then run in the same buffer at 80-120 V for ~40-60 minutes. For size comparison GeneRuler 1kb or 100bp DNA ladder (Thermo Fisher Scientific) were used. Desired PCR products were excised from the gel and purified using the Gel extraction kit (Qiagen) according to the manufacturer's manual. Concentration and purity of isolated DNA was determined by NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific).

Sequences encoding mRuby2 and Clover fluorescent proteins were cloned from plasmids (pCDNA3-mRuby2 and pCDNA3-Clover were a gift from Michael Lin (Addgene plasmids #40260 and #40259 respectively)) (Lam et al., 2012). Coding sequences for expression vectors were cloned from mESC cDNA; all other sequences from mESC gDNA.

Screening PCRs for plasmid inserts were conducted using Taq DNA polymerase (highQu), standard reaction setups followed:

Reagent	Final concentration/ amount	20 µl reaction
10x PCR buffer	1X	4.0 µl
12.5 mM dNTPs	250 µM	0.4 µl
30 µM Forward primer	600 nM	0.4 µl
30 µM Reverse primer	600 nM	0.4 µl
Tag DNA polymerase	0.05 U/µl	0.2 µl
DNA template	≤1000 ng	1.0 µl
ddH2O	to required volume	to 20.0 µl

PCR program was used as described earlier. Samples were consequently analyzed using agarose gel electrophoresis as described above.

The full list of used primers is provided in the Supplementary information section.

Ligation and transformation

Isolated PCR products (arms for homologous recombination and fluorescent protein sequences) were ligated into pJet1.2 plasmid (CloneJET PCR Cloning Kit – Thermo Fisher Scientific) using T4 DNA ligase (Thermo Fisher Scientific) according to the manufacturer's protocol in 1:3 (backbone: insert) molar ratio and total reaction volume of 20 µl. 4 µl of the ligation reaction were transformed to 50 µl of chemically competent Top10 or DH5α *E. coli* cells (prepared in-house according to Untergasser's protocol (A., 2008)). Transformation protocol: 20 min on ice, heat shock at 42°C for 1 min, 2 min on ice, added 600 µl of cold (4°C) LB media (10 g tryptone, 5 g

yeast extract, 10 g NaCl), shaking 45 min at 37°C and 600 RPM. Transformed bacteria were plated on ampicillin-containing (Amp) LB-agarose plates and grown overnight at 37°C.

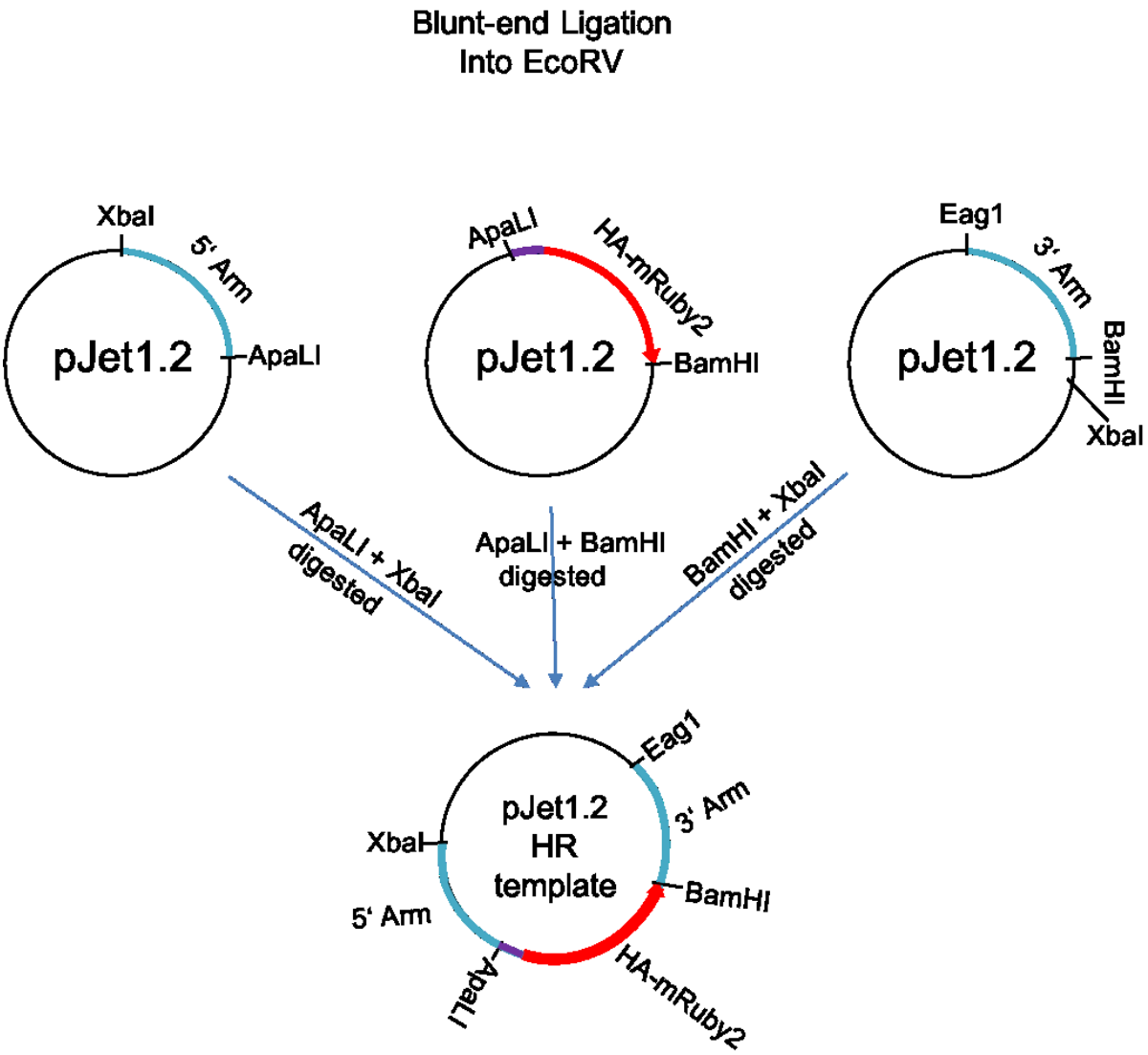
Colony screening and plasmid purification

The next day, colonies were selected, resuspended in 50 µl Amp LB media and screened as described in the section DNA amplification. Positive clones were inoculated into ~4.2 ml of Amp LB and grown overnight at 37°C and 250 RPM. Culture was centrifuged at 10 000 RCF for 6 min and plasmids were isolated using the QuickLyse MiniPrep kit (Qiagen) according to manufacturer's instructions. Plasmids were sequenced by GATC-biotech AG (later absorbed by Eurofins Genomics). For pJet1.2 inserts primer pJet_Fwd_seq and pJet_Rev2_seq were used, for mAgo2 expression vector mAgo2_screen_CDS_fwd and mAgo2_screen_CDS_rev were used. Samples were prepared according to GATC-biotech AG's protocol and send as 10 µl reactions containing 500-600 ng of plasmid, primer (final concentration 3 µM) and ddH_2O.

Positive plasmids were restricted using respective restriction enzymes (NEB or Thermo Fisher Scientific) in manufacturer's recommended buffers for two hours or overnight (Scheme 1) as follows:

Reagent	Amount (20 µl total)
10X Restriction buffer	2.0 µl
Restriction enzyme1	1.0 µl
Restriction enzyme2	1.0 µl
plasmid DNA	10 µg
ddH_2O	to 20.0 µl

Complete homologous templates were prepared by ligation using T4 DNA ligase (see above) of all parts together into restricted sticky-end pJet1.2 plasmids in molar ration 1:3 (backbone: inserts) for the HR templates. For the expression vectors pSV40 and pCDNA3 plasmids were used. Ligation reactions were used for transformation of chemically competent Top10 or DH5α *E. coli* cells as described. Colonies were then screened by PCR, positives were grown, plasmids were purified and sequenced. To amplify the verified plasmids bacterial cultures containing chosen plasmids were used for inoculating 100 ml of Amp LB and left shaking overnight at 37°C. Next day plasmids were purified using the MidiPrep kit (Qiagen) according to manufacturer's protocol. Plasmid maps for homologous recombination templates are provided in the Supplementary information section.



Scheme 1 Schematic visualization of the cloning of the homologous template plasmids on the example of HA-mRuby2-mAgo2 template. Each PCR amplicon was blunt-end ligated into the pJet1.2 vector, amplified in chemically competent E. coli and isolated plasmid was sequenced using the pJet_Fwd_seq primer. The inserts were cleaved by respective restriction enzymes as shown in the scheme in manufacturer's recommended buffers.

Preparation of sgRNA vectors

Single guide RNA (sgRNA) were designed using the CCTop prediction tool (https://crispr.cos.uni-heidelberg.de/ (Stemmer et al., 2015)).

sgRNA sequences were ordered as ssDNA oligos from Sigma-Aldrich. For each sgRNA expression vector two ssDNA oligos are necessary – one oligo contains the target sequence of the sgRNA and the second oligo is complementary to the first one. The complementary sequence of the oligos is flanked with nucleotides, which after annealing, create sticky ends for ligation of the duplex oligo into a U6-sgRNA plasmid. Complementary sense and antisense oligos were annealed in an

annealing reaction (50 µl reaction, 10x Annealing buffer 5.0 µl, 100 µM Sense oligo 3.0 µl, 100 µM Antisense oligo3.0 µl, ddH2O 49.0 µl). Reactions were heated at 95°C for 5 min and then left gradually cool down to RT.

After annealing 5 µl was directly used for T4 kinase (Thermo Fisher Scientific) reactions prepared according to manufacturer's manual. Phosphorylated samples were then diluted to 100 µl and 3 µl were used for ligation with a linearized dephosphorylated U6-sgRNA plasmid. The U6-sgRNA plasmid already contains the invariant part of the sgRNA and after successful insertion of the targeting sequence produces complete sgRNA from the Pol III U6 promoter.

preparation of CRISPR/Cas9 reporter plasmids

ssDNA oligos for CRISPR/Cas9 reporters containing sgRNA target sequences and their respective PAM in tandem were ordered from Sigma Aldrich.
Oligos were annealed, phosphorylated and diluted as described in the section Preparation of sgRNA vectors. Subsequently the duplex oligos were ligated into a linearized dephosphorylated pMB1610_pRR-Puro plasmid (pMB1610_pRR-Puro was a gift from Marc Bühler (Addgene plasmid # 65853) (Flemr and Buhler, 2015)). Full list of sgRNA oligo and reporter sequences is provided in the Supplementary information section.

Reporter plasmid produces puromycin resistance after cleavage within the cassette of the plasmid. Plasmids were obtained and verified as described earlier. As screening primers were used the antisense sgRNA oligos in combination with plasmid specific primers amplifying the second strand in the PCR reactions.

Cell culture

All cell culture work was done with R1 mESCs (unless stated otherwise) cultured in high glucose DMEM+ 15% fetal bovine serum + sodium pyruvate + non-essential amino acids + L-glutamine + 1x LIF (Sigma Aldrich) on gelatin-coated feeder-free tissue culture-grade plastic at 37°C and 5% CO_2. Medium was changed every 24 h and cells were regularly split when confluent (~1 to 10 ratio).

Generation of knock-in cell lines

R1 mESCs were plated on a 24-well plate (WP) (80 000 cells every well). After 24 h cells were transfected with mix of plasmids consisting of Cas9 expressing protein plasmid (450 ng), puromycin resistance reporter plasmid (200 ng), HA-mRuby2-mAgo2 HR template plasmid (350 ng) and mix of 2 sgRNA producing plasmids (400 ng in total, 200 ng each) using Lipofectamine 3000 Reagent (Thermo Fisher Scientific). 24 h after transfection cells were treated with puromycin (2 μg/ml) for 48 h (medium changed after 24 h and new puromycin containing medium added). Cells were expanded for 72 h and then washed with PBS, trypsinized and suspended in fresh culture medium. Fluorescently activated cell sorting (FACS) was used to select fluorescent positive cells but no fluorescence was observed. The concentration of the cells was determined by cell counting using the Bürker Chamber. Approximately 1500 cells each were transferred on three 15 cm dishes and cultured for 7 days to grow into colonies originating from single cells. The rest of the cells was collected and gDNA was isolated and screened for specific insertion of the desired KI by PCR with a locus specific primer and a sequence specific primer.

Individual single cell clones were transferred manually on 24-WP dishes, cultivated for 5 days and then samples were collected for gDNA which was screened for the KI sequence. PCR bands were sequenced (sequence reaction: 3 μM primer, 50 ng of DNA, ddH_2O to 10 μl).

Generation of HA-mRuby2-mAgo2_E1i1 endogenous knock-in cell lines

Cells were plated for transfection as described earlier, transfected with mix of plasmids consisting of Cas9 expressing protein plasmid (450 ng), puromycin resistance reporter plasmid (200 ng), HA-mRuby2-mAgo2 HR template plasmid (350 ng) and mix of 3 sgRNA producing plasmids (400 ng in total, 1:1:1 ratio of sgRNA-1, sgRNA-2 and sgRNA-3) following the same protocol as described earlier cells were selected for successful transfection and expanded. Cells were harvested and selected for fluorescent cells using FACS into 3 96-WP dishes. Cells were grown for 7 days, then trypsinized to spread the single cell colonies evenly on the whole well bottom surface and grown for three more days. After that samples of single cell clones were collected for gDNA and screened for specific knock-in. PCR products were sequenced as described earlier. Several positive clones were selected, transferred to 6-WP dishes and cultivated. Samples were collected and protein lysate were prepared to analyze by western blot analysis (described later). Positive clones were observed using fluorescent microscopy.

Fluorescently activated cell sorting

FACS was provided as a service by our in-house core microscopy facility. Cell sorting was done by Zdenek Cimburek using the BD Biosciences Influx sorter. Viable cells positive for the 585 nm and 610 nm fluorescent signal were collected into three gelatin coated 96-WP dishes.

Microscopy

Microscopes are provided by the in-house core microscopy facility. Images were obtained by the Leica DMI6000i inverted microscope with a confocal head Leica TCS SP8 using the HC PL APO 40X/1.30 oil objective and the Type-F immersion oil (Leica), with lasers set to 5% of power and by the Leica DMI6000i. All microscopy experiments were performed with live cells at 37°C with no special gasses supplied (air CO_2 concentration).

SDS-PAGE and western blotting (WB)

Cells were centrifuged at 10 000 RCF at 4°C for 5 min, supernatant was discarded, pelleted cells resuspended in ice cold PBS and centrifuged again. Cells were then resuspended in 1X RIPA buffer (25mM Tris, pH 7-8, 150 mM NaCl, 0.1% SDS (optional), 0.5% sodium deoxycholate, 1% Triton X-100) with a protease inhibitor cocktail (Sigma Aldrich) and lysed for 20 min on iced, mixed thoroughly every 5 min. Lysate was subsequently centrifuged at maximum speed for 15 min at 4°C.

Supernatants were collected and the protein concentration was determined by Bradford protein assay according to the Bio-Rad Bradford protein assay instruction manual. Dilution series from IgG protein standards and samples for measurements were prepared (1 μl of sample protein lysate, 200 μl of Bradford reagent (Bio-Rad) and 800 μl of water). After 10 min absorbance of samples was measured at 595 nm wavelength. The protein concentration was then calculated using linear regression.

Protein samples for sodium dodecyl sulphate poly-acryl amid gel electrophoresis (SDS-PAGE) were prepared by mixing 6 μl of 5X SDS sample loading buffer (250 mM Tris·HCl, pH 6.8, 10 mM

DTT, 0.05% Bromophenol Blue, 10% SDS and 30% Glycerol) with 70 μg of total protein and ddH_2O and heating them at 90°C for 10 min.

PAGE gels were prepared in a glass electrophoresis cassette (Dual Gel Caster – Hoefer). For 90+ kDA size proteins 6.0% separating gel was used. For lower sized proteins, 12% gel was used. In both cases, 5% stacking gels were used. Protein samples were loaded together with the ProteinRuler Plus Prestained Protein Ladder (Thermo Fisher Scientific) and resolved at 150-180 V in running buffer (pH 8.3, 25 mM Tris, 192 mM Glycine, 0.1% SDS in deionized H_2O).

For western blotting (WB) polyvinylidene difluoride (PVDF) membranes were activated in methanol, and then dipped in transfer semi-dry buffer (20% methanol, 1.25 mM Tris, $8 \leq pH \leq 10.0$, 9.6 mM Glycine in deionized H_2O). Gels were transferred into a blotting sandwich (3 filter papers – blotting sponge – gel – PVDF – membrane – blotting sponge – 3 filter papers; all sponges and papers were saturated in the semi-dry transfer buffer). Protein was transferred onto the membrane at 35 V for 60 min.

Membranes were marked by cutting lower right corner and if needed they were cut into pieces for detecting proteins of different sizes. Membranes were stabilized by dipping them in methanol, air dried and dipped in methanol again. Membranes were then blocked in 5% milk in TTBS buffer (pH 7.5, 150 mM NaCl, 10 mM Tris, 0.05% Twee 20 in deionized H_2O) for 60 min at room temperature. Membranes were drained and a primary antibody in 5% milk in TTBS was added for cultivation overnight at 4°C. Next day, membranes were washed 3-5 times in 5% milk in TTBS for 5 min at RT and then secondary antibody was added in 5% milk in TTBS for 60 min at room temperature. Membranes were washed 5 times in TTBS buffer for 5 min at room temperature. Membranes were developed using the SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific) according to manufacturer's manual. Exposure time typically ranged from 10 s to 10 min.

Antibodies

Primary antibodies:

anti-α-Tubulin	1:10 000 (T6074 Sigma Aldrich, Mouse, monoclonal antibody clone B-5-1-2)

anti-AGO2	1:1200 (EMBL ACE Mouse MA2 17F11C1 polyclonal antibody, kindly provided by Dónal O'Carroll)
anti-HA	1:2500 (12158167001 Roche, Rat, monoclonal antibody clone 3F10)

Secondary antibodies:

Anti-Mouse	1:50 000 (31430, Pierce (now Thermo Fisher Scientific) HRP, Goat, polyclonal antibody)
Anti-Rat	1:50 000 (31470 Pierce (now Thermo Fisher Scientific), HRP, Goat, polyclonal antibody)

Computational analysis

All annotated sequences were obtained from the mouse genome assembly GRCm38/mm10 using the UCSC genome browser (https://genome.ucsc.edu).

Next generation sequencing (NGS)

For GV oocyte transcriptome total RNA libraries were prepared from CB56Bl/6 mouse GV oocytes cloned into cDNA, amplified and sequenced on the Illumina IIx genome analyzer system. Mouse genome assembly GRCm38/mm10 was. used for mapping of 35 nt single-end and 76 nt pair-end reads. Sample preparation and NGS was done by our collaborators from Fugaku Aoki's laboratory and data were processed by Vedran Franke. Custom tracks were prepared from the data and can be visualized by the UCSC genome browser (https://genome.ucsc.edu). Datasets are available in the ArrayExpress database under reference (Abe et al., 2015).

Ribodepleted RNA was prepared from V6.5 mESCs using the mirVana miRNA isolation kit (Life technologies) by Richard A Young, cloned, amplified and sequenced by the Illumina HiSeq 2000.

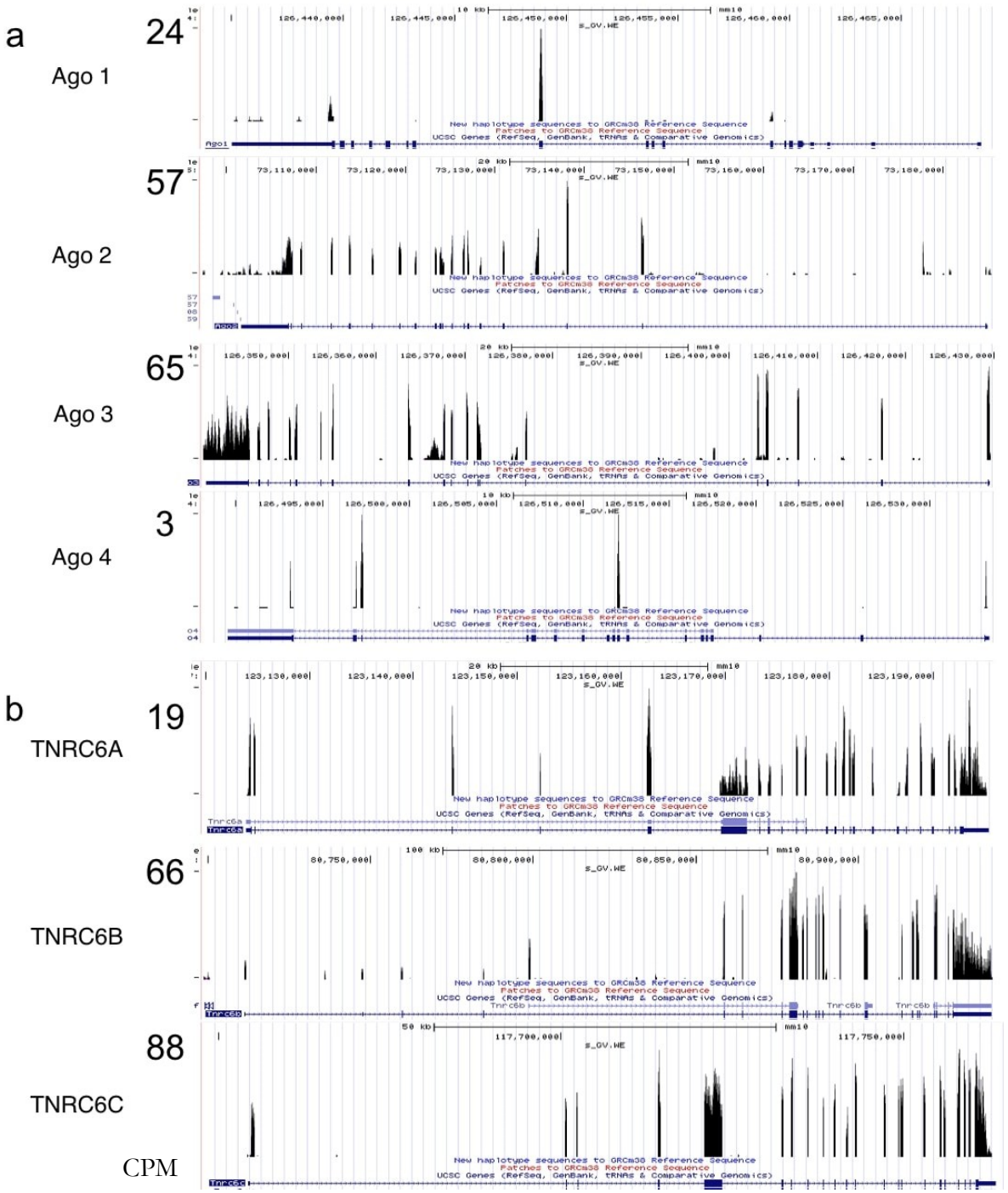
Data are available in the Gene Expression Omnibus database under the accession number GSM903663 . Data were mapped by Vedran Franke onto the NCBI37/mm9 mouse genome assembly and custom tracks were made to visualize the data using the UCSC genome browser.

Sanger sequencing results were processed and aligned with a consensus sequence using BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit.) or mapped on the GRCm38/mm10 genome assembly using BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

Results

Knock-in design

The first step in developing the miRNA pathway biosensor was to determine which specific factors to tag and use for the microscopy-based approach. The hallmark of the active miRNA pathways is the miRISC, which consist of an Argonaute protein, GW182 (in mice also called TNRC6) and the miRNA guide strand. Both Argonaute proteins and TNRC6 proteins in mice are encoded by several paralogs and labeling all the paralogs would not be feasible. There are four AGO proteins (AGO1-4) employed in miRNA pathway present in mice and three paralogs of TNRC6(A-C). Thus, I first checked available NGS data from GV oocytes to see the expression profiles of the Ago and Tnrc6 paralogs. From the NGS data alone the highest signal of any *Ago* had *Ago3* followed by *Ago2* (Figure 8a). The presence of AGO2 protein in the oocyte was confirmed in the laboratory by a western blot analysis by Shubhangini Kataruka (unpublished data, not shown). Based on the data and the fact that AGO2 is the catalytically active Ago protein facilitating the RNAi pathway in the oocytes and shows a phenotype in AGO2 knock-out mice (Kaneda et al., 2009), *Ago2* was chosen as a candidate for the biosensor. The NGS data were also used to determine the expression profiles of the *Tnrc6* paralogs. *Tnrc6a* is the least expressed in mice oocytes. *TNRC6B* and *TNRC6C* have the same mRNA expression levels (Figure 8b). Using the UCSC genome browser I compared the number and characteristics of the annotated isoforms of paralogs TNRC6B/C (Figure 8c). The TNRC6C has fewer annotated isoform with a conserved C-terminal end (with the exception of one short truncated isoform). The TNRC6B isoforms seem less conserved with larger variance in length and N- and C-terminal ends. I opted to first tag the TNRC6C isoform but also prepare the constructs needed for the tagging of the TNRC6B isoform. I also used NGS data from mESCs to verify that *Ago2* and *TNRC6C* are expressed in mESCs (Figure 8d).



Continues on the next page

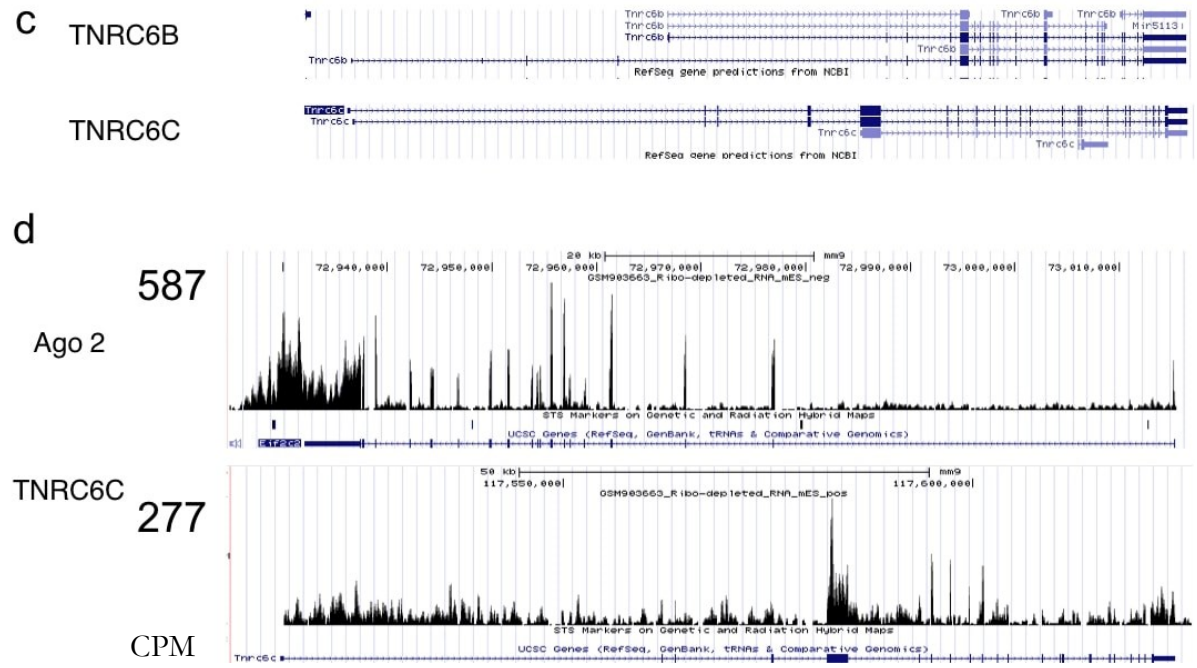


Figure 8 ***Expression profiles and characteristics of Argonaute and TNRC6 proteins in mouse GV oocyte and mESCs.***

Snaphots from the UCSC genome browser. On the X axis is the position within the gene, Y axis shows counts per milion (CPM). The peaks represent the density of the signal. Each snapchot is scaled to its maximum CPM. ***(a)*** *Expression of Argonaute proteins. Ago2 and Ago3 have by far the highest levels of mRNA expression of the Ago proteins in mouse GV oocytes. Ago2 was selected for the knock-in.* ***(b)*** *Expression of three homologs of Tnrc6 (also known as GW182) proteins in mouse GV oocytes. Tnrc6a has low levels of mRNA, expression of Tnrc6b and Tnrc6c are approximately the same.* ***(c)*** *Isoforms of TNRC6B and TNRC6C produced in mice, based on the number of isoform and conserved C-terminal exon the TNRC6C isoform was selected for the knock-in.* ***(d)*** *Ago2 and TNRC6C selected for the knock-ins are both expressed in mESCs. All Argonaute proteins are showed from 3'-end to 5'-end.*

Next, I proceeded to design the constructs for CRISPR/Cas9 driven homologous recombination knock-ins into the endogenous loci of *Ago2* and *Tnrc6c*. N-terminally tagged Ago showed stable previously, in the laboratory prepared by Matyas Flemr. From discussion with Gunter Meister we were aware that the C-terminus is problematic to tag because it is hidden inside the structure of the protein. Therefore, I decided to tag the N-terminal end of AGO2. TNRC6C has several annotated isoforms with alternative N-terminal ends however the C-terminal end seems more conserved therefore I chose to tag it.

Next, the fluorescent proteins were selected to be able to act as a FRET pair with the donor's emission spectra overlapping with the absorption spectra of the acceptor. A bright and stable pair of Clover and mRuby2 was chosen.

The constructs for homologous recombination consist of two homologous arms which direct the recombination and secure precise integration of the sequence they flank the genome. For efficient

homologous recombination I aimed for ~1000 nt in length of each arm. The arms are flanking the unique sequence of the knock-in – in this case the coding sequence of an epitope tag for easier selection and future needs, coding sequence of a fluorescent protein and a flexible glycine-serine linker to prevent steric hindrance in the fused proteins (sequence of the linker provided in the Suplementary information). Homologous arms were designed in a way that the coding sequence of the knock-in would be inserted precisely after the start codon (in case of *Ago2*) or the stop codon (in case of *Tnrc6c*) and therefore the resulting sequence codes fused proteins (Figure 9).

For each knock-in two sgRNAs were selected. The sgRNA sequences were design so they target the gDNA sequence but not the sequence of the homologous recombination constructs. This was done by mutating a guanine which corresponds to the PAM sequence (NGG to NGA) in the homologous recombination arm or by designing the sgRNA in a way that the sequence is disrupted in the homologous recombination template by the sequence of the fluorescent protein. It is key that CRISPR/Cas9 would not cleave the homologous recombination template, otherwise the efficiency of successful knock-in would be greatly diminished.

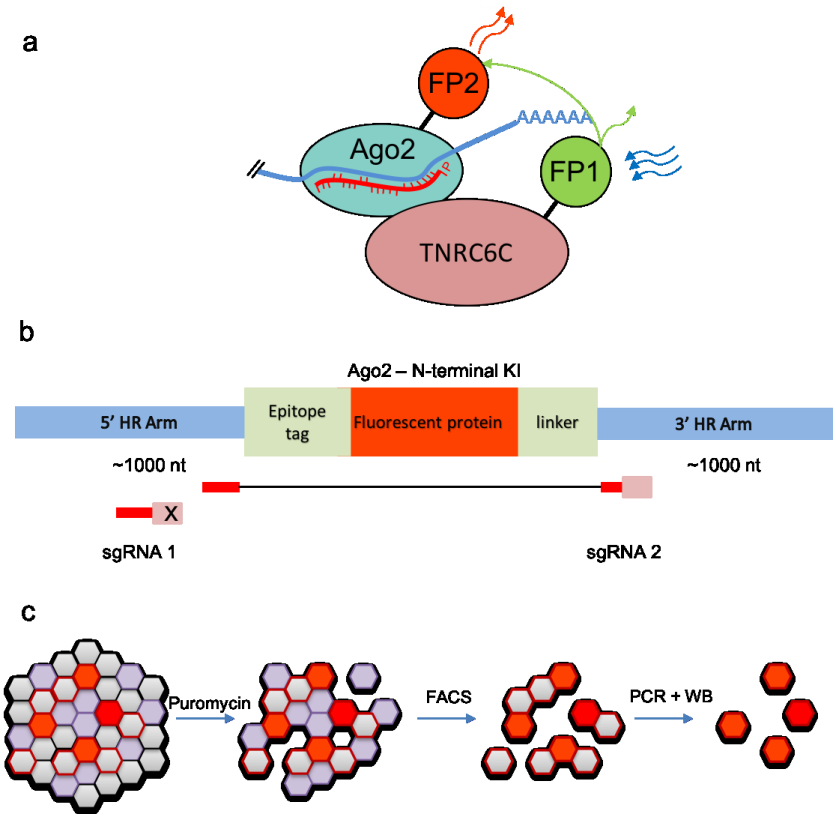


Figure 9 ***Schematic visualization of the miRNA pathway biosensor and the knock-in design.***
(a) *Schematic visualization of the miRNA pathway biosensor with working FRET between two fluorescent proteins. Clover (FP1) act as a donor of energy for the acceptor mRuby2 (FP2) which emits red light (excitation, FRET and emission*

represented by arrows). ***(b)*** *Schematic view of the knock-in design. The homologous arms are flanking an epitope tag sequence, fluorescent protein sequence and a GS flexible linker sequence. sgRNA1 is targeting the gDNA in the sequence homologous to the 5' HR arm, however in the arm the PAM region was mutated, sgRNA2 targeting sequence was in the HR template disrupted by the knock-in sequence.* ***(c)*** *Schematic representation of the knock-in generating pipeline.*

Argonaute 2 knock-in in mESCs

The cloning of *Ago2* homologous arms proved to be extremely difficult because of the high GC content (in the first 500 nt from the start codon approximately 78%) and low complexity of the sequence. Many attempts with various conditions with multiple sets of primers were tested but neither yield a desired product. After several months of unsuccessful attempts, it was decided to synthesise the sequence and a synthetic construct was ordered from Sigma Aldrich.

In parallel, I prepared expression vector of the tagged protein (HA-mRuby2-linker-mAgo2). After transfection I obtained fluorescent cells however I did not see a protein of the expected size on western blot (anti-HA antibody was used). Only after I examined smaller proteins range I observed at approximately 30kDa a fragment which would correspond to the size of HA-mRuby2 (Figure 10b). This suggests that Ago2 was separated from the HA-mRuby2 most likely in the linker sequence region. This was surprizing and unpredictable because the particular GS linker is routinely used in fused proteins. A few other researches personally (namely Ralph Grand from Friedrich Miescher Institute and Radek Malik from our group) confirmed that it has happened to their proteins too and that deletion of the linker generally remedies this effect. Thus, I re-cloned all of my homologous recombination templates without the linker sequence.

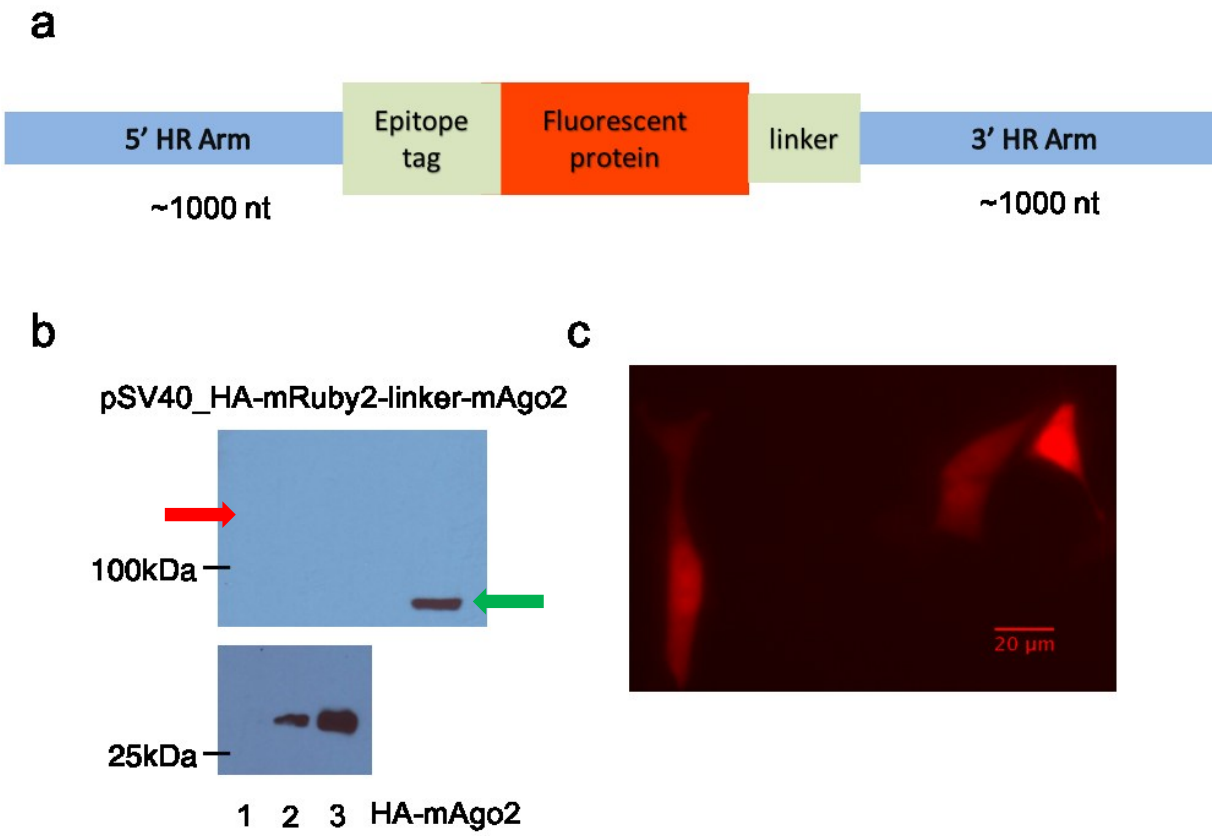


Figure 10 ***The HA-mRuby2 is separated from Ago2 in the linker sequence region***
(a) *Schematic visualization of the homologous recombination template for Ago2 knock-in* ***(b and c)*** *HA-mRuby2 is separated from AGO2 most likely by degradation of the linker.* ***(b)*** *Western blot analysis of transiently expressed HA-mRuby2-linker-mAgo2 construct shows missing of the expected signal (red arrow) in all three samples however positive signal is at size corresponding to the size of HA-mRuby2. As positive control was used HA-mRuby2 producing vector (green arrow).* ***(c)*** *Picture of the transient expression from the HA-mRuby2-linker-mAgo2 expression vector in NIH_3T3 cell line of mouse fibroblasts (snapshot was made using the Leica DMI6000i microscope, live cells).*

After transfection of new constructs for *Ago2* and *Tnrc6c* knock-ins and selection of transfected cells, surviving cells were analysed using flow cytometry. To my disappointment, no cells showed any fluorescent signal. In any case, single cell-originating clones were prepared and screened for the knock-in with positive result in 50% of the clones. The PCR amplicons from the screened clones were sequenced and showed integration of the sequence into *Ago2* locus with the sequence of the epitope tag directly following the start codon as desired. However, at the site of the sgRNA1 in the 5' UTR of *Ago2*, the sequences were disrupted by deletions which likely caused loss of the rest of the 5' UTR and disrupted the promotor region of the gene (Figure 11b). The finding suggests that the homologous recombination template was cleaved in the cells after transfection. The disruption of the homologous recombination template causing this is surprizing because the PAM region was mutated in the template and should have not been recognized by Cas9. The effect was likely brought about by the low complexity of the sequence which caused off targeting in the

region. This motion was supported by the fact that transfection of only sgRNA2 did not yielded any fluorescent positive knock-ins. It seems that a half of the homologous recombination template containing the 3' arm and the knock-in sequence gets incorporated into the genome by HR on the side with the 3' arm and the 5' end is joined by nonhomologous end joining. These findings pointed to the need of an alternative strategy for introducing the double stranded breaks to the gDNA in a way that would not disrupt the HR template. The same results were observed in the TNRC6C knock-ins.

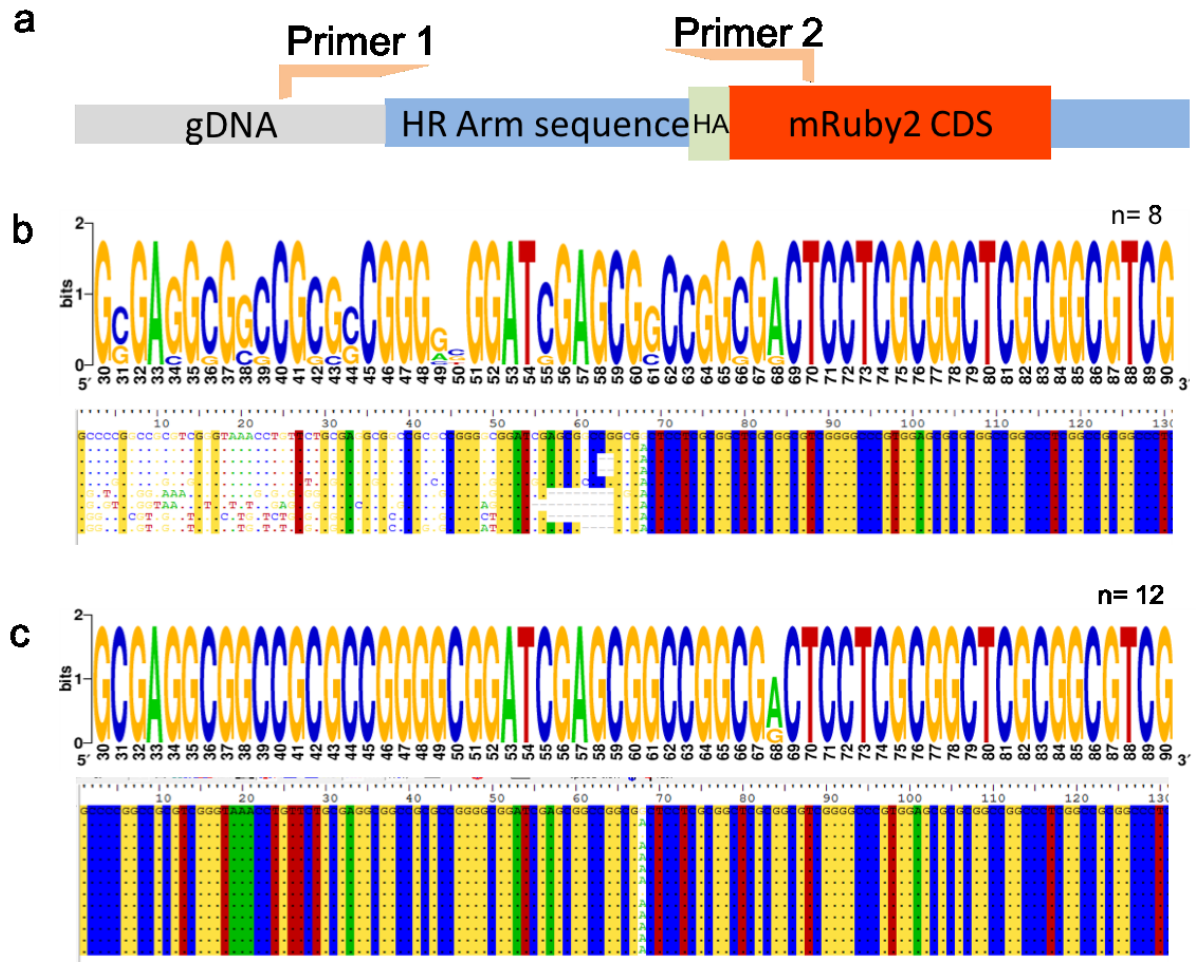


Figure 11 ***Obstructions during the knock-in preparation***

(a) *Schematic visualization of the locus and sequence specific PCR used to confirm the presence of the knock-in sequence.* ***(b)*** *Visual representation of the sequence clones positive for the knock-in sequence however negative for fluorescence after transfection using the sgRNA1. Position 68 shows the mutated PAM nucleotide of sgRNA1 and upstream of this nucleotide the mutation and deletion start while downstream the sequence is identical to the knock-in consensus of the prepared template. The mutations are likely cause by a partial degradation of the HR template by Cas9 resulting in partial homologous recombination. The beginning of the inserted coding sequence is not shown however it shows perfect complementarity to the consensus of the designed knock-in. First nt of the HA-tag would be at position 146. Represented as a sequence logo (top, n= 8) and aligned sequences to the consensus sequence (bottom).* ***(c)*** *Similar visualization of sequence characteristics as in (d) by sequence logo (n =12) and aligned sequences. Sequence and fluorescence positive clones showed in (d) originate from cells transfected by the new E1i1 HR template. Surprisingly in two sequenced clones the PAM has its original sequence found in WT cells. This might be due to*

termination of the homologous heteroduplex joint. Otherwise all other positions have perfect complementarity of base pairs to the consensus sequence. Sequence logos were generated by WebLogo, https://weblogo.berkeley.edu (Crooks et al., 2004).

New Ago2 knock-in template

After the problems with the homologous recombination constructs, which were described earlier and caused a significant delay of the project, I decided to re-design the template so that the whole guide sequence of any sgRNA would not be present in the template and the sgRNA target sequences would have higher complexity. I decided to make a short deletion of 117 nt in the intron 1 of *Ago2*. The deletion starts 250 nt downstream of the exon-intron junction to minimize the potential effect on the gene regulation and splicing (Figure 11a). I used the sequence of this deletion to target there all three sgRNAs that I used later on. After I cloned all the necessary plasmids I transfected R1 mESCs with the new HR template (which I call HA-mRuby2-mAgo2_E1i1) together with the new sgRNAs.

After selection I used flow cytometry and FACS and finally obtained fluorescent cells (Figure 12b) which represented ~6% of the total population. Next, many single-cell originating clones were cultured and screened. Gel-isolated DNA from several positive PCR clones was sequenced and analysed. Most of the samples showed perfect complementarity to the consensus sequence throughout the length of the knock-in and the homologous arms (Figure11c). Because of the way the new HR template was prepared the former PAM substitution is present in the new template as well. Surprisingly in two (out of 12) sequenced clones the substitution is excluded, and the original guanine is still present (Figure 11c).

Few selected clones were used for western blot analysis with anti-HA antibody, the analysis confirmed the presence of the fused HA-mRuby2-mAgo2 protein in most of the clones (Figure 12c). Western blot analysis was repeated with AGO2 specific antibody to distinguish heterozygous and homozygous clones (Figure 12d). Six homozygous clones were obtained which were used for microscopy (Figure 12e)

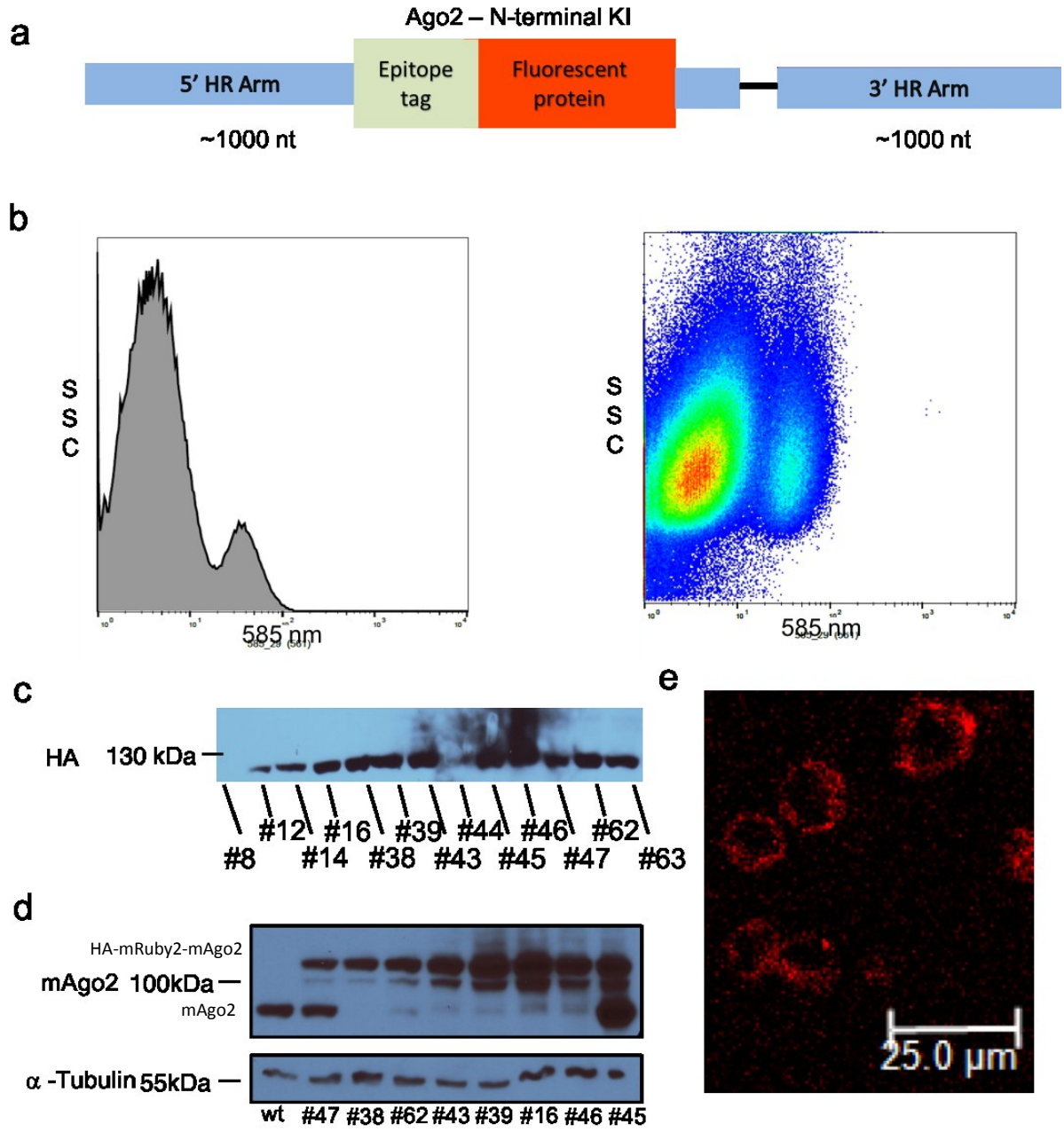


Figure 12 ***HA-mRuby2-mAgo2 knock-in of the endogenous locus in mESCs.***

(a) *Schematic visualization of the new HA-mRuby2-mAgo2_E1i1 HR template with the deletion in the template (represented by a black bar).* ***(b)*** *Results of FACS visualized by two plots. The distinct population of fluorescently positive cells equate to ~6% of the cells. X axis is relative fluorescence emission of a 585 nm light, Y axis shows the side scatter (SSC) parameter.* ***(c)*** *Western blot analysis of fluorescently positive clones. Clones #8, #12 and #44 were found negative for the AGO2 knock-in protein and not cultured anymore, rest further analysed. Anti-HA antibody used.* ***(d)*** *Western blot analysis of several fluorescent positive clones. Far right sample is a WT R1 mESCs control followed by 8 clones. clone clones #47 and #45 are likely heterozygotes, remaining clones are likely heterozygotes. The signal in the likely homozygotic clones at the WT size is most likely brought about by products of degradation. Anti-AGO2 antibody used.* ***(e)*** *Microscopic snapshot of the HA-mRuby2-mAgo2_E1i1 #38 clone (taken by Leica DMI6000i with confocal head, live cells).*

TNRC6C knock-in in mESCs

The *Tnrc6c* knock-in experienced similar pitfalls as the *Ago2* knock-in. The linker was removed as a precaution from the homologous template based on the results from the fusion AGO2 expression vector. The homologous template without the linker yielded cells, which were positive for the knock-in sequence in the PCR screen (Figure 13b-c), however did no show and fluorescence similarly to the *Ago2* knock-in, because the integration of the knock-in sequence into the locus was not perfect (Figure 13d).

Thus, I designed new homologous template, which includes complete deletion of the intron21 (727 nt) resulting in the last two exons of *Tnrc6c* being joined together. Six sgRNAs were designed to target the sequence of the intron21 in the gDNA. At the time of writing the construct was, however, not completely ready.

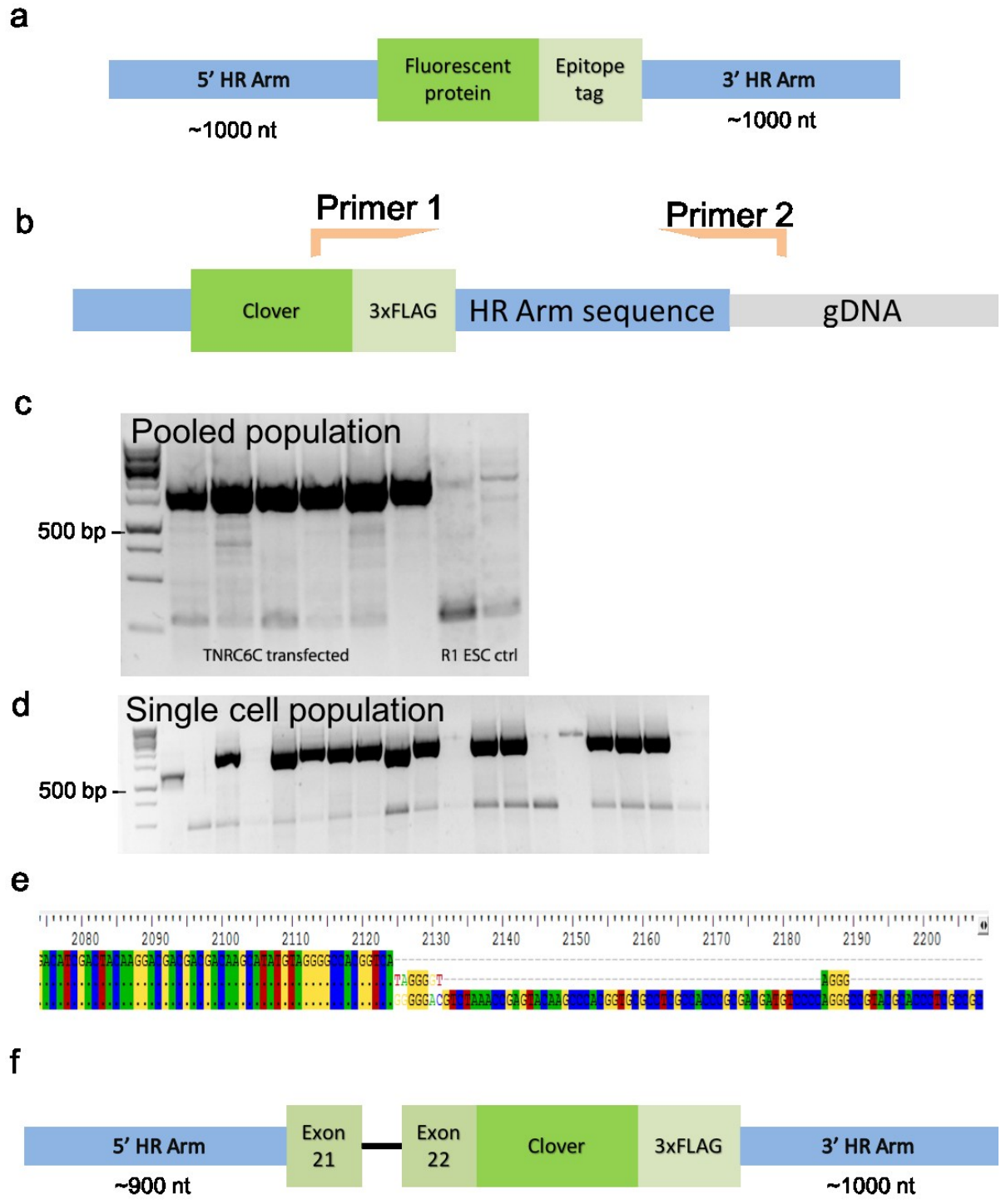


Figure 13 ***Tnrc6c knock-in***

(a) *Schematic visualization of the Tnrc6c homologous template without the linker, which was transfected in mESCs.* ***(b)****Schematic visualization of the PCR design used for knock-in verification* ***(c-d)*** *PCR screen of gDNA using sequence specific and locus specific primers. gDNA was isolated from pooled mESCs populations* ***(c)*** *and later from single cell originating populations* ***(d)****.* ***(e)*** *Visualization of the sequencing results of two gDNA PCR amplicons from single-cell populations of the Tnrc6c knock-in. First line represents the consensus sequence of the knock-in. The results show integration of undesired sequences originating from the cloning plasmid backbone after the 3' end of the 3xFLAG coding sequence.* ***(f)*** *Schematic visualization of the newly designed template with the last intron of Tnrc6c deleted resulting in joined exon21 and exon22.*

Discussion

In this thesis I present results of a project aimed at tagging endogenously expressed genes encoding miRISC components. I design a strategy for generating cell lines with stable expression of fusion fluorescent proteins primarily on HA-mRuby2 tagged AGO2 and discuss potential pitfalls of the endeavor. The fluorescently tagged AGO2 will be a part of the miRNA pathway biosensor, which will help to elucidate the lack of functionality ad significance of the miRNA pathway in the mouse oocyte. At the time of writing the TNRC6C knock-in was not ready. The TNRC6C knock-in procedure faced similar troubles as the Ago2 knock-in throughout the process and the last set of constructs was delayed until the procedure was troubleshooted in Ago2. Accordingly, I focused on the Ago2 constructs to test if the newly adopted approach would bring the desired goal. In the end, the new strategy worked well for Ago2, thus the *Tnrc6c* knock-in should be ready soon. The results presented here are important, because, to my knowledge, fluorescent knock-ins of the endogenous miRISC components were not generated so far.

Achieved aims

The design of the miRISC biosensor started by choosing paralogs of the miRISC components to tag. For that I used available NGS data. I examined NGS data from GV oocytes and mESCs avaiable in the lab (Figure 8). *Ago2* paralog was chosen because of the expression levels and its parallel role in RNAi as the only AGO protein involved in the pathway which is essential in the oocyte (Kaneda et al., 2009; Schmitter et al., 2006). The oocyte-specific knock-out mice of *Ago2* results in oocytes with abnormal spindles causing developmental arrest. The phenotype is embryonically lethal with majority of fertilized oocytes failing to divide into two-cell stage (Kaneda et al., 2009). For the dominance of *Ago2* over other paralogs and reasons mentioned above, having endogenously tagged Ago2 mouse model available would be beneficial for future research of the laboratory on both the miRNA and the RNAi pathways. For Ago2 tagging the N-terminal end was chosen because it was shown that it is stable (Pillai et al., 2004) and because the C-terminal end is hidden inside the protein making it impossible to fuse anything to it (Schirle and MacRae, 2012). Tnrc6c was chosen based on the expression levels and the annotated isoforms. The characteristics of those isoforms suggested that tagging might be more beneficial of the C-terminal end because it is shared among most of the isoforms. It is in our interest to have as much as possible of the protein tagged to obtain strong signal of fluorescence. Also, some isoforms might lack function.

For the successful production of fusion proteins from endogenous loci, it is necessary to have a single nucleotide precision of the introduced coding sequence so the knock-in would not result in frameshifts and nonsense mediated decay of the newly produced mRNA. This was achieved by HDR where, in the case of the Ago2 knock-in, the coding sequence for the HA-mRuby2 tag was fused to the exon1 of the *Ago2* gene and the endogenous 5'UTR in the homologous template. During the HDR, the template sequence replaces the homologous sequence at the locus including the exon1 and the 5'UTR. The newly integrated sequence acts as new exon 1 of the fusion protein, it is transcribed by Pol II from the endogenous promoter of *Ago2* and is spliced to *Ago2* exon2 creating an mRNA which codes for the full-length HA-mRuby2-Ago2 protein. To increase the occurrence of HDR, I introduced DSB of DNA in a targeted manner by CRISPR/Cas9 (Wang et al., 2015). For every knock-in I used more than one targeting sgRNA to ensure at least one functional targeting sgRNA per locus. In some cases, the sgRNA guide sequences overlapped each other, which should further increase the efficiency of the knock-in even more (Jang et al., 2018).

The knock-in procedure turned out to be much more difficult than expected initially and it was riddled with obstructions though the process. This lead to a significant delay of the project. As a result, at the time of writing the thesis, the final biosensor cell line was still not ready. The cloning of necessary homologous arms from the gDNA for HDR was delayed because of the low complexity of the sequence, which caused the failure of numerous PCR cloning attempts. The homologous arms were essential for the specificity of integration of the knock-in sequence in the genome. The sequence with the lowest complexity causing most of the difficulties was directly flanking the start codon of the *Ago2* gene and included a part of the 5' UTR. Therefore, it had to be included in the construct. Bypassing this sequence by introducing an alternative 5' UTR would inevitably alter the endogenous regulation of *Ago2* expression, which would be undesirable if we want the biosensor as a technique for studying the miRNA pathway in near endogenous conditions. For this reason, it was not possible to skip the sequence and extensive work was done to obtain it. However, as I did not succeed to amplify it, we overcame this problem and had the first construct for the Ago2 knock-in synthetized by Sigma Aldrich. Notably, the chemical synthesis of the construct took the company several weeks as the sequence did not pass the quality control multiple times.

An unpleasant surprise was the separation of the HA-mRuby2 protein from the Ago2 protein when expressed in cultured cells caused by the instability of the fused protein in the linker region; (the chosen used GS linker was developed by (Waldo et al., 1999)) and there was no indication in the

literature of problem with this linker (Figure 10). A GS flexible linker was used because GS linkers have a long history of use in fusion proteins and the proteins are generally reported stable (reviewed in (Chen et al., 2013). While my case might be rather unique, my experience shows the importance of careful validation of all prepared fusion proteins before using them for experiments and drawing conclusion based on the data obtained with them. The linker issue caused another delay of the project because I had to redesigning the constructs and additional cloning and verification of the templates to get rid of the linker sequence. The sequence was previously introduced in the constructs as a non-complementary sequence in the primer for the fluorescent protein and there was no restriction site between the fluorescent protein sequence and the linker (similarly in the synthetized plasmid). I did not want to use available kits for short deletion in plasmids because they are based on PCR amplification of the sequence and that proved to be a pitfall before in preparation of the homologous arm. I therefore cloned the sequences of the fluorescent proteins (without the linker but with the epitope tags on the other site) and exchanged the whole fluorescent protein sequences in respective constructs. The separation in the linker region was observed in cells transiently expressing the fused protein from an expression vector and unfortunately, at the time I stopped the cultivation of cells transfected with the old linker-containing constructs before I screened them. Because of that, I realized that the sgRNAs are cleaving the homologous template only after the transfection of the new linker-free constructs.

While designing the first constructs and the sgRNAs I did not fully appreciate how bad is the sequence which I had been working with for targeting by CRISPR/Cas9. I argued that the mutated PAM or the disruption of the sequence by the HA-mRuby2 sequence in the HR templates would suffice to prevent the cleavage of the HR template plasmids by Cas9 after transfection (Jinek et al., 2012). The 18 nt targeting sequence of sgRNA2 for *Ago2* was split to 14 and 4 nt – in the middle of the sgRNA seed region. However, in spite of that, the HR templates were getting cleaved in cells which led to the separation of the 5' Arm from the template and ultimately to none or incomplete homologous recombination. That produced cells with missing 5'UTR and possibly part of the promoter region of the *Ago2* locus and therefore not expressing any fluorescent protein (Figure 11). Use of only one of the two sgRNAs did not improve the results. Therefore, I decided to rebuild my constructs again. I suggest that the cleavage is due to the high GC content and low complexity of the sequence and hence the sgRNAs bind the sequence imprecisely at the locus. Higher GC content in sgRNAs corelates with higher off-targeting rate on the genome level (Lin et al., 2014; Morgens et al., 2017; Peng et al., 2016) which is consistent with what we observed in small-scale in the locus.

Finally, short deletion of 117 nt in the intron 1 of *Ago2* (which was located in the 3' arm of the previous construct) and extension of the 3'arm downstream was prepared in combination with three new sgRNAs. These target the gDNA in the *Ago2* gene corresponding to the deletion in the HR template. Therefore, the HR construct does not get cleaved. In addition, the sequence of the deletion has higher complexity which should improve the targeting of the sgRNAs to the locus and reduce off-targeting. Because the deletion is 250 nt downstream of the exon-intron junction the impact on the regulation of expression and splicing should be minimal. However as with any other change to the genome some effect can unavoidably occur (for more information see (Chorev et al., 2017; Jo and Choi, 2015)).

For the delivery of the HR template I used circular dsDNA in a form of a plasmid. It is not the most efficient method and it has been shown that single-stranded DNA (ssDNA) templates yield higher percentage of positive cells. However, preparing long (~2-3kb) ssDNA is not trivial and would take additional time. Linearized plasmid DNA or linearized template sequence (e. g. PCR products) could potentially improve the yield as well. However, because I was working with fast dividing mESCs the number of cells was not a limiting factor for me. Also, the integration of the template via the homology directed repair was still relatively high (~6% after puromycin selection) Because the reasons mentioned above I decided to use the circular plasmid and not prepare the ssDNA or linearize the template. The availably of FACS at my disposal was of course a factor in this decision as well. The mouse model will be prepared by injection of mESCs expressing the fused proteins into blastocysts and generating chimeric mouse so the efficiency of the HDR won't matter much either. If for some reason we decide to introduce the knock-in by direct injections of DNA and RNA into the zygote the newly developed methods using the ssDNA templates would be beneficial (Miura et al., 2018; Quadros et al., 2017). This approach could however prove challenging as the introduced KI sequence is relatively long (735 nt) which will lower the efficiency of the HDR.

After FACS the generated cell lines were selected by locus specific PCR however they would benefit from a southern blot analysis to see whether there are any random integrations of the template in the genome. The fluorescent protein is almost definitely produced only in the form of the fused protein with Ago2 otherwise signal would be detected during the western blot analysis with the HA antibody (Figure 12c). The possibility of integrations to regulating sequences or in

coding regions cannot be eliminated completely and the integrations might alter the genome regulation in the cells, nonetheless the probability is relatively low.

Because of the way the new homologous recombination template was prepared (the 5' arm was used from the older template) the former PAM substitution is present in the new template as well. Surprising in two (out of 12) sequenced clones the substitution is excluded, and the original guanine is present instead in the sequence. This is not an artefact of the sequencing, I hypothesize that the heterologous duplex joint during the homologous recombination was terminated before it reached the locus and therefore the guanine is not mutated. The whole HA-mRuby2 sequence is present in the genome and the start of the sequence is only 78 nt from the former PAM so the termination had to happen almost immediately after the HA-mRuby2 sequence.

The PCR selected clones were analyzed by WB. The anti-AGO2 antibody was used to determine homozygotic and heterozygotic clones. Indeed, six clones, from the first tested sample of 8, are likely homozygotic and two heterozygotic. The clones were previously tested using anti-HA antibody so all clones negative for HA signal were eliminated prior to the experiment. There is a weak signal at the WT AGO2 size which is probably a result of degradation. The signal is marginal to influence the results of any planned experiments. Right below the signal of the expected size in the samples positive for the fused protein there is another signal that could be a result of degradation as well. This signal was not observed with the use of anti-HA antibody which together with the size of the detected protein suggest that the degradation might occur within the mRuby2 protein. It is not clear at which point the potential degradation occurs as it could be in the living cells or during the sample preparation. Nonetheless the degradation signals are weak in comparison to the full-size fused protein signal therefore the fused protein is relatively stable for the planed experiments.

As it is endogenously regulated protein the fluorescence of the fused protein is relatively low. Which was however expected and won't influence the microscopic experiments such as FRET or FCCS. For the single-molecule methods low levels of fluorescence are beneficial as it is easier to track the molecules. Mouse ESCs which were used to generate the knock-ins are not an optimal cell line for microscopy of cytoplasmic proteins. The cells are relatively small and most of the volume of the cells is occupied by the nucleus nonetheless the esthetics of the model system are not our major concern.

Future plans

At the time of writing the biosensor was not ready due to the circumstances mentioned earlier. Once the complete biosensor cell line is finished it will be stimulating to see if FRET will work in the setup of AGO2 and TNRC6C. The method is sensitive to the distance of the two fluorophores and efficiency of the energy transfer decreases inversely with the distance to the power of six. This in practical term means that the two fluorophores have to be approximately 10 nm or closer from each other for detectable FRET (Sekar and Periasamy, 2003). This has to be addressed by trial after the biosensor is ready. Apart from the distance of the fluorophores FRET is dependable on the characteristics of the used fluorophores such as the quantum yield and extinction coefficient (Ishikawa-Ankerhold et al., 2012). I consulted the literature available on fluorescent FRET-pairs and selected fluorescent proteins mRuby2 and Clover. The pair was chosen based on the quantum yield, extinction coefficient, fluorescent stability and their larger distance from each other that is tolerable for FRET as well as their accessibility from Addgene ((Lam et al., 2012). If FRET proves to work, we can measure the standard FRET based on the intensity of the fluorescence which can however be affected by experimental environment. We can also use combination of fluorescence lifetime imaging microscopy (FLIM) and FRET to overcome the limitations and artefact that intensity-based FRET brings. FLIM measures the time needed for the fluorophore to decrease its intensity to 1/e and this method is insensitive to cross-contamination of signal, concentration of fluorophores or the variation in excitation intensity and time (Ishikawa-Ankerhold et al., 2012). Second method to see the interaction between Ago2 and TNRC6C is the FCCS which is FRET independent and should work even if FRET doesn't. The method is based on tracking photons emitted from both fluorophores at spatiotemporal resolution and correlating their appearance using statistical analysis (Schwille et al., 1997) to predict their interaction. We are planning to combine these methods or use only FCCS in case FRET is not obtainable. The methods can with very low concentration of the fluorophores which probably will be the case of the oocyte model.

Once the biosensor is developed it has to be verified by pull-down experiments of each of the fused proteins followed by western blotting and preferably also mass spectrometry analysis to see if the components of the full RICS and its binding partners (such as CCR4-NOT and PABPC) coprecipitate with the pulled-down protein. The effect of the fused proteins on the miRNA pathway should be assessed by luciferase assays. We hope that the fluorescent proteins won't have any major effect and the tag proteins will still be active but thorough verification is needed.

If the biosensor passes the verification it can be used for screenings of miRNA inhibitors in the mouse embryonic stem cell line. Inhibitors can be used to study the miRNA pathway further and perhaps could lead to interesting applications in the clinical research as well. The main purpose is to study the interaction of the AGO2 and TNRC6C in the mouse oocytes where the pathway ceases to be active during the growth of the oocyte. From a technical standpoint the protein levels of the biosensor components could be an issue for microscopy however single-molecule FCCS (and possibly FRET) should work well. The stoichiometry of the components of the pathway is a major candidate for the impaired functionality in fully grown oocytes. The biosensor would be the most sensitive technique for monitoring the miRNA pathway and it should lower the threshold for sensing the activity. It will be interesting to see if the biosensor detects any residual miRNA pathway activity in the fully grow oocyte.

By microinjecting miRNAs into the GV oocytes and studying if the miRNA pathways increases its function we could determine whether the miRNA pathway is regulated by the number of miRNAs present. It was reported that even the most abundant endogenous miRNAs in the oocyte are not capable of suppressing their targets as mentioned earlier (Ma et al., 2010) however, the biosensor will provide better sensitivity for the testing. Alternatively to the miRNAs levels, the effect could be brought about by the low abundancy of the protein components of the pathway. The low abundancy might be caused by the inflation of the cytoplasmic volume and the disruption of the stoichiometry or, as mentioned earlier, by alternative isoforms of the components (Freimer et al., 2018). Injection of the fused proteins or their mRNA (alone or with miRNAs) into the oocyte and studying whether it rescues the effect might be a step as well. If the activity would not be rescued that would indicate a presence of some outside inhibitor. On the other hand, if the activity would be rescued it would not help reject any hypothesis because of the introduction of not modified proteins and the disruption of the stoichiometry. Moreover, the inflated protein levels could dilute the potential inhibitors and mask the true nature of the regulation.

By looking at the earlier stages of the developing oocyte and mapping the miRNA pathway activity in them should provide us with a temporal resolution of the activity. A gradual decrease in the activity would point towards the stoichiometry hypothesis rather than to a rapid inhibitory miRNA switch-off. However, gradual replacement of the protein components by their miRNA-pathway impaired isoforms would not be detected this way. The biosensor alone cannot answer all the questions we have regarding the miRNA pathway in the oocytes, but it will be a strong technique to elucidate some of them. I personally think that the inactivity of the miRNA pathway is caused

by combination of many aspects including the stoichiometry, protein modifications and possibly by other regulation factors.

Even eight years after the genetic evidence that miRNAs are not important for development and fertilization of oocytes and the development of early embryos we still do not understand why exactly the pathways is non-functional and is dispensable. This is mostly given by the biology of the oocyte and by the unique environment which it represents and by the fact that it cannot be simulated in any other model system. In the laboratory my colleagues are extensively working on the topic of the stoichiometry and Ago regulations with promising results. Other groups are also working on the miRNA pathway regulations so hopefully the next few years will elucidate some of the questions and bring us closer to understanding the fascinating processes happening in the of oocyte biology and the oocyte to embryo transition.

List of References

A., U. (2008). Preparation of Chemical Competent Cells (Untergasser's Lab).

Abe, K., Yamamoto, R., Franke, V., Cao, M., Suzuki, Y., Suzuki, M.G., Vlahovicek, K., Svoboda, P., Schultz, R.M., and Aoki, F. (2015). The first murine zygotic transcription is promiscuous and uncoupled from splicing and 3' processing. EMBO J *34*, 1523-1537.

Ameres, S.L., Horwich, M.D., Hung, J.H., Xu, J., Ghildiyal, M., Weng, Z., and Zamore, P.D. (2010). Target RNA-directed trimming and tailing of small silencing RNAs. Science *328*, 1534-1539.

Ameres, S.L., and Zamore, P.D. (2013). Diversifying microRNA sequence and function. Nat Rev Mol Cell Biol *14*, 475-488.

Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature *513*, 569-573.

Andersson, M.G., Haasnoot, P.C., Xu, N., Berenjian, S., Berkhout, B., and Akusjarvi, G. (2005). Suppression of RNA interference by adenovirus virus-associated RNA. J Virol *79*, 9556-9565.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T.*, et al.* (2006). A novel class of small RNAs bind to MILI protein in mouse testes. Nature *442*, 203-207.

Aravin, A.A., Hannon, G.J., and Brennecke, J. (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science *318*, 761-764.

Auyeung, V.C., Ulitsky, I., McGeary, S.E., and Bartel, D.P. (2013). Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. Cell *152*, 844-858.

Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P., and Blelloch, R. (2008). Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes Dev *22*, 2773-2785.

Ballarino, M., Pagano, F., Girardi, E., Morlando, M., Cacchiarelli, D., Marchioni, M., Proudfoot, N.J., and Bozzoni, I. (2009). Coupled RNA processing and transcription of intergenic primary microRNAs. Mol Cell Biol *29*, 5632-5638.

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. Cell *136*, 215-233.

Bartel, D.P. (2018). Metazoan MicroRNAs. Cell *173*, 20-51.

Bellutti, F., Kauer, M., Kneidinger, D., Lion, T., and Klein, R. (2015). Identification of RISC-associated adenoviral microRNAs, a subset of their direct targets, and global changes in the targetome upon lytic adenovirus 5 infection. J Virol *89*, 1608-1627.

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

Bernstein, E., Kim, S.Y., Carmell, M.A., Murchison, E.P., Alcorn, H., Li, M.Z., Mills, A.A., Elledge, S.J., Anderson, K.V., and Hannon, G.J. (2003). Dicer is essential for mouse development. Nat Genet *35*, 215-217.

Bogerd, H.P., Karnowski, H.W., Cai, X., Shin, J., Pohlers, M., and Cullen, B.R. (2010). A mammalian herpesvirus uses noncanonical expression and processing mechanisms to generate viral MicroRNAs. Mol Cell *37*, 135-142.

Bohnsack, M.T., Czaplinski, K., and Gorlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. RNA *10*, 185-191.

Braun, J.E., Huntzinger, E., Fauser, M., and Izaurralde, E. (2011). GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. Mol Cell *44*, 120-133.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089-1103.

Brustikova, K., Sedlak, D., Kubikova, J., Skuta, C., Solcova, K., Malik, R., Bartunek, P., and Svoboda, P. (2018). Cell-Based Reporter System for High-Throughput Screening of MicroRNA Pathway Inhibitors and Its Limitations. Front Genet *9*, 45.

Bussing, I., Slack, F.J., and Grosshans, H. (2008). let-7 microRNAs in development, stem cells and cancer. Trends Mol Med *14*, 400-409.

Cai, X., Hagedorn, C.H., and Cullen, B.R. (2004). Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. RNA *10*, 1957-1966.

Catalanotto, C., Azzalin, G., Macino, G., and Cogoni, C. (2000). Gene silencing in worms and fungi. Nature *404*, 245.

Cazalla, D., Xie, M., and Steitz, J.A. (2011). A primate herpesvirus uses the integrator complex to generate viral microRNAs. Mol Cell *43*, 982-992.

Chekulaeva, M., Mathys, H., Zipprich, J.T., Attig, J., Colic, M., Parker, R., and Filipowicz, W. (2011). miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. Nat Struct Mol Biol *18*, 1218-1226.

Chen, X., Zaro, J.L., and Shen, W.C. (2013). Fusion protein linkers: property, design and functionality. Adv Drug Deliv Rev *65*, 1357-1369.

Chiang, H.R., Schoenfeld, L.W., Ruby, J.G., Auyeung, V.C., Spies, N., Baek, D., Johnston, W.K., Russ, C., Luo, S., Babiarz, J.E*., et al.* (2010). Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev *24*, 992-1009.

Chorev, M., Joseph Bekker, A., Goldberger, J., and Carmel, L. (2017). Identification of introns harboring functional sequence elements through positional conservation. Sci Rep *7*, 4201.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188-1190.

Czech, B., and Hannon, G.J. (2016). One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. Trends Biochem Sci *41*, 324-337.

Czech, B., Zhou, R., Erlich, Y., Brennecke, J., Binari, R., Villalta, C., Gordon, A., Perrimon, N., and Hannon, G.J. (2009). Hierarchical rules for Argonaute loading in Drosophila. Mol Cell *36*, 445-456.

Dayeh, D.M., Kruithoff, B.C., and Nakanishi, K. (2018). Structural and functional analyses reveal the contributions of the C- and N-lobes of Argonaute protein to selectivity of RNA target cleavage. J Biol Chem *293*, 6308-6325.

De La Fuente, R. (2006). Chromatin modifications in the germinal vesicle (GV) of mammalian oocytes. Dev Biol *292*, 1-12.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev *15*, 188-200.

Elkayam, E., Faehnle, C.R., Morales, M., Sun, J., Li, H., and Joshua-Tor, L. (2017). Multivalent Recruitment of Human Argonaute by GW182. Mol Cell *67*, 646-658 e643.

Eulalio, A., Behm-Ansmant, I., Schweizer, D., and Izaurralde, E. (2007). P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. Mol Cell Biol *27*, 3970-3981.

Fabian, M.R., Cieplak, M.K., Frank, F., Morita, M., Green, J., Srikumar, T., Nagar, B., Yamamoto, T., Raught, B., Duchaine, T.F.*, et al.* (2011). miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT. Nat Struct Mol Biol *18*, 1211-1217.

Fabian, M.R., Mathonnet, G., Sundermeier, T., Mathys, H., Zipprich, J.T., Svitkin, Y.V., Rivas, F., Jinek, M., Wohlschlegel, J., Doudna, J.A.*, et al.* (2009). Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. Mol Cell *35*, 868-880.

Fang, W., and Bartel, D.P. (2015). The Menu of Features that Define Primary MicroRNAs and Enable De Novo Design of MicroRNA Genes. Mol Cell *60*, 131-145.

Farh, K.K., Grimson, A., Jan, C., Lewis, B.P., Johnston, W.K., Lim, L.P., Burge, C.B., and Bartel, D.P. (2005). The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. Science *310*, 1817-1821.

Felix, M.A., Ashe, A., Piffaretti, J., Wu, G., Nuez, I., Belicard, T., Jiang, Y., Zhao, G., Franz, C.J., Goldstein, L.D.*, et al.* (2011). Natural and experimental infection of Caenorhabditis nematodes by novel viruses related to nodaviruses. PLoS Biol *9*, e1000586.

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

Flemr, M., and Buhler, M. (2015). Single-Step Generation of Conditional Knockout Mouse Embryonic Stem Cells. Cell Rep *12*, 709-716.

Flemr, M., Ma, J., Schultz, R.M., and Svoboda, P. (2010). P-body loss is concomitant with formation of a messenger RNA storage domain in mouse oocytes. Biol Reprod *82*, 1008-1017.

Flemr, M., Malik, R., Franke, V., Nejepinska, J., Sedlacek, R., Vlahovicek, K., and Svoboda, P. (2013). A retrotransposon-driven dicer isoform directs endogenous small interfering RNA production in mouse oocytes. Cell *155*, 807-816.

Franke, V., Ganesh, S., Karlic, R., Malik, R., Pasulka, J., Horvat, F., Kuzman, M., Fulka, H., Cernohorska, M., Urbanova, J.*, et al.* (2017). Long terminal repeats power evolution of genes and gene expression programs in mammalian oocytes and zygotes. Genome Res *27*, 1384-1394.

Freimer, J.W., Krishnakumar, R., Cook, M.S., and Blelloch, R. (2018). Expression of Alternative Ago2 Isoform Associated with Loss of microRNA-Driven Translational Repression in Mouse Oocytes. Curr Biol *28*, 296-302 e293.

Friedman, R.C., Farh, K.K., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res *19*, 92-105.

Fromm, B., Billipp, T., Peck, L.E., Johansen, M., Tarver, J.E., King, B.L., Newcomb, J.M., Sempere, L.F., Flatmark, K., Hovig, E.*, et al.* (2015). A Uniform System for the Annotation of Vertebrate microRNA Genes and the Evolution of the Human microRNAome. Annu Rev Genet *49*, 213-242.

Ghildiyal, M., Xu, J., Seitz, H., Weng, Z., and Zamore, P.D. (2010). Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway. RNA *16*, 43-56.

Gregory, R.I., Chendrimada, T.P., Cooch, N., and Shiekhattar, R. (2005). Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. Cell *123*, 631-640.

Grosswendt, S., Filipchyk, A., Manzano, M., Klironomos, F., Schilling, M., Herzog, M., Gottwein, E., and Rajewsky, N. (2014). Unambiguous identification of miRNA:target site interactions by different types of ligation reactions. Mol Cell *54*, 1042-1054.

Guo, S., and Kemphues, K.J. (1995). par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. Cell *81*, 611-620.

Guo, Y., Liu, J., Elfenbein, S.J., Ma, Y., Zhong, M., Qiu, C., Ding, Y., and Lu, J. (2015). Characterization of the mammalian miRNA turnover landscape. Nucleic Acids Res *43*, 2326-2341.

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

Haas, G., Cetin, S., Messmer, M., Chane-Woon-Ming, B., Terenzi, O., Chicher, J., Kuhn, L., Hammann, P., and Pfeffer, S. (2016). Identification of factors involved in target RNA-directed microRNA degradation. Nucleic Acids Res *44*, 2873-2887.

Hagan, J.P., Piskounova, E., and Gregory, R.I. (2009). Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. Nat Struct Mol Biol *16*, 1021-1025.

Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science *286*, 950-952.

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

Han, B.W., Hung, J.H., Weng, Z., Zamore, P.D., and Ameres, S.L. (2011). The 3'-to-5' exoribonuclease Nibbler shapes the 3' ends of microRNAs bound to Drosophila Argonaute1. Curr Biol *21*, 1878-1887.

Han, J., Lee, Y., Yeom, K.H., Nam, J.W., Heo, I., Rhee, J.K., Sohn, S.Y., Cho, Y., Zhang, B.T., and Kim, V.N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell *125*, 887-901.

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

Heo, I., Ha, M., Lim, J., Yoon, M.J., Park, J.E., Kwon, S.C., Chang, H., and Kim, V.N. (2012). Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. Cell *151*, 521-532.

Heo, I., Joo, C., Kim, Y.K., Ha, M., Yoon, M.J., Cho, J., Yeom, K.H., Han, J., and Kim, V.N. (2009). TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. Cell *138*, 696-708.

Hubstenberger, A., Courel, M., Benard, M., Souquere, S., Ernoult-Lange, M., Chouaib, R., Yi, Z., Morlot, J.B., Munier, A., Fradet, M.*, et al.* (2017). P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. Mol Cell *68*, 144-157 e145.

Hundley, H.A., and Bass, B.L. (2010). ADAR editing in double-stranded UTRs and other noncoding RNA sequences. Trends Biochem Sci *35*, 377-383.

Huntzinger, E., and Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat Rev Genet *12*, 99-110.

Ishikawa-Ankerhold, H.C., Ankerhold, R., and Drummen, G.P. (2012). Advanced fluorescence microscopy techniques--FRAP, FLIP, FLAP, FRET and FLIM. Molecules *17*, 4047-4132.

Iwasaki, S., Kobayashi, M., Yoda, M., Sakaguchi, Y., Katsuma, S., Suzuki, T., and Tomari, Y. (2010). Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. Mol Cell *39*, 292-299.

Jang, D.E., Lee, J.Y., Lee, J.H., Koo, O.J., Bae, H.S., Jung, M.H., Bae, J.H., Hwang, W.S., Chang, Y.J., Lee, Y.H.*, et al.* (2018). Multiple sgRNAs with overlapping sequences enhance CRISPR/Cas9-mediated knock-in efficiency. Exp Mol Med *50*, 16.

Jankele, R.S., P. (2015). Analysis of short Argonaute isoforms from mouse oocytes. In [Online] https://iscunicz/webapps/zzp/detail/144561 (Prague: Charles University in Prague).

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Jinek, M., and Doudna, J.A. (2009). A three-dimensional view of the molecular machinery of RNA interference. Nature *457*, 405-412.

Jo, B.S., and Choi, S.S. (2015). Introns: The Functional Benefits of Introns in Genomes. Genomics Inform *13*, 112-118.

Johnston, M., Geoffroy, M.C., Sobala, A., Hay, R., and Hutvagner, G. (2010). HSP90 protein stabilizes unloaded argonaute complexes and microscopic P-bodies in human cells. Mol Biol Cell *21*, 1462-1469.

Kamenska, A., Simpson, C., Vindry, C., Broomhead, H., Benard, M., Ernoult-Lange, M., Lee, B.P., Harries, L.W., Weil, D., and Standart, N. (2016). The DDX6-4E-T interaction mediates translational repression and P-body assembly. Nucleic Acids Res *44*, 6318-6334.

Kaneda, M., Tang, F., O'Carroll, D., Lao, K., and Surani, M.A. (2009). Essential role for Argonaute2 protein in mouse oogenesis. Epigenetics Chromatin *2*, 9.

Kawahara, Y., Megraw, M., Kreider, E., Iizasa, H., Valente, L., Hatzigeorgiou, A.G., and Nishikura, K. (2008). Frequency and fate of microRNA editing in human brain. Nucleic Acids Res *36*, 5270-5280.

Ketting, R.F., Haverkamp, T.H., van Luenen, H.G., and Plasterk, R.H. (1999). Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell *99*, 133-141.

Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell *115*, 209-216.

Klattenhoff, C., and Theurkauf, W. (2008). Biogenesis and germline functions of piRNAs. Development *135*, 3-9.

Krol, J., Busskamp, V., Markiewicz, I., Stadler, M.B., Ribi, S., Richter, J., Duebel, J., Bicker, S., Fehling, H.J., Schubeler, D.*, et al.* (2010). Characterizing light-regulated retinal microRNAs reveals rapid turnover as a common property of neuronal microRNAs. Cell *141*, 618-631.

Kwak, P.B., and Tomari, Y. (2012). The N domain of Argonaute drives duplex unwinding during RISC assembly. Nat Struct Mol Biol *19*, 145-151.

Lam, A.J., St-Pierre, F., Gong, Y., Marshall, J.D., Cranfill, P.J., Baird, M.A., McKeown, M.R., Wiedenmann, J., Davidson, M.W., Schnitzer, M.J.*, et al.* (2012). Improving FRET dynamic range with bright green and red fluorescent proteins. Nat Methods *9*, 1005-1012.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843-854.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S.*, et al.* (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415-419.

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

Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004b). Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. Cell *117*, 69-81.

Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15-20.

Li, Y., Lu, J., Han, Y., Fan, X., and Ding, S.W. (2013). RNA interference functions as an antiviral immunity mechanism in mammals. Science *342*, 231-234.

Liang, C., Wang, Y., Murota, Y., Liu, X., Smith, D., Siomi, M.C., and Liu, Q. (2015). TAF11 Assembles the RISC Loading Complex to Enhance RNAi Efficiency. Mol Cell *59*, 807-818.

Lin, Y., Cradick, T.J., Brown, M.T., Deshmukh, H., Ranjan, P., Sarode, N., Wile, B.M., Vertino, P.M., Stewart, F.J., and Bao, G. (2014). CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. Nucleic Acids Res *42*, 7473-7485.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437-1441.

Liu, J., Valencia-Sanchez, M.A., Hannon, G.J., and Parker, R. (2005). MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. Nat Cell Biol *7*, 719-723.

Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.E., Smith, D.P., and Wang, X. (2003). R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. Science *301*, 1921-1925.

Liu, Y., Ye, X., Jiang, F., Liang, C., Chen, D., Peng, J., Kinch, L.N., Grishin, N.V., and Liu, Q. (2009). C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. Science *325*, 750-753.

Luo, Y., Na, Z., and Slavoff, S.A. (2018). P-Bodies: Composition, Properties, and Functions. Biochemistry *57*, 2424-2431.

Lykke-Andersen, K., Gilchrist, M.J., Grabarek, J.B., Das, P., Miska, E., and Zernicka-Goetz, M. (2008). Maternal Argonaute 2 is essential for early mouse development at the maternal-zygotic transition. Mol Biol Cell *19*, 4383-4392.

Ma, H., Wu, Y., Choi, J.G., and Wu, H. (2013). Lower and upper stem-single-stranded RNA junctions together determine the Drosha cleavage site. Proc Natl Acad Sci U S A *110*, 20687-20692.

Ma, J., Flemr, M., Stein, P., Berninger, P., Malik, R., Zavolan, M., Svoboda, P., and Schultz, R.M. (2010). MicroRNA activity is suppressed in mouse oocytes. Curr Biol *20*, 265-270.

Maniataki, E., and Mourelatos, Z. (2005). A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. Genes Dev *19*, 2979-2990.

Mathys, H., Basquin, J., Ozgur, S., Czarnocki-Cieciura, M., Bonneau, F., Aartse, A., Dziembowski, A., Nowotny, M., Conti, E., and Filipowicz, W. (2014). Structural and biochemical insights to the role of the CCR4-NOT complex and DDX6 ATPase in microRNA repression. Mol Cell *54*, 751-765.

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

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell *15*, 185-197.

Meister, G., Landthaler, M., Peters, L., Chen, P.Y., Urlaub, H., Luhrmann, R., and Tuschl, T. (2005). Identification of novel argonaute-associated proteins. Curr Biol *15*, 2149-2155.

Melton, C., Judson, R.L., and Blelloch, R. (2010). Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. Nature *463*, 621-626.

Miura, H., Quadros, R.M., Gurumurthy, C.B., and Ohtsuka, M. (2018). Easi-CRISPR for creating knock-in and conditional knockout mouse models using long ssDNA donors. Nat Protoc *13*, 195-215.

Montgomery, M.K., Xu, S., and Fire, A. (1998). RNA as a target of double-stranded RNA-mediated genetic interference in Caenorhabditis elegans. Proc Natl Acad Sci U S A *95*, 15502-15507.

Moretti, F., Kaiser, C., Zdanowicz-Specht, A., and Hentze, M.W. (2012). PABP and the poly(A) tail augment microRNA repression by facilitated miRISC binding. Nat Struct Mol Biol *19*, 603-608.

Morgens, D.W., Wainberg, M., Boyle, E.A., Ursu, O., Araya, C.L., Tsui, C.K., Haney, M.S., Hess, G.T., Han, K., Jeng, E.E.*, et al.* (2017). Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. Nat Commun *8*, 15178.

Mullokandov, G., Baccarini, A., Ruzo, A., Jayaprakash, A.D., Tung, N., Israelow, B., Evans, M.J., Sachidanandam, R., and Brown, B.D. (2012). High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. Nat Methods *9*, 840-846.

Murchison, E.P., Stein, P., Xuan, Z., Pan, H., Zhang, M.Q., Schultz, R.M., and Hannon, G.J. (2007). Critical roles for Dicer in the female germline. Genes Dev *21*, 682-693.

Murphy, D., Dancis, B., and Brown, J.R. (2008). The evolution of core proteins involved in microRNA biogenesis. BMC Evol Biol *8*, 92.

Nakanishi, K., Ascano, M., Gogakos, T., Ishibe-Murakami, S., Serganov, A.A., Briskin, D., Morozov, P., Tuschl, T., and Patel, D.J. (2013). Eukaryote-specific insertion elements control human ARGONAUTE slicer activity. Cell Rep *3*, 1893-1900.

Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. Plant Cell *2*, 279-289.

Nejepinska, J., Flemr, M., Svoboda, P. (2012). The canonical RNA interference pathway in animals. In In: Mallick B, Ghosh Z (eds) Regulatory RNAs Springer, Berlin Heidelberg, pp. 111-149.

Nguyen, T.A., Jo, M.H., Choi, Y.G., Park, J., Kwon, S.C., Hohng, S., Kim, V.N., and Woo, J.S. (2015). Functional Anatomy of the Human Microprocessor. Cell *161*, 1374-1387.

Noland, C.L., Ma, E., and Doudna, J.A. (2011). siRNA repositioning for guide strand selection by human Dicer complexes. Mol Cell *43*, 110-121.

Obbard, D.J., Gordon, K.H., Buck, A.H., and Jiggins, F.M. (2009). The evolution of RNAi as a defence against viruses and transposable elements. Philos Trans R Soc Lond B Biol Sci *364*, 99-115.

Okamura, K., Liu, N., and Lai, E.C. (2009). Distinct mechanisms for microRNA strand selection by Drosophila Argonautes. Mol Cell *36*, 431-444.

Pal-Bhadra, M., Leibovitch, B.A., Gandhi, S.G., Chikka, M.R., Bhadra, U., Birchler, J.A., and Elgin, S.C. (2004). Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. Science *303*, 669-672.

Park, M.S., Phan, H.D., Busch, F., Hinckley, S.H., Brackbill, J.A., Wysocki, V.H., and Nakanishi, K. (2017). Human Argonaute3 has slicer activity. Nucleic Acids Res *45*, 11867-11877.

Patrick, D.M., Zhang, C.C., Tao, Y., Yao, H., Qi, X., Schwartz, R.J., Jun-Shen Huang, L., and Olson, E.N. (2010). Defective erythroid differentiation in miR-451 mutant mice mediated by 14-3-3zeta. Genes Dev *24*, 1614-1619.

Peng, R., Lin, G., and Li, J. (2016). Potential pitfalls of CRISPR/Cas9-mediated genome editing. FEBS J *283*, 1218-1231.

Pfeffer, S., Sewer, A., Lagos-Quintana, M., Sheridan, R., Sander, C., Grasser, F.A., van Dyk, L.F., Ho, C.K., Shuman, S., Chien, M.*, et al.* (2005). Identification of microRNAs of the herpesvirus family. Nat Methods *2*, 269-276.

Pillai, R.S., Artus, C.G., and Filipowicz, W. (2004). Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. RNA *10*, 1518-1525.

Quadros, R.M., Miura, H., Harms, D.W., Akatsuka, H., Sato, T., Aida, T., Redder, R., Richardson, G.P., Inagaki, Y., Sakai, D.*, et al.* (2017). Easi-CRISPR: a robust method for one-step generation of mice carrying conditional and insertion alleles using long ssDNA donors and CRISPR ribonucleoproteins. Genome Biol *18*, 92.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat Protoc *8*, 2281-2308.

Rand, T.A., Ginalski, K., Grishin, N.V., and Wang, X. (2004). Biochemical identification of Argonaute 2 as the sole protein required for RNA-induced silencing complex activity. Proc Natl Acad Sci U S A *101*, 14385-14389.

Rissland, O.S., Hong, S.J., and Bartel, D.P. (2011). MicroRNA destabilization enables dynamic regulation of the miR-16 family in response to cell-cycle changes. Mol Cell *43*, 993-1004.

Ruby, J.G., Jan, C.H., and Bartel, D.P. (2007a). Intronic microRNA precursors that bypass Drosha processing. Nature *448*, 83-86.

Ruby, J.G., Stark, A., Johnston, W.K., Kellis, M., Bartel, D.P., and Lai, E.C. (2007b). Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. Genome Res *17*, 1850-1864.

Schirle, N.T., and MacRae, I.J. (2012). The crystal structure of human Argonaute2. Science *336*, 1037-1040.

Schirle, N.T., Sheu-Gruttadauria, J., and MacRae, I.J. (2014). Structural basis for microRNA targeting. Science *346*, 608-613.

Schmitter, D., Filkowski, J., Sewer, A., Pillai, R.S., Oakeley, E.J., Zavolan, M., Svoboda, P., and Filipowicz, W. (2006). Effects of Dicer and Argonaute down-regulation on mRNA levels in human HEK293 cells. Nucleic Acids Res *34*, 4801-4815.

Schurmann, N., Trabuco, L.G., Bender, C., Russell, R.B., and Grimm, D. (2013). Molecular dissection of human Argonaute proteins by DNA shuffling. Nat Struct Mol Biol *20*, 818-826.

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell *115*, 199-208.

Schwille, P., Meyer-Almes, F.J., and Rigler, R. (1997). Dual-color fluorescence cross-correlation spectroscopy for multicomponent diffusional analysis in solution. Biophys J *72*, 1878-1886.

Sekar, R.B., and Periasamy, A. (2003). Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. J Cell Biol *160*, 629-633.

Shenoy, A., and Blelloch, R.H. (2014). Regulation of microRNA function in somatic stem cell proliferation and differentiation. Nat Rev Mol Cell Biol *15*, 565-576.

Stark, A., Brennecke, J., Bushati, N., Russell, R.B., and Cohen, S.M. (2005). Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. Cell *123*, 1133-1146.

Stein, P., Rozhkov, N.V., Li, F., Cardenas, F.L., Davydenko, O., Vandivier, L.E., Gregory, B.D., Hannon, G.J., and Schultz, R.M. (2015). Essential Role for endogenous siRNAs during meiosis in mouse oocytes. PLoS Genet *11*, e1005013.

Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J., and Mateo, J.L. (2015). CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. PLoS One *10*, e0124633.

Suh, N., Baehner, L., Moltzahn, F., Melton, C., Shenoy, A., Chen, J., and Blelloch, R. (2010). MicroRNA function is globally suppressed in mouse oocytes and early embryos. Curr Biol *20*, 271-277.

Svoboda, P. (2010). Why mouse oocytes and early embryos ignore miRNAs? RNA Biol *7*, 559-563.

Svoboda, P. (2017). Long and small noncoding RNAs during oocyte-to-embryo transition in mammals. Biochem Soc Trans *45*, 1117-1124.

Svobodova, E., Kubikova, J., and Svoboda, P. (2016). Production of small RNAs by mammalian Dicer. Pflugers Arch *468*, 1089-1102.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C. (1999). The rde-1 gene, RNA interference, and transposon silencing in C. elegans. Cell *99*, 123-132.

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

Tang, F., Kaneda, M., O'Carroll, D., Hajkova, P., Barton, S.C., Sun, Y.A., Lee, C., Tarakhovsky, A., Lao, K., and Surani, M.A. (2007). Maternal microRNAs are essential for mouse zygotic development. Genes Dev *21*, 644-648.

Tants, J.N., Fesser, S., Kern, T., Stehle, R., Geerlof, A., Wunderlich, C., Juen, M., Hartlmuller, C., Bottcher, R., Kunzelmann, S.*, et al.* (2017). Molecular basis for asymmetry sensing of siRNAs by the Drosophila Loqs-PD/Dcr-2 complex in RNA interference. Nucleic Acids Res *45*, 12536-12550.

Tomari, Y., Matranga, C., Haley, B., Martinez, N., and Zamore, P.D. (2004). A protein sensor for siRNA asymmetry. Science *306*, 1377-1380.

Tsuboyama, K., Tadakuma, H., and Tomari, Y. (2018). Conformational Activation of Argonaute by Distinct yet Coordinated Actions of the Hsp70 and Hsp90 Chaperone Systems. Mol Cell *70*, 722-729 e724.

Ustianenko, D., Hrossova, D., Potesil, D., Chalupnikova, K., Hrazdilova, K., Pachernik, J., Cetkovska, K., Uldrijan, S., Zdrahal, Z., and Vanacova, S. (2013). Mammalian DIS3L2 exoribonuclease targets the uridylated precursors of let-7 miRNAs. RNA *19*, 1632-1638.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833-1837.

Waldo, G.S., Standish, B.M., Berendzen, J., and Terwilliger, T.C. (1999). Rapid protein-folding assay using green fluorescent protein. Nat Biotechnol *17*, 691-695.

Wang, B., Li, K., Wang, A., Reiser, M., Saunders, T., Lockey, R.F., and Wang, J.W. (2015). Highly efficient CRISPR/HDR-mediated knock-in for mouse embryonic stem cells and zygotes. Biotechniques *59*, 201-202, 204, 206-208.

Wang, Q., and Carmichael, G.G. (2004). Effects of length and location on the cellular response to double-stranded RNA. Microbiol Mol Biol Rev *68*, 432-452, table of contents.

Wang, Y., Sheng, G., Juranek, S., Tuschl, T., and Patel, D.J. (2008). Structure of the guide-strand-containing argonaute silencing complex. Nature *456*, 209-213.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T*., et al.* (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature *453*, 539-543.

Wen, J., Ladewig, E., Shenker, S., Mohammed, J., and Lai, E.C. (2015). Analysis of Nearly One Thousand Mammalian Mirtrons Reveals Novel Features of Dicer Substrates. PLoS Comput Biol *11*, e1004441.

Westholm, J.O., Ladewig, E., Okamura, K., Robine, N., and Lai, E.C. (2012). Common and distinct patterns of terminal modifications to mirtrons and canonical microRNAs. RNA *18*, 177-192.

Wilson, R.C., and Doudna, J.A. (2013). Molecular mechanisms of RNA interference. Annu Rev Biophys *42*, 217-239.

Xie, M., Li, M., Vilborg, A., Lee, N., Shu, M.D., Yartseva, V., Sestan, N., and Steitz, J.A. (2013). Mammalian 5'-capped microRNA precursors that generate a single microRNA. Cell *155*, 1568-1580.

Yan, K.S., Yan, S., Farooq, A., Han, A., Zeng, L., and Zhou, M.M. (2003). Structure and conserved RNA binding of the PAZ domain. Nature *426*, 468-474.

Yang, J.S., Maurin, T., Robine, N., Rasmussen, K.D., Jeffrey, K.L., Chandwani, R., Papapetrou, E.P., Sadelain, M., O'Carroll, D., and Lai, E.C. (2010). Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. Proc Natl Acad Sci U S A *107*, 15163-15168.

Yang, Y., Bai, W., Zhang, L., Yin, G., Wang, X., Wang, J., Zhao, H., Han, Y., and Yao, Y.Q. (2008). Determination of microRNAs in mouse preimplantation embryos by microarray. Dev Dyn *237*, 2315-2327.

Yang, Z., Jakymiw, A., Wood, M.R., Eystathioy, T., Rubin, R.L., Fritzler, M.J., and Chan, E.K. (2004). GW182 is critical for the stability of GW bodies expressed during the cell cycle and cell proliferation. J Cell Sci *117*, 5567-5578.

Ye, X., Huang, N., Liu, Y., Paroo, Z., Huerta, C., Li, P., Chen, S., Liu, Q., and Zhang, H. (2011). Structure of C3PO and mechanism of human RISC activation. Nat Struct Mol Biol *18*, 650-657.

Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev *17*, 3011-3016.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R.*, et al.* (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917-1920.

Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell *101*, 25-33.

Zekri, L., Huntzinger, E., Heimstadt, S., and Izaurralde, E. (2009). The silencing domain of GW182 interacts with PABPC1 to promote translational repression and degradation of microRNA targets and is required for target release. Mol Cell Biol *29*, 6220-6231.

Zeng, Y., and Cullen, B.R. (2005). Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. J Biol Chem *280*, 27595-27603.

Zeng, Y., Yi, R., and Cullen, B.R. (2005). Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. EMBO J *24*, 138-148.

Supplementary information

Primer list

Lower case letters show restriction enzyme sites, which are not complementary to the template DNA and are used for introducing the sites into the amplicons.

HA-mRuby2-linker-mAgo2 primers

mAgo2_5-arm_out_Fwd	GATGCCTGGCAATTTACCTACTTC
mAgo2_5-arm_out_Rev	CGTCCGTGTTTTTAAGTTTTCTGG
mAgo2_5-arm_out_2_Fwd	AGCTGTTTTTAGGCATCTCTGAGC
mAgo2_5-arm_out_2_Rev	GGAAATGCGTCCGTGTTTTTAAGT
mAgo2_5Arm_out_3_Fwd	TATTGGTACCGCAAGGGCTAGGTA
mAgo2_5Arm_out_3_Rev	CACAGCTATGCGCTCGGGAA

mAgo2_5Arm_In_Xbal_Fwd	atctagaGCTCCACTGCTGAGTGAAGTCTAA
mAgo2_5Arm_in_ApaLI_Rev	AgtgcacAGTCGCCGGCCGCTCGATCCgcc
mAgo2_5Arm_In_Xbal_2_Fwd	atctagaTGGCTCAGGGCCAAGCACACCGCTGC
mAgo2_5Arm_in_XhoI_Fwd2	ctcgagTGCTGTGTCTAGTTTCCTCCTCCT
mAgo2_5Arm_in_SacI_Rev2	gagctcAGTCGCCGGCCGCTCGATCC
mAgo2_5Arm_in_XhoI_3_Fwd	ctcgagTATTGGTACCGCAAGGGCTAGGTA
mAgo2_3Arm_out_Fwd	AGGATTATCTAGGTTGTGGAGTCG
mAgo2_3Arm_out_Rev	TCTTACTTCCACCCTAGCAAAGAG
mAgo2_3-arm_out_2_Fwd	AATAACTTAAAGAGCCAGGCCGA
mAgo2_3-arm_out_2_Rev	GATCCAATGGTCTTCCATTCTCCT
mAgo2_3Arm_out_3_Fwd	GTCGCTTCTGGGGTCCCTCATTC
mAgo2_3Arm_out_3_Rev	CCCGGATGTTAAACCTCTCTGAGC
mAgo2_3Arm_out_4_Fwd	GGGGTCCCTCATTCAGGGCAAG
mAgo2_3Arm_out_4_Rev	CTCCCCTTGCAGAACACACAGTA
mAgo2_3Arm_In_BamHI_Fwd	AggatccTACTCGGGAGCCGGCCCCGGTGAG
mAgo2_3Arm_In_EagI_Rev	AcggccgCGTCCCCGAGGAAGTCACTAGAAA
mAgo2_3Arm_In_EagI_2_Rev	AcggccgCAGATCCAATGGTCTTCCATTCTCCT
mAgo2_3Arm_In_XbaI_3_Rev	tctagaGCTAAGATTTGCTGGGTTCCTCCA
mAgo2_3Arm_In_XbaI_4_Rev	tctagaAGCGTCTTCAAGAAAGACCCACTG
HA-mRuby2_ApaLI_Fwd	TgtgcacCCTCGCGGCTCGCGGCGTCGTGGCCCGTGGAGCGCGCGGCCGGCCCTCGGCCGCGGCCCTCGGCAACGCCACCATGTACCCATACGATGTTCCAGATTACGCTGTGTCTAAGGGCGAAGAGCTG
mRuby2_SacI_Fwd	gagctcCCTCGCGGCTCGCGGCGTCGTGGCCCGTGGAGCGCGCGGCCGGCCCTCGGCCGCGGCCCTCGGCAACGCCACCATGTACCCATACGATGTTCCAGATTACGCTGTGTCTAAGGGCGAAGAGCTG
mRuby2_BamHI_linker_Rev	aGGATCCAAACTCGCCAGAACCGGCAGCAGAACCAGCAGACCCCTTGTACAGCTCGTCCATCCCACC
mRuby2_BamHI_rev	ggatccCTTGTACAGCTCGTCCATCCCACC

Screening primers

Ago2_5screen_Fwd	ACAAGTGAAGAAGGGACATACGTG
Ago2_5screen_Rev	CTCGGCCTGGCTCTTTAAGTTATT
Ago2_3screen_Fwd	GGTTGAGAAAACAAGTGTTGGGGA
Ago2_3screen_Rev	TCTTCAAATCGTTGGGAAGCCTTT
mRuby2_screen_Fwd	CCAATTCAAATGCACAGGTGAAGG
mRuby2_screen_Rev	GGAAAGTTTACCCCTCTGACTTGG
mAgo2_2_Fwd	ATAACTTAAAGAGCCAGGCCGAGG
mRuby2_Rev	CATGGTTTGAGTTCCCATGTACGG

HA-mRuby2-(linker)-mAgo2 expression vector (Ago2 coding sequence)

HindIII_HA_mRuby2_Fwd

aagcttATGTACCCATACGATGTTCCAGATTACGCTGTGTCTAAGGGCGAAGA GCTGATC

mRuby2_BamHI_linker_Rev

aGGATCCAAACTCGCCAGAACCGGCAGCAGAACCAGCAGACCCCTTGTACA GCTCGTCCATCCCACC

mRuby2_BamHI_rev	ggatccCTTGTACAGCTCGTCCATCCCACC
BamHI_mAgo2_CDS_Fwd	ggatccTACTCGGGAGCCGGCCCCGTTCTT
BamHI_mAgo2_CDS_Fwd2	ggatccTACTCGGGAGCCGGCCCCGT
NotI_mAgi2_CDS_Rev	gcggccgcTCAAGCAAAGTACATGGTGCGCAGTGTGtcc
NotI_mAgi2_CDS_Rev2-	gcggccgcTCAAGCAAAGTACATGGTGCGCAGTGTG

Screening primers

mAgo2_screen_CDS_fwd	CTGAGAAATGCCCTCGGAGAGTGA
mAgo2_screen_CDS_rev	CAGTGAAGAAGGAACGGCCAACAG

HA-mRuby2-Ago2_E1i1

Only the 3' Arm was amplified new, rest of the template was used from the previous version of the homologous template.

mAgo2_3Arm_In_BamHI_Fwd	AggatccTACTCGGGAGCCGGCCCCGGTGAG
Ago2_i1_5'_Rev_NheI	gctagcCCCCACAGCTATGCGCTCGGGAAC
Ago2_i1_3'_FWD_NheI	gctagcCGCCCTCCCCATTCAAGTGCTAAT
Ago2_i1_3'_Rev_NcoI	ccatggAAAGCTCCCCTTGCAGAACACACA
Ago2_i1_3'_FWD_NheI	gctagcCCCATTCAAGTGCTAATCGCCCTG
Ago2_i1_3'_Rev_NcoI	ccatggGGAGTGCTCAAACAGGTTGCCCTA

TNRC6C-Clover-3xFLAG

C-TNRC6C_5_out_fwd	ACTGTGGAACCCAGAATGGGACTT
C-TNRC6C_5_out_rev	GACTGGACAGAGGTACTCCAAGGG
C-TNRC6C_NotI_5_fwd	gcggccgcCCAGAACATAGCAAGGGCCTGGAG
C-TNRC6C_BamHI_5_rev	ggatccGATGGACTCGCCGCTGAGCAGGTC
C-TNRC6C_3_out_fwd	CACGCCTCTCAACACTCTGCTG
C-TNRC6C_3_out_rev	TGGGAAGTGGTGCAATACTCTGGT
C-TNRC6C_NdeI_3_fwd	catatgTAGGGGCCACGGTCACTGGGGGAA
C-TNRC6C_HindIII_3_rev	aagcttGAATTGTCACCAGCCCTGTCCCGA

Clover_BamHI_linker_Fwd
ggatccGGGTCTGCTGGTTCTGCTGCCGGTTCTGGCGAGTTTGTGAGCAAGG
GCGAGGAGCTGTTC
Clover_BamHI_fwd ggatccGTGAGCAAGGGCGAGGAGCTGTTC
Clover_NdeI_3xFLAG_rev
catatgCTTGTCGTCGTCGTCCTTGTAGTCGATGTCGTGGTCCTTGTAGTCACC
GTCGTGGTCCTTGTAGTCCTTGTACAGCTCGTCCATGCCATG

Screening primers

C-TNRC6C_screen_5_fwd	CAGTGTTAGGAGGGGATGGGGAAG

C-TNRC6C_screen_5_rev	GAGATCTGACGCCCTCTTCTGGAG
C-TNRC6C_screen_3_fwd	TTCAGATTGTGGGCTGGAAAGGGT
C-TNRC6C_screen_3_rev	CGGATGTCAACGTGGAAGTGTTCA

TNRC6C-Clover-3xFLAG_e21e22

TNRC6C_i19_out_Fwd	GCCTGGGAGCATCTTATACTGGCA
TNRC6C_i21_out_Rev	GTAGCACTTCCCCATCCCCTCCTA
TNRC6C_i19_Fwd_NotI	gcggccgcCCTTGCCCGATGCACTCACCTTAT
TNRC6C_i19_Fwd2_NotI	gcggccgcGTGTGTGGATTCCGAGTGAAGGCT
TNRC6C_E21_Rev_PciI	ACATGTGCAGAGACTTCTGGGCCT
TNRC6C_E22_FWD_PciI	acatGTGTGTACTGGGAAACACCACCATCTT
C-TNRC6C_BamHI_5_rev	ggatccGATGGACTCGCCGCTGAGCAGGTC

Screening primers

TNRC6C_i19_screen_Fwd	CCCTCTCAGCTCTATGGTCAGCCT
TNRC6C_E21_screen_Rev	GCCTTGGCGGCTTCTTCCTTAGAA

Plasmid screening primers

pJet1.2_Fwd_seq	CGACTCACTATAGGGAGAGCGGC
pJet1.2_Rev2_seq	CGGTTCCTGATGAGGTGGTTAGC
pPuro_qPCR_Rev	CATGGCCGAGTTGAGCGGTTC
pPuro_Fwd_seq	CCAAACTGGAACAACACTCAACCC

sgRNA oligos and reporters

First line of sgRNA called "Sequence" shows the recognized (targeted) sequence, "Oligo pair" the sequence of oligos with overhangs for ligation that were ordered.

HA-mRuby2-Ago2

mAgo2_sgRNA_1

Sequence: GGCGGATCGAGCGGCCGG

Oligo pair:	mAgo2_sgRNA_1_sense	CACCgGGCGGATCGAGCGGCCGG
	mAgo2_sgRNA_1_antisense	AAACCCGGCCGCTCGATCCGCCC

mAgo2_sgRNA_2

Sequence:		GCAACGCCACCATGTACT
Oligo pair:	mAgo2_sgRNA_2_sense	CACCgGCAACGCCACCATGTACT
	mAgo2_sgRNA_2_antisense	AAACAGTACATGGTGGCGTTGCC

Reporter:

mAgo2_reporter_sense

cGGCGGATCGAGCGGCCGGCGGGCAACGCCACCATGTACTCGGgacgt

mAgo2_reporter_antisense

cCCGAGTACATGGTGGCGTTGCCCGCCGGCCGCTCGATCCGCCgagct

HA-mRuby2-mAgo2-E1i1

Ago2_i1_sg1

Sequence:		CAGCACGGCGGGCCTGGGGC
Oligo pair:	Ago2_i1_sg1_T	CACCgCAGCACGGCGGGCCTGGGGC
	Ago2_i1_sg2_B	AAACGCCCCAGGCCCGCCGTGCTGC

Ago2_i1_sg2

Sequence:		TGCGTCCAAGCGACCGATCG
Oligo pair:	Ago2_i1_sg2_T	CACCgTGCGTCCAAGCGACCGATCG
	Ago2_i1_sg2_B	AAACCGATCGGTCGCTTGGACGCAC

Ago2_i1_sg3

Sequence:		GCTTGGACGCAGTTTCACGC
Oligo pair:	Ago2_i1_sg3_T	CACCGCTTGGACGCAGTTTCACGC
	Ago2_i1_sg3_B	AAACGCGTGAAACTGCGTCCAAGC

Reporter

Ago2_i1_rep_sense

cCAGCACGGCGGGCCTGGGGCGGGTGCGTCCAAGCGACCGATCGCGGGCT

TGGACGCAGTTTCACGCCGGgacgt

Ago2_i1_rep_antisense

cCCGGCGTGAAACTGCGTCCAAGCCCGCGATCGGTCGCTTGGACGCACCCG

CCCCAGGCCCGCCGTGCTGgagct

TNRC6C-Clover-3xFLAG

C-TNRC6C_sg1

Sequence:		AGCGGCGAGTCCATCTAG
Oligo pair	TNRC6C_sg1_fwd	CACCgAGCGGCGAGTCCATCTAG
	TNRC6C_sg1_rev	AAACCTAGATGGACTCGCCGCTc

C-TNRC6C_sg2

Sequence:		TCAGCGGCGAGTCCATCT
Oligo pair	TNRC6C_sg2_fwd	CACCgTCAGCGGCGAGTCCATCT
	TNRC6C_sg2_rev	AAAc AGATGGACTCGCCGCTGAc

Reporter

TNRC6C_reporter_sense

cAGCGGCGAGTCCATCTAGGGGTCAGCGGCGAGTCCATCTAGGgacgt

TNRC6C_reporter_antisense

cCCTAGATGGACTCGCCGCTGACCCCTAGATGGACTCGCCGCTgagct

TNRC6C-Clover-3xFLAG_e21e22

Sequence:		CATCAGGGCACTTGCTACCTGGG
Oligo pair	TNRC6C_i20_sg1_ fwd	CACCgCATCAGGGCACTTGCTACCT
	TNRC6C_i20_sg1_ rev	AAACAGGTAGCAAGTGCCCTGATGc
Sequence:		GGTGCCCAACAGTGTTAGGAGGG
Oligo pair	TNRC6C_i20_sg2_ fwd	CACCGGTGCCCAACAGTGTTAGGA
	TNRC6C_i20_sg2_ rev	AAACTCCTAACACTGTTGGGCACc
Sequence:		AGGTGCCCAACAGTGTTAGGAGG
Oligo pair	TNRC6C_i20_sg3_fwd	CACCgAGGTGCCCAACAGTGTTAGG
	TNRC6C_i20_sg3_rev	AAACCCTAACACTGTTGGGCACCTc
Sequence:		TAAAAGGATCGCCACGTCCCAGG
Oligo pair	TNRC6C_i20_sg4_ fwd	CACCgTAAAAGGATCGCCACGTCCC
	TNRC6C_i20_sg4_ rev	AAACGGGACGTGGCGATCCTTTTAc

Sequence:		GTAGTACAGACTCAGGACGAAGG
Oligo pair	TNRC6C_i20_sg5_ fwd	CACCGTAGTACAGACTCAGGACGA
	TNRC6C_i20_sg5_ rev	AAACTCGTCCTGAGTCTGTACTAc
Sequence:		CTCCACACATGAGTCTTCTGTGG
Oligo pair	TNRC6C_i20_sg6_ fwd	CACCgCTCCACACATGAGTCTTCTG
	TNRC6C_i20_sg6_ rev	AAACCAGAAGACTCATGTGTGGAGc

Reporter

The reporter does not contain all of the sgRNA targeting sequences, in the interest of length of the reporter. For the production of puromycin resistance is sufficient one cleavage of the reporter sequence.

TNRC6C-e21e22-reporter_sense

cCATCAGGGCACTTGCTACCTGGGGGTGCCCAACAGTGTTAGGAGGGAGG
TGCCCAACAGTGTTAGGAGGTAAAAGGATCGCCACGTCCCAGGgacgt

TNRC6C-e21e22_reporter_antisense

cGACGTGGCGATCCTTTTACCTCCTAACACTGTTGGGCACCTCCCTCCTAACACTGT
TGGGCACCCCCAGGTAGCAAGTGCCCTGATGgagct

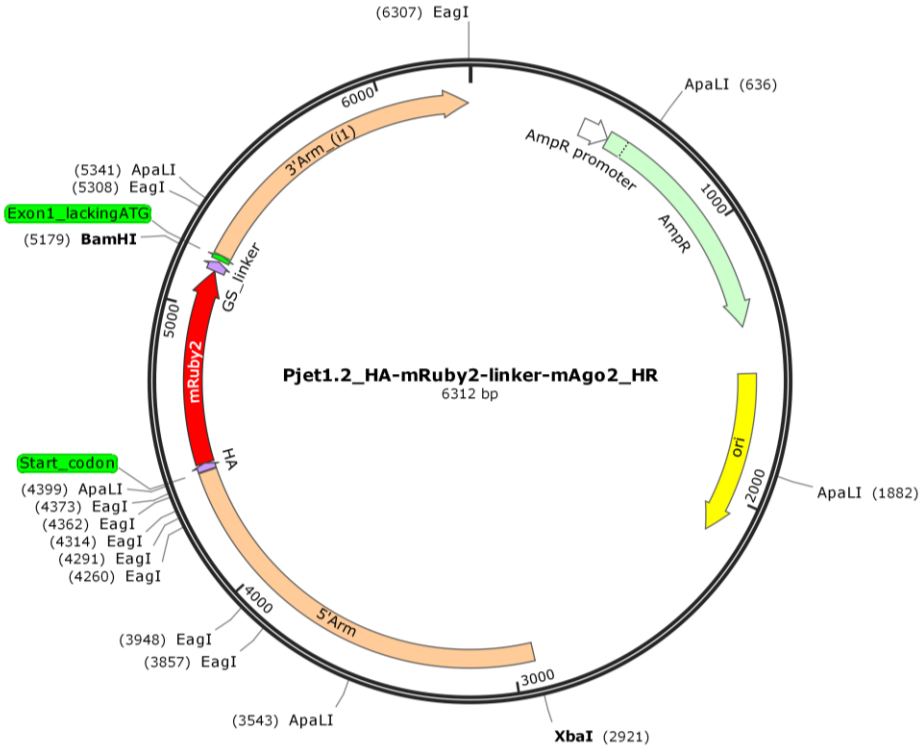
GS flexible linker sequence

GS flexible linker

DNA	5’-GGGTCTGCTGGTTCTGCTGCCGGTTCTGGCGAGTTT-3’
Peptide	N-GSAGSAAGSGEF-C

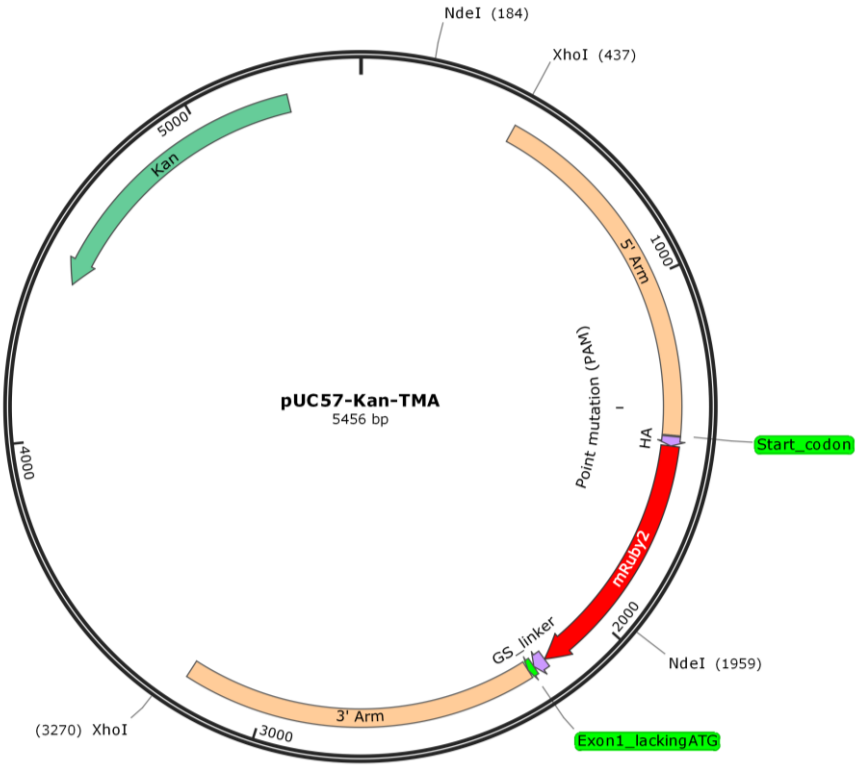
Plasmid maps

pJet1.2_HA-mRuby2-linker-mAgo2



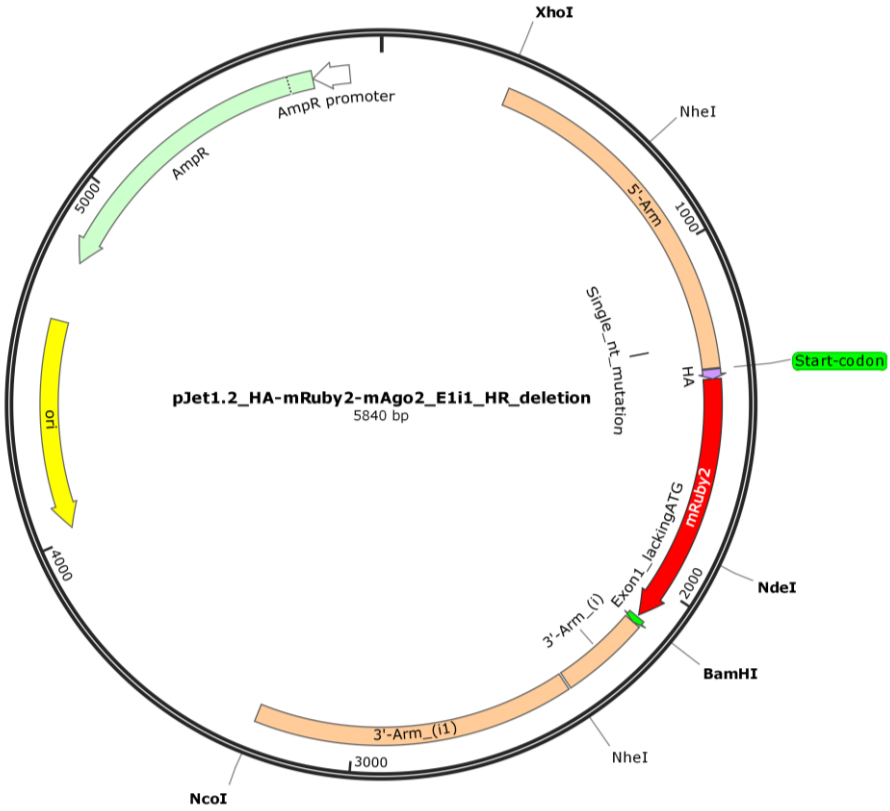
This plasmid was not finished, because of the problematic amplification of 5' arm described earlier.

pUC57_HA-mRuby2-linker-mAgo2 (synthetized)

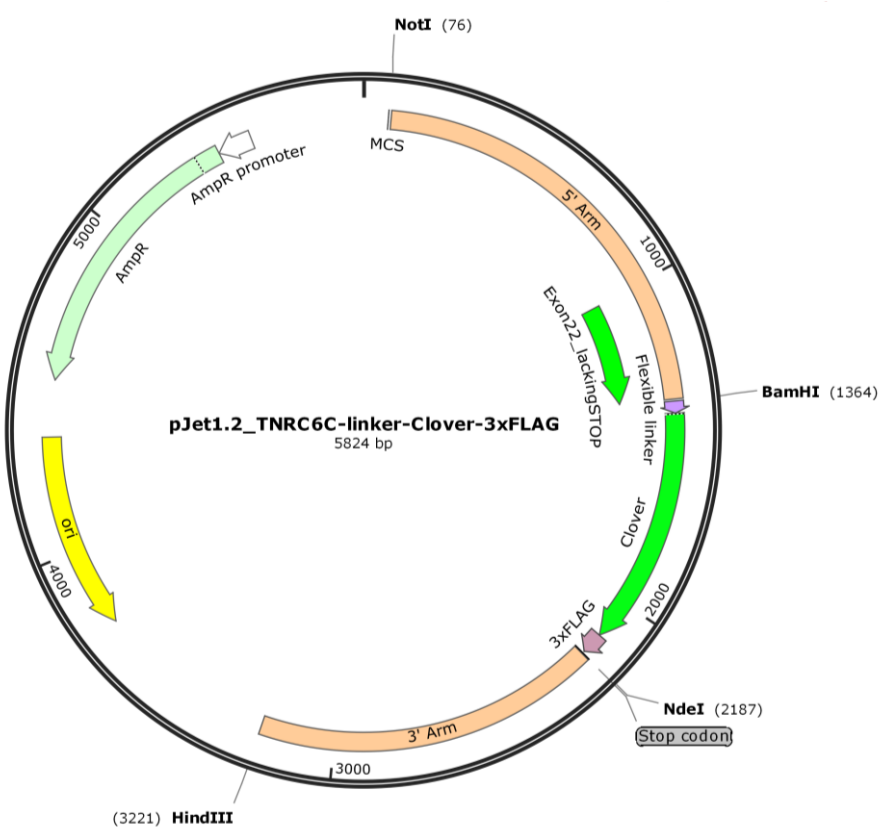


The version of this plasmid without the linker is not shown, it was prepared by amplifying the 3' Arm (including the exon1) by PCR and ligated into a pJet1.2 vector as described earlier. The 5' Arm together with half of the HA-mRuby2 sequence was restricted from the pUC57 synthetized plasmid (XhoI + NdeI), second part of the mRuby2 (from NdeI to the linker) was restricted from a pJet1.2_HA-mRuby2 plasmid. The three obtained fragments were ligated together in pJet1.2 plasmid forming a complete homologous recombination template similar to the one showed here, only without the linker.

pJet1.2_HA-mRuby2-mAgo2_E1i1

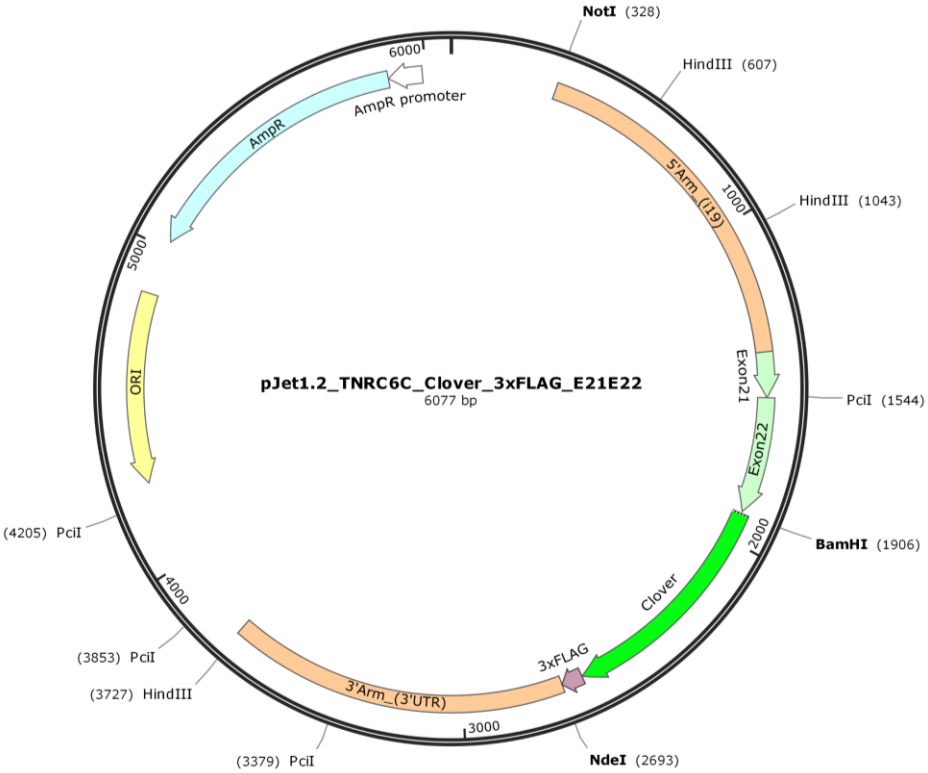


pJet1.2_TNRC6C-linker-Clover-3xFLAG



Similar plasmid without the linker sequence was prepared by excision of the linker-Clover-3xFLAG sequence from the pJet1.2_TNRC6C-linker-Clover-3xFLAG by restriction reaction (BamHI + NdeI) and subsequent ligation of the linearized plasmid with Clover-3xFLAG fragment restricted with BamHI + NdeI. The Clover-3xFLAG was previously prepared by PCR as described in the section Materials and Methods.

pJet1.2_TNRC6C-Clover_3xFLAG_e21e22



pCDNA3.1(+)_HA-mRuby2-linker-mAgo2_CDS (expression vector)