Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

The nuclear export of siRNA precursors via the dsRBD protein Blanks in *Drosophila melanogaster*

Volker Fabian Nitschko

aus

Nürtingen, Deutschland

2020

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

The nuclear export of siRNA precursors via the dsRBD protein Blanks in *Drosophila melanogaster*

Volker Fabian Nitschko

aus

Nürtingen, Deutschland

2020

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Klaus Förstemann betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 14.9.2020

Volker Nitschko

.....

Dissertation eingereicht am	29.9.2020
1. Gutachter:	Prof. Dr. Klaus Förstemann
2. Gutachter:	PD Dr. Dietmar Martin
Mündliche Prüfung am	29.10.2020

Abstract

The RNA interference (RNAi) pathway and small RNAs (sRNAs) in general have emerged as important regulators in diverse and essential cellular processes over the last decades. Among these processes are fundamental pathways such as genome surveillance, gene regulation and virus defence in higher eukaryotes.

The siRNA biogenesis pathway has been extensively studied and most of it is thereby well understood. Biogenesis starts with a long double-stranded precursor which is processed into the 21 nt long product by Dicer proteins. These shorter double-strands are then loaded into Argonaute proteins in which one strand is degraded while the other one becomes active and guides the Argonaute to its target for gene regulation. The long double-stranded precursors are either originating from the cytoplasm for example in virus defence or are produced in the nucleus in case of transposon defence and others. It is so far not completely clear if the precursors that are generated in the nucleus are processed directly in the nucleus by a Dicer protein or if they are exported to the cytoplasm before processing. In contrast to similar pathways like the miRNA biogenesis pathway, it is currently not known which protein (complex) facilitates the nuclear export of the double-stranded siRNA precursors.

The *Drosophila* nuclear protein Blanks has been shown to negatively affect spermatogenesis in mutants. Additionally, several studies found Blanks to promote RNAi. However, these studies could not propose a molecular mechanism with which Blanks could influence spermatogenesis or RNAi. In the experiments performed for this thesis we could now show that Blanks can bind double-stranded RNA via the second of its two double-stranded RNA binding domains (dsRBDs). Blanks had no influence on cytosolic RNAi. However, we observed that Blanks can shuttle between the cytoplasm and the nucleus. These observations and detected interactions with proteins from the nuclear export and import pathways led us to the hypothesis that Blanks is a novel factor involved in the nuclear export of dsRNAs. In deep sequencing experiments we were able to show that numerous genomic loci which are convergently transcribed produce siRNAs both *in vitro* and *in vivo*. The biogenesis of these mature siRNAs was dependent on Blanks. Additionally, we could also observe that alterations in Blanks localization and mutations in the dsRNA-binding domain negatively affected the production of these siRNAs. Due to these characteristics we named these endo-siRNAs “Blanks exported precursor siRNAs” or short “bepsiRNAs”. Since bepsiRNAs originate not only from exonic regions of genes but also from introns and 3’UTRs we propose that the precursors are formed from nascent transcripts of loci with convergent transcription.

Our *in vivo* studies also showed that a mutation in the dsRBD2 of Blanks alone does not lead to an infertility phenotype like a knock-out mutant. This observation is consistent with the results from previous studies that attributed the fertility defect to the first dsRBD. This means that Blanks' role in spermatogenesis and the nuclear export of dsRNAs are unrelated.

It has been shown in several species that siRNAs are also originating from DNA double-strand breaks (DSBs). The biogenesis and molecular function of these damage-induced siRNA (diRNAs) has been the subject of many studies over that last years and still remains elusive in most parts. A genome-wide screen performed in our lab showed that many *Drosophila* proteins influence the biogenesis of diRNAs. One group of these proteins were the components of the replication factor C (RFC) complex. The canonical function of the RFC complex is to load and unload PCNA during DNA replication and thereby to ensure genome stability. We set out to investigate a potential additional role of the RFC complex in the biogenesis of diRNAs via chromatin IP (ChIP). However, we were not able to get reproducible or strong signals for RFC complex recruitment to sites near an induced DSB via this method. We concluded that major adjustments to the experimental setup had to be made to investigate the potential role of the RFC complex in diRNA biogenesis further that were beyond the scope of this thesis.

Table of Contents

1 Small RNAs in *Drosophila*

1.1 RNAi and small RNA biogenesis in *Drosophila*

Thirty years ago Napoli, Lemieux, and Jorgensen (1990) described a reduction of expression levels for mRNAs transcribed from both the wildtype allele and a chimeric transgene of the gene CHS in petunia after the transgene was inserted. They named this phenomenon co-suppression but could not describe how this suppression was caused on a mechanistic level. After thirty years of research we know the described effect as RNA interference (RNAi). One major step in RNAi research was undoubtedly a *C. elegans* study showing that double-stranded RNA can silence gene expression very efficiently on the RNA level (Fire et al. (1998)). This discovery was rewarded with the Nobel prize in physiology or medicine in 2006. Today we know that RNAi is highly conserved in eukaryotes. We also have knowledge of a multitude of different small RNA species that are at the center of several distinct RNAi pathways. RNAi pathways do not only govern gene regulation but have been shown to play important roles in antiviral defence and genome defence e.g. by repression of transposable elements (TEs) (Ghildiyal and Zamore (2009); Liu and Paroo (2010)).

Apart from the RNA, two protein families play a major role in RNA interference. Dicer proteins – RNase III nuclease family members – have been shown to process longer double-stranded RNAs (dsRNAs) into functional smaller usually 21-22 nt long oligonucleotides (Bernstein et al. (2001); Elbashir, Lendeckel, and Tuschl (2001)). Argonaute family members in complex with a single-stranded small RNAs make up the functional unit for transcript regulation and repression called the RNA-induced silencing complex (RISC) (Hammond et al. (2000); Hammond et al. (2001); Azlan, Dzaki, and Azzam (2016); Meister (2013); Wilson and Doudna (2013)).

Many discoveries about the different RNAi pathways have been made in the model organism *Drosophila melanogaster* due to the fact that the proteins in the small interfering RNA (siRNA), microRNA (miRNA) and Piwi interacting RNA (piRNA) biogenesis pathways are distinct for each pathway. This allows an examination of a single pathway in detail without the possibility of cross-talk from components of one of the other pathways. In contrast, humans for example both the miRNA- and the siRNA-biogenesis is carried out by the same Dicer and Argonaute protein. Since the studies described in this thesis are investigating siRNAs and their biogenesis, *Drosophila melanogaster* was the ideal organism to perform the experiments in. In the following the biogenesis and molecular function of siRNAs, miRNAs and piRNAs in *Drosophila* is summarized to highlight similarities and differences between these different regulatory pathways.

1.1.1 siRNAs: *Drosophila's* defence against both exogenous and endogenous threats

In contrast to mammals *Drosophila* is lacking the interferon system as an innate immunity response pathway (Sabin, Hanna, and Cherry (2010)). Both DNA- and RNA-viruses produce dsRNAs as part of their replication cycle (Schuster, Miesen, and van Rij (2019); Bronkhorst and van Rij (2014); Weber et al. (2006)). These long exogenous dsRNA are detected by the cell and processed by Dcr-2 (Sabin et al. (2013)) into 21 nt long exogenous siRNAs (exo-siRNAs). These small dsRNAs have very defined characteristics. Of the 21 nucleotides 19 form a perfectly paired duplex. This results in a two-nucleotide overhang at each 3' end with a free hydroxyl group. The 5' ends of the duplex carry a phosphate (Ghildiyal et al. (2008); Elbashir et al. (2001); Kim, Han, and Siomi (2009); van Rij and Berezikov (2009)). After processing the siRNAs are loaded into AGO2 via the RISC-loading complex (RLC) made up of Dcr-2 and its cofactor R2D2. The thermodynamic stability at the 5' ends of the duplex determines which strand becomes the guide or passenger strand while loading (Schwarz et al. (2003); Khvorova, Reynolds, and Jayasena (2003); Liu et al. (2003); Tomari et al. (2004)). AGO2 cleaves the passenger strand and the endonuclease C3PO facilitates removal from the RISC (Matranga et al. (2005); Meister (2013); Kim, Lee, and Carthew (2007); Miyoshi et al. (2005); Rand et al. (2005)). The guide strand is methylated by Hen1 and targets the RISC to a perfectly complementary mRNA which is cleaved by AGO2 (Horwich et al. (2007); Liu et al. (2004); Meister et al. (2004)). After cleavage the mRNA fragments are degraded by exonucleases.

In flies transposable elements (TEs) pose a severe threat to genomic integrity and have to be tightly regulated. This regulation of transposon transcripts is done via endogenous siRNAs (endo-siRNAs) (Chung et al. (2008)). However, we now know that these endo-siRNAs are not only derived from TE transcription products but can originate from inverted repeats, transcripts of pseudo- and protein coding genes, self-complementary mRNAs and bidirectional transcription (Okamura and Lai (2008); Liu and Paroo (2010)). The long dsRNAs are processed by Dcr-2 similar to exo-siRNAs (Okamura, Chung, et al. (2008); Okamura, Balla, et al. (2008); Ghildiyal and Zamore (2009)). However, in case of the endo-siRNAs Dcr-2 partners with Loqs-PD (Hartig et al. (2009); Zhou et al. (2009); Hartig and Forstemann (2011)). It has been shown that R2D2 and Loqs-PD functions partially overlap and that there is a certain degree of redundancy between these two Dcr-2 cofactors (Mirkovic-Hosle and Forstemann (2014)).

1.1.2 miRNAs: post-transcriptional gene expression regulators

pri-miRNAs – the initial miRNA precursors - are transcribed by RNA polymerase II from specific genomic loci that are commonly located in clusters (Lee et al. (2002); Lee et al. (2004); Bartel (2004); Ghildiyal and Zamore (2009)). The first processing step is performed by a RNase III protein called Drosha in the nucleus. Together with its partner protein Pasha (DGCR8 in humans) it forms the so-

called Microprocessor complex. The resulting processing product is the 60-70 nt long hairpin pre-miRNA (Lee et al. (2003); Denli et al. (2004); Gregory et al. (2004); Han et al. (2004)). In addition, pre-miRNA can arise from introns of protein coding genes. These so called mirtrons are debranched by the splicing machinery before undergoing the same processing steps as regular pre-miRNAs (Okamura et al. (2007); Ruby, Jan, and Bartel (2007)). Nuclear export of the pre-miRNAs is facilitated by Exportin-5 and Ran-GTP (Yi et al. (2003); Bohnsack, Czaplinski, and Gorlich (2004); Lund et al. (2004)). In the cytoplasm processing continues with another cleavage by the RNase III protein Dcr-1 in a complex with Loqs-PB (Forstemann et al. (2005); Jiang et al. (2005); Saito et al. (2005)). One of the strands of the resulting miRNA/miRNA*-duplex is loaded into AGO1 to form the RISC. This RISC does not facilitate nucleolytic cleavage of the target mRNA. Base pairing between the miRNA and the target sequence is not perfect but a short “seed” sequence near the 5’ end of the miRNA needs to base pair with the target for efficient repression (Forstemann et al. (2007); Brennecke et al. (2005)). The RISC usually targets mRNAs in their 3’-UTR and leads to translation inhibition and deadenylation. Many of these effects rely on proteins from the GW-family (Bartel (2009); Fukaya and Tomari (2012); Eulalio, Tritschler, and Izaurralde (2009)).

1.1.3 piRNAs: repressors of transposable elements in the germline

Long single-stranded piRNAs precursors originate from genomic piRNA clusters likely derived from defective transposon sequences. The mature ~25-30 nt long piRNAs are generated by an amplification loop called ping pong cycle that is driven by the two Argonaute family proteins AGO3 and Aub. In this amplification loop AGO3 is associated with sense piRNAs and Aub with the antisense piRNAs. A characteristic 10 nt overlap between the sense and antisense piRNA is the result of the ping pong amplification (Aravin et al. (2001); Brennecke et al. (2007); Gunawardane et al. (2007)). (Hartig, Tomari, and Forstemann (2007)). Additionally, Aub is directed by the antisense piRNAs to transposon transcripts to facilitate their cleavage or translational inhibition (Hartig, Tomari, and Forstemann (2007); Khurana and Theurkauf (2010); Siomi, Miyoshi, and Siomi (2010); Siomi et al. (2011)). The antisense piRNAs can also bind to another Argonaute family member called Piwi (hence their name). Piwi in complex with the piRNA causes transcriptional silencing of transposon loci (Wang et al. (2015)).

1.2 siRNA response to a DSB

DNA double-strand breaks (DSB) are the most severe threat to genome integrity. Therefore, correct recognition and repair are one of the most important tasks a cell has to deal with. Incorrect repair that is not recognized by the cell can lead to cancer by disruption of tumor suppressor genes or chromosomal rearrangements (Ceccaldi, Rondinelli, and D'Andrea (2016)). If cells fail to repair or to

recognize erroneous repair events, they undergo apoptosis. The two main pathways responsible for DSB-repair are homologous recombination (HR) and non-homologous end joining (NHEJ) which in addition to the DSB response pathway have been studied extensively over the last decades.

In recent years small RNAs originating from DSBs have been detected in various species like *Neurospora crassa* (Lee et al. (2009)), *Arabidopsis thaliana* (Wei et al. (2012)), *Drosophila melanogaster* (Michalik, Bottcher, and Forstemann (2012)) and in human cell culture ((Francia et al. 2012)). The biogenesis and possible pathways in which these damage-induced RNAs (diRNAs) are involved with are studied by the small RNA and DSB repair scientific community since their discovery. The presence of diRNAs in such a wide variety of species suggests a conserved mechanism for both biogenesis and function.

Links between the diRNAs and HR have been proposed but a study from our lab (Schmidts et al. (2016)) could show that siRNAs do not affect DSB-repair via HR in *Drosophila melanogaster*.

Apart from the function of the diRNAs their biogenesis is still not fully explored. There are studies that show that there might be different species of diRNAs that depend on separate proteins for their biogenesis (Bonath et al. (2018)). Our lab studied factors involved with diRNA biogenesis on a genome-wide scale in *Drosophila*. The screen showed that siRNA biogenesis near a DSB is affected by a variety of proteins like splicing factors, *blanks* (see chapter 2) and of course factors from the siRNA biogenesis pathway like Dcr-2 (Merk et al. (2017)).

2 The molecular function of *blanks* in the nuclear export of siRNA precursors

Parts of this chapter have been previously published in Nitschko et al. (2020) which is marked at the corresponding positions.

2.1 Introduction and aim of the project

Small RNA biogenesis has been extensively studied in *Drosophila* and the processing of the initial RNA precursors into functional siRNAs, miRNAs and piRNAs is reasonably well understood. However, one step in endo-siRNAs biogenesis has not yet been conclusively addressed: export of the double-stranded RNA precursors from the nucleus to the cytoplasm. For the miRNA processing pathway, the export of pre-miRNAs has been shown to be facilitated by Exportin-5 (Lund and Dahlberg (2006), Lund et al. (2004)) and XPO1 (*emb* in *Drosophila*) (Bussing et al. (2010)). It seems reasonable to assume that the dsRNA-precursors of endo-siRNAs are exported by the same factors but the question has not yet been addressed conclusively.

In a genome-wide screen performed in our lab the dsRNA binding domain containing protein Blanks (also described as *CG10630/lump*) was found to have an influence on the siRNA generation following a DNA double-strand break (Merk et al. (2017)). The similarity of the two predicted dsRNA binding domains to the ones present in the Dicer cofactors R2D2 and Loqs led us to further study this protein and its potential to act in a sRNA processing pathway. Blanks has been described in previous studies to contribute to RNA silencing (Gerbasi et al. (2011); Zhou et al. (2008)) and to be essential for sperm individualization and thereby male fertility in *Drosophila* (Sanders and Smith (2011)). Furthermore, it was described to be involved in chromatin regulation (Schneiderman, Goldstein, and Ahmad (2010), Swenson et al. (2016)). A recent study also found it to cause misregulation of several mRNAs in testes of *blanks* mutants (Liao, Ai, and Fukunaga (2018)). Although these studies gave us insights into what processes *blanks* might be involved with, no molecular mechanism was proposed. Our lab wanted to address this blind spot and started an investigation of *blanks*.

Initial results were described in another dissertation (Kunzelmann (2017)) and will be summarized here shortly to provide a summary of what was known at the beginning of the studies performed for this manuscript. An initial bioinformatic analysis of Blanks revealed three domains of interest. The first one is a lysine rich amino acid stretch that is predicted to be a nuclear localization sequence (NLS) by several web tools. During the course of experiments done for this manuscript this analysis was reconfirmed with another prediction tool called NLStradamus (Nguyen Ba et al. (2009)). It predicts amino acids 86 to 101 (GRKKQKKKENKKAKIR) to constitute an NLS. Additionally, two dsRNA

binding domains (dsRBDs) are found in the Blanks sequence. DsRBD2 shows strong sequence homology to the dsRBD1 and dsRBD2 found in Dicer cofactors R2D2 and Loqs (PACT and TRBP in humans). This similarity lead to our initial hypothesis that Blanks might be involved in siRNA biogenesis. The amino acid sequence of the first dsRBD (dsRBD1) diverges from the canonical motive and is more similar to dsRBD3 in Loqs (or PACT). These domains have lost their RNA-binding capability and have evolved to be protein-protein interaction domains (Jakob et al. (2016)).

It is difficult to find an orthologous gene in other species than *Drosophila* since all amino acid sequence-based search results are dominated by the similarities of the dsRBDs to the ones found in other proteins. Structural homology prediction tools like HHpred show homology of Blanks to ILF3/NF90. However even this homology is restricted to the dsRBDs and a small basic stretch upstream of dsRBD1.

The homology of the dsRBDs to Loqs and R2D2 and *blanks*' appearance in our screen which identified it as a factor involved in siRNA biogenesis at the DNA double-strand break led to the initial hypothesis that it might be another factor interacting with Dcr-2 to facilitate siRNA processing and AGO2 loading. However, no interaction could be found between Blanks and Dicer-2 in both a mass spectrometry and in co-immunoprecipitation experiments. Blanks also did not contribute to cytosolic RNAi, a result that was confirmed by another experimental approach in this study.

Initial sequencing experiments performed in the lab also revealed that Blanks does not influence AGO2 loading of mature siRNAs and repression of transposon transcripts (s. also the additional analysis performed for this the manuscript). However, these sequencing results also revealed the presence of blanks-dependent siRNA loci (later called bepsiRNAs), which were further analysed and characterized for this manuscript.

The mass spectrometry experiments mentioned above also showed Blanks to interact with proteins involved in nuclear export, namely *Bj1* (*Rcc*), *Kap-α3* and *Ran*. This discovery was the basis to the hypothesis that Blanks might be involved in nuclear export of dsRNAs that is addressed below.

Based on the knowledge that was gained by our lab and others the aim for the experiments described in this manuscript was to elucidate Blanks' molecular role and to uncover which pathways it interacts with.

2.2 Results

2.2.1 Cytosolic RNAi is not influenced by Blanks

In previous work performed in the lab *blanks* was shown to have no influence on cytosolic RNAi (Kunzelmann (2017)). The experimental setup was a treatment of cells with dsRNA against GFP and a

subsequent transfection with a GFP-encoding plasmid. The dsRNA is processed to siRNAs which represses GFP expression. The experiment was performed in the context of cell lines with a copper-inducible expression of *Dcr-2* or *blanks*. Expression of these genes is thereby controllable and can be turned off completely when no copper is added to the cell culture medium (Kunzelmann et al. (2016)). *Blanks* expression levels did not change the efficiency of repression of the GFP reporter. However, if *Dcr-2* expression is turned off GFP repression was strongly diminished which is expected since the dsRNAs can no longer be processed into the functional siRNAs.

Since the GFP-expressing plasmid was under the control of a ubiquitin promotor and thus strongly expressed itself, it could lead to the repression of its transcripts. We thus wanted to validate the results with an alternative experimental setup in which we can avoid this indirect effect. The mentioned inducible N-terminally FLAG-tagged Blanks cell line was created via our lab's CRISPR/Cas9 tagging protocol (Bottcher et al. (2014); Kunzelmann et al. (2016)) and constitutively expresses Cas9. We chose Cas9 as our RNAi target since changes in expression between the FLAG-Blanks and a non-tagged parental cell line should be minimal and not have an influence on the experiment.

For the experiments cells were treated with dsRNA against Cas9 and *Renilla* Luciferase (RLuc) as a control. Before the knockdown cells were either cultured in regular cell culture medium or medium with 200 µM $CuSO_4$ to induce expression for a week. The concentration of 200 µM $CuSO_4$ was previously determined to give similar *blanks* expression levels as in non-induced wildtype cells (Kunzelmann (2017)). A protein extract of the samples was used for Western blot analysis. The efficiency of the Cas9 knockdown can be clearly seen in Figure 2.1A for both the FLAG-Blanks and the parental cell line whereas Cas9 levels are not affected in the control knockdown conditions. Cas9 levels of three replicates were quantified and normalized to β-tubulin and RLuc knockdown levels. The quantification in Figure 2.1B shows that there is no difference in the Cas9 knockdown efficiency between the samples with and without induction of *blanks*. The difference in overall knockdown efficiency between the copper-inducible and the parental cell line can be explained by small differences that are present between different clonal cell lines and should not have an influence on the conclusion drawn from the results.

With this experiment we were able to corroborate the result from the previously performed experiment that *blanks* does not influence cytosolic RNAi while also avoiding indirect RNAi effects inherent in the other experimental setup.

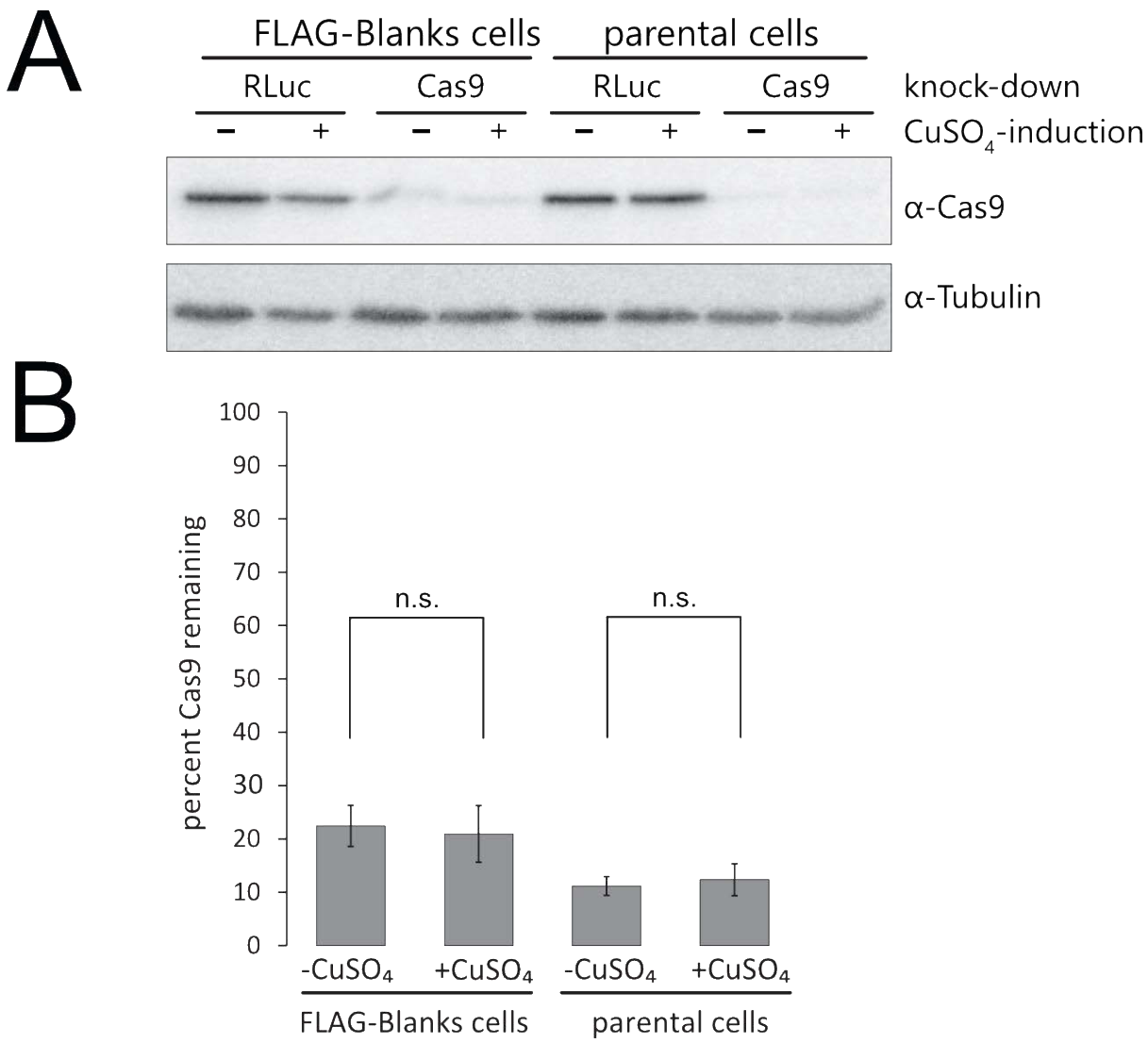

Figure 2.1 Blanks does not influence on cytosolic RNAi
(A) Western blot of Cas9 and tubulin (control) levels in RNAi treated cells (Cas9 or Renilla Luciferase (RLuc)) from copper-inducible FLAG-Blanks cell line and the parental control line. *blanks* expression was induced with 200 µM $CuSO_4$. One blot representative of all three replicates is shown.
(B) Quantification of Cas9 Western blot band intensity normalized to tubulin bands and RLuc levels with the same experimental conditions.
Figure adapted from Nitschko et al. (2020)

2.2.2 Blanks shuttles between nucleus and cytoplasm

Blanks has been reported to be almost exclusively localized in the nucleus of S2 cells (Gerbasi et al. 2011). The initial proteomics screen from our lab (Kunzelmann (2017)) showed a significant enrichment of factors involved in nuclear trafficking like *Bj1* (*Rcc*), *Kap-α3* and *Ran* in association with Blanks. This led to the hypothesis, that Blanks might be involved in the nuclear export of dsRNAs. To test this hypothesis a S2 cell line with endogenously GFP-tagged Blanks was generated

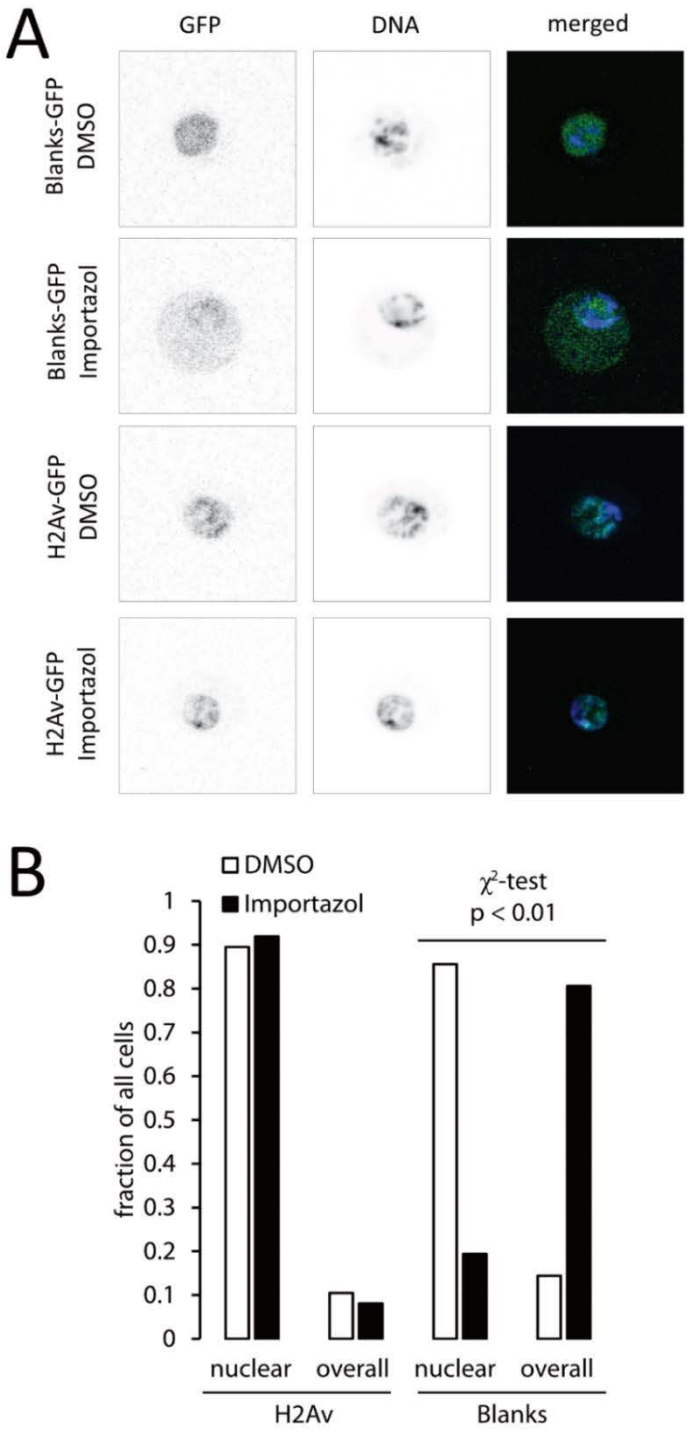

Figure 2.2 Blanks can shuttle between nucleus and cytoplasm
(A) Fluorescence microscopy images of the endogenously GFP-tagged Blanks- and H2Av-cell lines in the Importazole assay. Cells have been treated with 200 µM Importazole or the same volume of DMSO in the control conditions. DNA staining was performed with Hoechst33342.
(B) Quantification of cells in the Importazole assay that show a nuclear localized or a fluorescence signal over the whole cell ("overall") originating from the GFP-tagged Blanks or H2Av.
Figure was already published in Nitschko et al. (2020)

which allows the direct observation of Blanks localization via fluorescence microscopy. Our cell line showed the majority of Blanks to be localized to the nucleus (Figure 2.2), reconfirming the previously

described localization. To disturb the nuclear export and import pathways the small molecule Importazole was employed. Importazole blocks nuclear import via β-importin family members (Soderholm et al. (2011)). When the Blanks-GFP cell line was treated with Importazole, the localization of Blanks changed from mainly nuclear to distributed over the whole cell (Figure 2.2). This shows that Blanks can indeed leave the nucleus and accumulates in the cytoplasm when nuclear import is blocked via Importazole.

To confirm that this effect is not resulting from the accumulation of newly translated protein which cannot enter the nucleus after Importazole treatment the assay was repeated in a cell line with GFP-tagged histone H2Av. Under the DMSO control conditions H2Av shows the nuclear localization that is to be expected from a histone protein. When the cell line is treated with Importazole analogous to the Blanks-GFP cell line, the localization of H2Av is still almost exclusively nuclear. The localization behaviour of the two nuclear proteins differs significantly if nuclear import is blocked (Figure 2.2). This indicates that the effect of newly translated proteins on a changed localization pattern can be neglected for the relatively short time frame in which the assay was conducted.

Comparing the results of the assay for the two nuclear localized proteins shows that Blanks seems to have the ability to shuttle between nucleus and cytoplasm. One resulting hypothesis would be that it is a novel export factor or adapter protein for nuclear export of dsRNAs.

2.2.3 Interaction of Blanks with export related proteins

The results from the Importazole assay show that Blanks can shuttle between the nucleus and the cytoplasm. Additionally, the knowledge from the crosslink MS-experiments performed in Kunzelmann (2017) show that it might interact with export related factors like Bj1 (Rcc), Ran and Kap-α3. We wanted to validate the MS results with an independent method and study these potential interactors. Although they did not show significant interaction with Blanks in the MS experiments, we decided to include Karyβ3 and Exportin-5 to the proteins to be tested for interaction with Blanks since they belong to the same protein family as Kap-α3 and especially Exportin-5 has been shown in human cell lines to interact with dsRBPs (Gwizdek et al. (2004)). Since we already had C- and N-terminally tagged Blanks cell lines, we decided to add an epitope tag to the proteins mentioned above in these cell lines and subsequently perform co-immunoprecipitation to identify interactors.

Efforts were made to tag the five proteins (Kap-α3, Karyβ3, Bj1 (Rcc), Ran and Exp-5) in the background of both C- and N-terminally tagged Blanks with a Strep tag at the C- or N-terminus, a Strep-His-Strep-tag at the N-terminus or GFP at either terminus in both cell lines. However, tagging efficiencies were overall low. Although PCR results were positive for 31 of the 40 combinations

tested, only 11 cell lines showed a protein signal in the Western blots with antibodies targeting the epitope tags on the export related factors. Overall Ran and Karyβ3 showed detectable levels of tagged protein in most combinations in contrast to the other three proteins that could not be detected in any of the Western blots. Since Ran was one of the strongest interactors in the MS experiments and one of the most important factors in nuclear export, we decided to test the interaction via Co-IP before making efforts to improve tagging of the other factors. The Co-IP was performed after a treatment with 0,1 % formaldehyde to stabilize weak interactions via crosslinks. Further details on the IP can be found in the methods section of this manuscript.

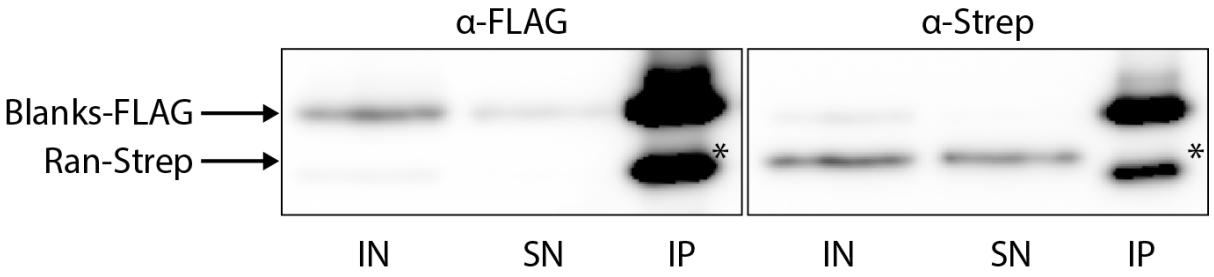

Figure 2.3 Western blot of Co-IP experiment from lysate of Blanks-FLAG (bait) and Ran-Strep cell line
Input (IN), supernatant (SN) and IP fractions of the Co-IP experiment were used for analysis on the Western blot. For detection of the epitope tagged proteins the blot was treated with a α-FLAG antibody (left) and an α-Strep (right). The Blanks-FLAG bands can be seen on the Strep blot because stripping after α-FLAG treatment that was done first did not work efficiently. The two bands marked with an * are unspecific bands from the α-FLAG antibody.

As can be seen from Figure 2.3 there was no interaction between Ran and Blanks in the experiment. The IP with Blanks-FLAG as a bait itself was very efficient and could recover most of the protein in the lysate (see IP lane on the right). However, Ran can only be detected in the input and supernatant fractions and not bound to the beads. This initial result indicates that the interaction between Blanks and Ran is either very weak and under the detection limit of the Strep blot or non-existent.

Since we were not as successful in tagging and/or detecting the export related proteins with the GFP- and Strep-epitope tags, we decided to approach the problem from the other side. We know that the interaction of FLAG-epitope and antibody we use is very strong and that we can detect even small amounts of tagged proteins very efficiently. Additionally, we knew from the past experiments that Blanks was relatively easy to tag. This is why we decided to start over in creating a double-tagged cell line. This time we started by tagging the five export related proteins with a FLAG-tag which could be achieved without many problems. We decided to do our Co-IP experiment in the cell lines in which we had the FLAG-tag at the N-terminus of the export related factors since this also

gave us the opportunity to overexpress the proteins if needed via the mtnDE-promotor that is part of our N-terminal tagging cassette. In this background, we then tagged Blanks at the C-terminus with a V5-tag which from past experience gives strong signals on a Western blot even with relatively low protein concentrations. It also has the benefit of being resistant to crosslinks via formaldehyde since it does not contain any lysine residues in contrast to the Strep- or GFP-tags tested above which might also have affected sensitivity of the initial round of Co-IPs.

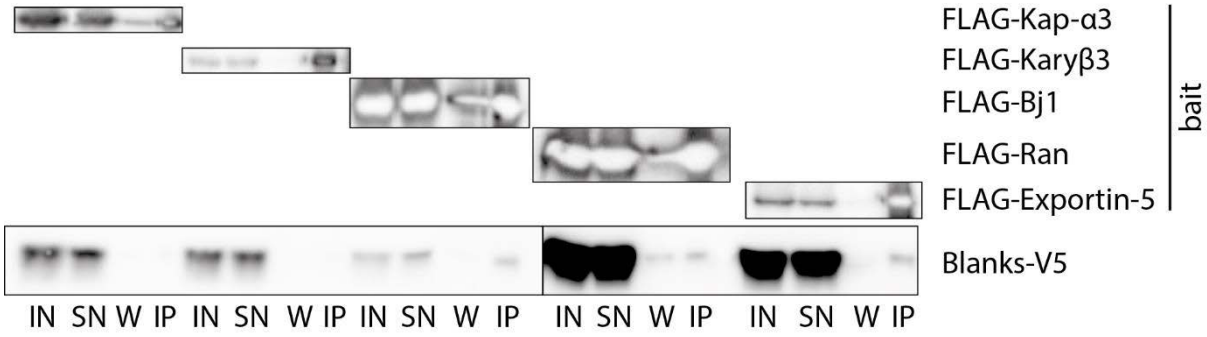

Figure 2.4 Western blot of Co-IP experiments from lysate of cell lines with N-terminally tagged export related proteins (bait) and C-terminally V5-tagged Blanks
Input (IN), supernatant (SN), wash (W) and IP fractions of the Co-IP experiment were used for analysis on the Western blot. Due to the high amounts of export related proteins used in the IPs the substrate used for Western blot development was used up after a short while which results in the hollow bands visible on the blots. The black and white levels for the image of the Ran and Exportin-5 blot (right) had to be adjusted to visualize the Blanks bands in the W and IP fractions. No additional band was visible on the image for the left part of the figure when levels were tweaked and no adjustment was done.

On the Western blot in Figure 2.4 showing the results from the Co-IPs a Blanks band can be seen in the IP lanes for the IPs with the Bj1, Ran and Exp-5 baits. This means there might be an interaction between these proteins and Blanks. This interaction seems to be weak as the bands are barely detectable. In addition, we performed the Co-IP and Western blot from the same cell line in two more replicates with varying results. IP efficiencies of the export related bait proteins remained unchanged, but the signal for Blanks differed from the initial replicate with no change in the experimental setup or the expression levels. In the second replicate Blanks only associated with Ran and the interaction to Bj1 and Exp-5 that was detected in the first replicate could not be confirmed. In the third replicate Blanks did not associate to any of the three proteins for which it showed interaction in the first.

All in all, we could not validate the results from the MS experiment in regards to an interaction between Blanks and export related proteins. Although we had some Co-IPs in which Blanks seemed to associate with Bj1, Exp-5 and Ran, these results varied over the three performed replicates. This could mean that the interaction between Blanks and the tested proteins might be very weak and

thereby hard to detect. However, we cannot confidently exclude the possibility that these detected interactions might have been false positives.

2.2.4 RNA-binding is dependent on the second dsRBD

Blanks carries two domains that are annotated as double-stranded RNA binding domains (dsRBDs).

The first dsRBD shows a degenerate amino acid sequence to the canonical dsRBD motive which resembles the amino acid sequence of the dsRBD3 Loqs or its human homolog TRBP. In the case of Blanks two lysines that confer the RNA-binding activity are mutated to $V^{165}N^{166}$. It is thus unlikely that Blanks binds RNA via this dsRBD1 domain. However, it might have evolved to have a different role and has been previously shown to be important for male fertility in *Drosophila* (Sanders and Smith (2011)).

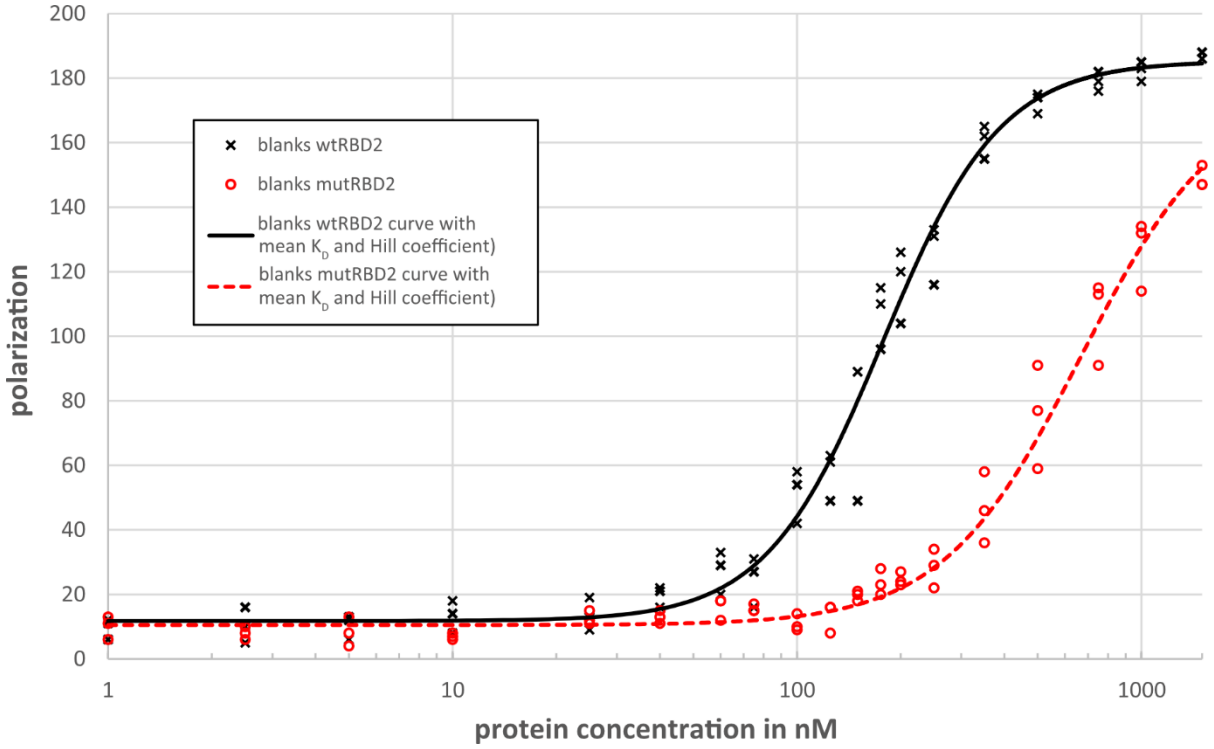

Figure 2.5 Blanks binds dsRNA via its dsRBD2
Quantification of the anisotropy measurement with the two recombinantly expressed Blanks variants with either a wildtype or mutant dsRBD2. Data points of all three replicates per condition and the fitted curves calculated from the mean K_D and Hill coefficients are shown.
Figure was adapted from Nitschko et al. (2020).

The amino acid sequence of the second dsRBD of Blanks shows strong amino acid sequence homology to the canonical dsRBD motive and dsRBDs from sRNA processing factors like Loqs (dsRBD1/2) or R2D2 (dsRBD1/2). It is thus more likely that RNA-binding by Blanks is conferred by dsRBD2. To test if Blanks binds dsRNA and if the potential binding is via the dsRBD2, two Blanks variants were recombinantly expressed and purified. The first one has the wildtype sequence whereas for the second variant the two lysine residues that are predicted to be responsible for RNA-

binding according to the homologs were replaced with alanines ($K^{301}K^{302}$ → AA). RNA-binding to a fluorescently labelled 23 nt long double stranded synthetic RNA (21 base pairs and 2 nt overhang at each 3' end) of both purified Blanks variants was measured via fluorescence anisotropy.

The wildtype protein shows RNA-binding in the nanomolar range (K_D=177 ± 22 nm, n=3) (Figure 2.5) which is comparable to RNA-binding via a single dsRBD of other dsRBPs like Loqs (Tants et al. (2017)). The variant with the mutation in dsRBD2 shows an almost four-fold decrease in binding affinity (K_D=666 ± 197 nm, n=3) (Figure 2.5).

These results show that Blanks can bind dsRNAs and that binding is dependent on an intact second dsRBD.

2.2.5 RNA-binding is not related to male fertility defect of mutants

It has been previously reported that *blanks* mutant *Drosophila* males are sterile (Sanders and Smith (2011)). To study if this fertility defect is related to Blanks' ability to bind dsRNA or its localization, we generated transgene flies carrying the following Blanks variants via microinjection of plasmids coding for the corresponding constructs: Blanks wildtype ("wildtype"), Blanks with an additional nuclear localization sequence (NLS) ("NLS-Blanks") and Blanks with the additional NLS and a mutant dsRBD2 ("NLS-Blanks mut. dsRBD2").

The extra NLS is thought to limit the time that Blanks is localized to the cytoplasm after nuclear export and to facilitate quick reimport. The mutation in the second dsRBD abolishes dsRNA binding as seen in 2.2.4. The transgene is inserted into the background of an intact wildtype version of *blanks*. The check if the transgene itself has a dominant negative effect on male fertility, a fertility assay was conducted (Figure 2.6A). In the background of a wildtype copy of *blanks*, the transgenes did not have a negative effect on male fertility.

The transgenic flies were used to create flies carrying the transgene in the background of a knockout mutant of *blanks* (*blanks*^{MI10901}) to potentially rescue this phenotype. The mutant itself was sterile and all five crosses in the fertility assay resulted in no F1 flies (Figure 2.6B). The wildtype construct was able to rescue this fertility defect completely. The rescue with NLS-Blanks showed a strongly reduced fertility indicating that the altered localization has an effect on fertility. When the NLS is combined with a mutant dsRBD2 to rescue the mutant, the decreased fertility caused by the NLS alone is partially restored. These results show that the dsRBD2 is not required for male fertility. The decrease in fertility observed for NLS-Blanks could result from unspecific binding of the overexpressed transgene to structured RNAs that are thereby prevented from being exported to the

cytoplasm. Indeed, in past studies alterations to hairpin RNA levels in testes had detrimental effects on male fertility in *Drosophila* ((Wen et al. (2015)).

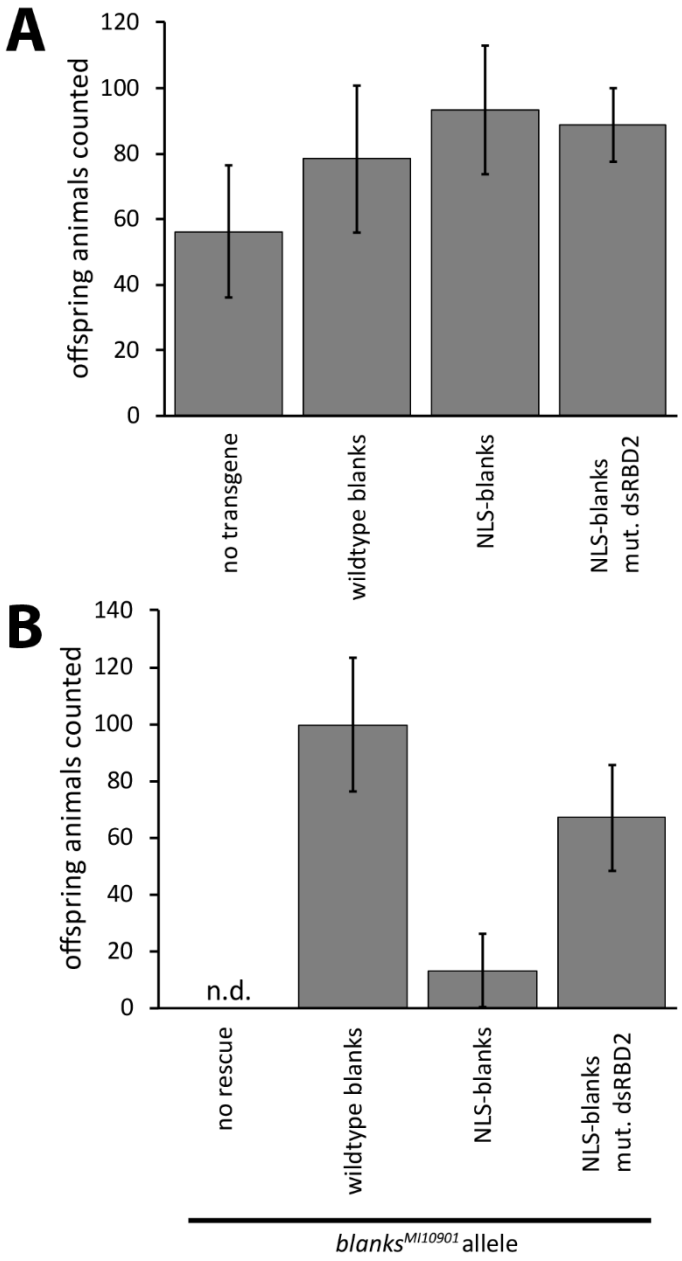

Figure 2.6 Male fertility is not dependent on the dsRNA-binding ability of Blanks
For each condition the mean of five replicates is shown. The error bars indicate the standard deviation
(A) Quantification of the fertility of the *blanks*-transgene flies with a wildtype copy of *blanks* in the background.
(B) Quantification of homozygous males carrying the *blanks* transgene in the background of a *blanks* knockout mutant (*blanks*^{MI10901}). For the no rescue condition, no pupae developed (none detected, n.d.).
Figure was adapted from Nitschko et al. (2020).

2.2.6 Nuclear export of structured RNAs is impaired in dsRBD2-mutants

In the fertility assay we observed that an overexpression of NLS-Blanks showed a limited ability to rescue the sterility of males in a *blanks* mutant background. Since this effect was not observed for NLS-Blanks with one mutant dsRBD2 a hypothesis is that NLS-Blanks might bind to structured RNAs and trap them in the nucleus.

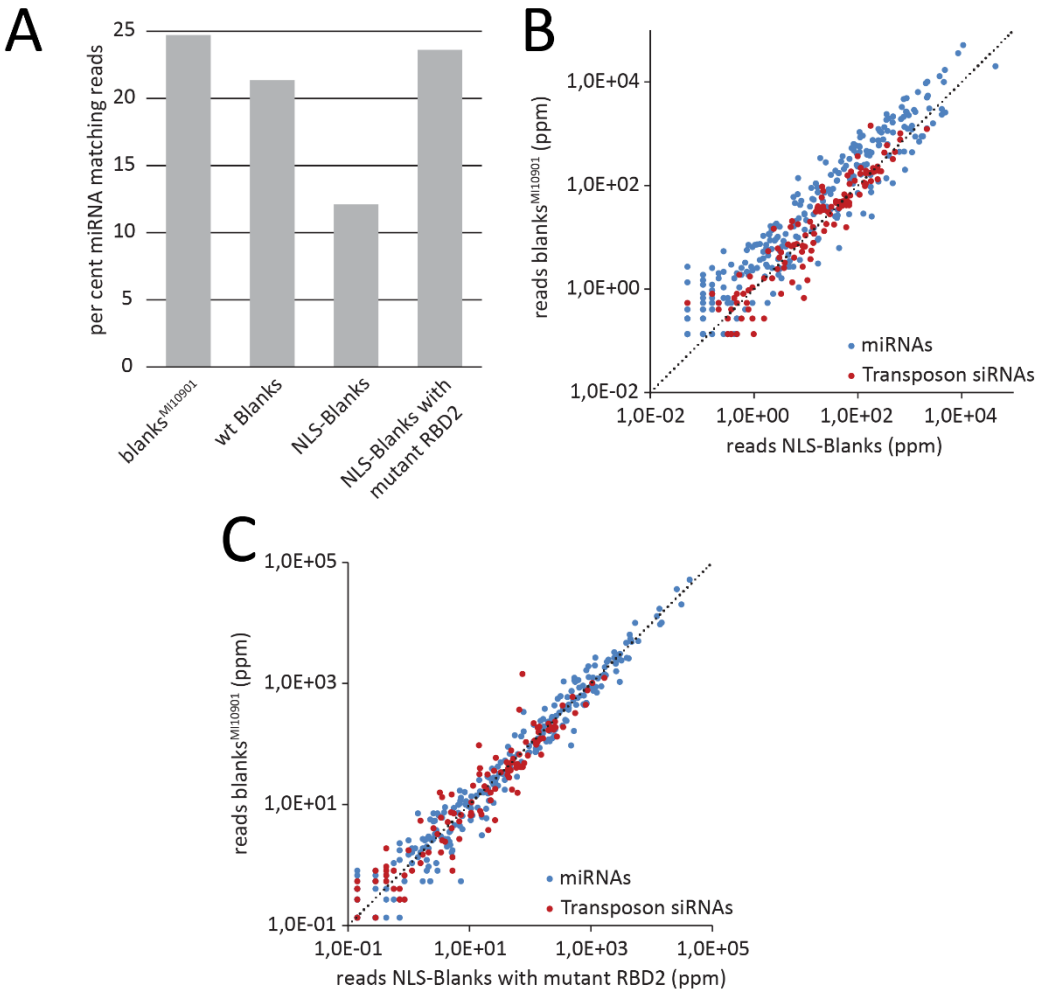

Figure 2.7 Overexpression of NLS-blanks unspecifically traps miRNA precursors in the nucleus
(A) Quantification of relative mature miRNA reads in the libraries generated from RNA from testes of *blanks* transgenic rescue fly lines.
(B) Comparison of reads counts for single miRNAs and transposon derived siRNAs between the mutant and the NLS-*blanks* rescue library.
(C) Comparison of reads counts for single miRNAs and transposon derived siRNAs between the mutant and the NLS-*blanks* mut. dsRBD2 rescue library.
Figure was adapted from Nitschko et al. (2020).

Deep sequencing data from testes was analysed assess at miRNA levels. Since miRNAs are processed from a double-stranded structured precursor they might be affected by trapping to the nucleus via overexpressed NLS-Blanks. Indeed, when we compared the relative levels of mature miRNAs between the libraries, we saw that they did not differ between the mutant, the wildtype rescue and

the rescue with NLS-Blanks with a mutant dsRBD2. However, in the NLS-Blanks rescue library the relative number of miRNAs was roughly halved (Figure 2.7A). This is consistent with the hypothesis formulated in the fertility assay: The structured miRNA precursors are likely trapped by overexpressed NLS-Blanks in the nucleus preventing nuclear export and processing by Dcr-1 to mature miRNAs. This decrease in mature miRNA levels is not resulting from a specific decrease of a small group of miRNAs but observed over all miRNAs (Figure 2.7B). In the case of the additional dsRBD2 mutation the unspecific binding and trapping cannot take place and mature miRNA levels are the same as in the mutant (Figure 2.7C) and wildtype. In contrast to the levels of mature miRNAs transposon derived siRNAs are not affected.

2.2.7 Blanks-dependent siRNAs are originating from genomic loci with convergent transcription

Previous sRNA sequencing experiments performed in the lab found sRNAs originating from several genomic loci that are dependent on *blanks* expression (Kunzelmann (2017)). Moreover, these sRNAs also showed *Dcr-2* dependency and were mainly 21 nt long. Reads for these genomic loci were detected in both sense and antisense direction. Taken together this implies that these *blanks*-dependent sRNAs are siRNAs. One common feature of these *blanks*-dependent siRNAs was their origin at loci with convergent transcription of overlapping genes or genes with close proximity to each other. Transcription from both genes at these loci would result in RNAs that can form a dsRNA precursor which can be processed into siRNAs in the cytoplasm.

To test the hypothesis that *blanks*-dependent siRNAs arise from loci with convergent transcription, the sequencing data was re-analysed. We created a list of genomic regions for which convergent transcription could lead to the formation of dsRNAs. The list was created by extending the annotated genes from *Drosophila* genome version r6.02 (Flybase FB2014_05) by 300 nt at the 3'UTR and then checking them for overlap with another gene in antisense direction (details s. 4.1.3). The extension by 300 nt was added since transcription termination does not necessarily coincide with the polyA-site which usually marks the end of the annotated 3'UTR. Short lived transcripts exceeding the annotated 3'UTR could form dsRNAs with other RNAs stabilizing it in the process and creating a potential siRNA precursor. The final list contained 4366 genomic regions with the potential to produce these dsRNAs.

The small RNA libraries were generated (Kunzelmann (2017)) from S2 cell lines with a copper-inducible expression of either *blanks* or *Dcr-2* (Kunzelmann et al. (2016)). In these cell lines

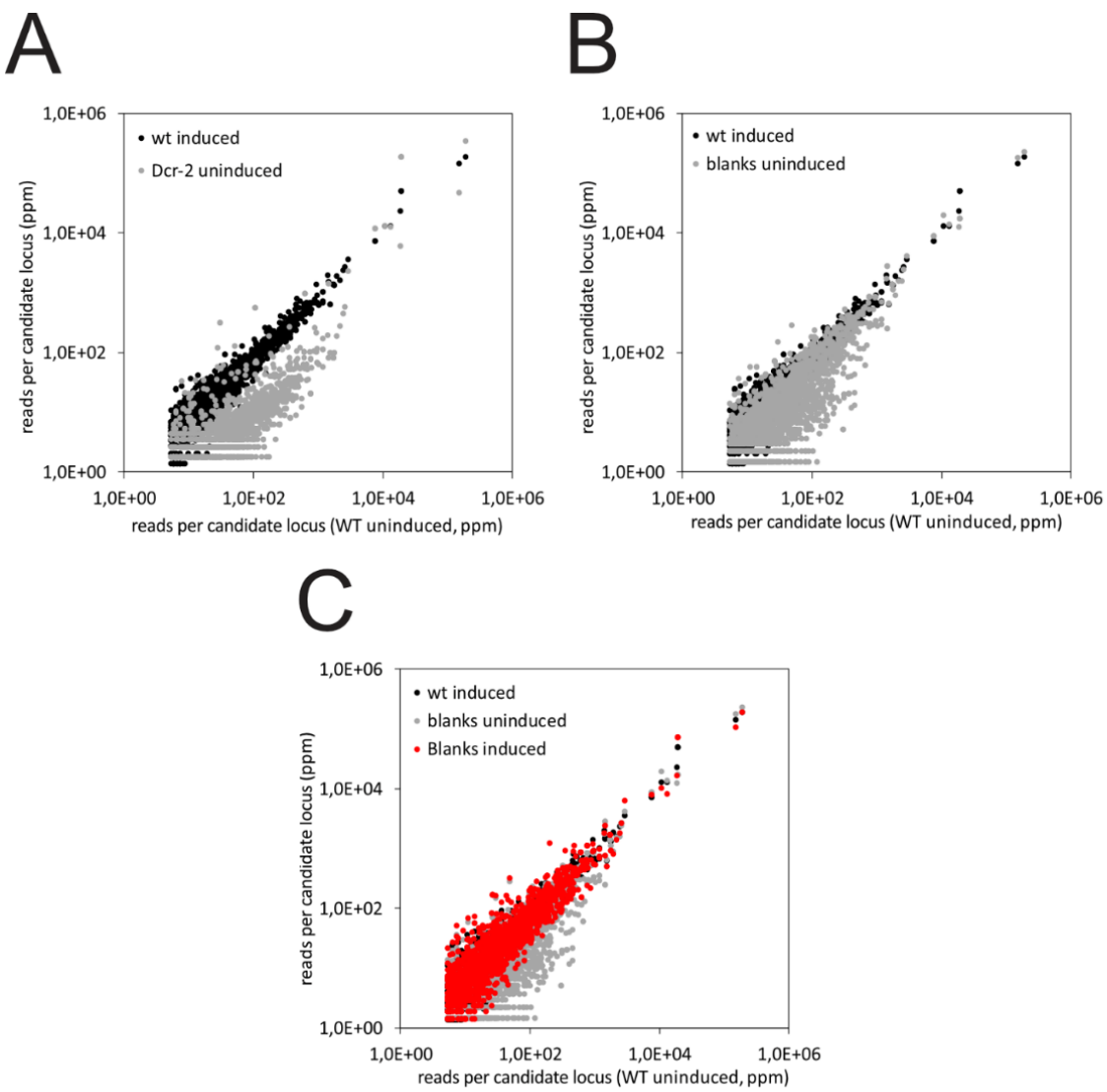

Figure 2.8 A population of genomic loci with convergent transcription produces siRNAs only in the presence of Blanks

sRNA sequencing reads from S2 cell lines with copper-inducible expression of *Dcr-2* or Blanks and a wildtype control cell line were mapped to genomic loci with the potential for dsRNA production via convergent transcription. Reads were normalised to number of genome matching reads (ppm).
(A) The majority of analysed loci shows a reduction in siRNA reads in the absence of Dcr-2 showing that the sRNAs observed are siRNAs.
(B) In the absence of Blanks a subset of loci shows a reduction in siRNA reads.
(C) The loci that showed reduced siRNA levels in the absence of Blanks return to wildtype levels when *blanks* is expressed.
Figure was adapted from Nitschko et al. (2020).

expression of the genes is fully dependent on induction via Cu^{2+}-ions. Both inducible cell lines and a wild type line (as a control) were treated with 0 or 200 µM $CuSO_4$ before RNA isolation for 4 days. The library reads were size selected to 21 nt to enrich for mature siRNAs, depleted of transposon matching reads and then mapped via bowtie to the list of potential overlapping convergent transcription that was described above. Of the 4366 loci contained on the list, 1748 showed >5 ppm reads which was used as a cut-off in the analysis. In the wildtype cell line sRNA levels were

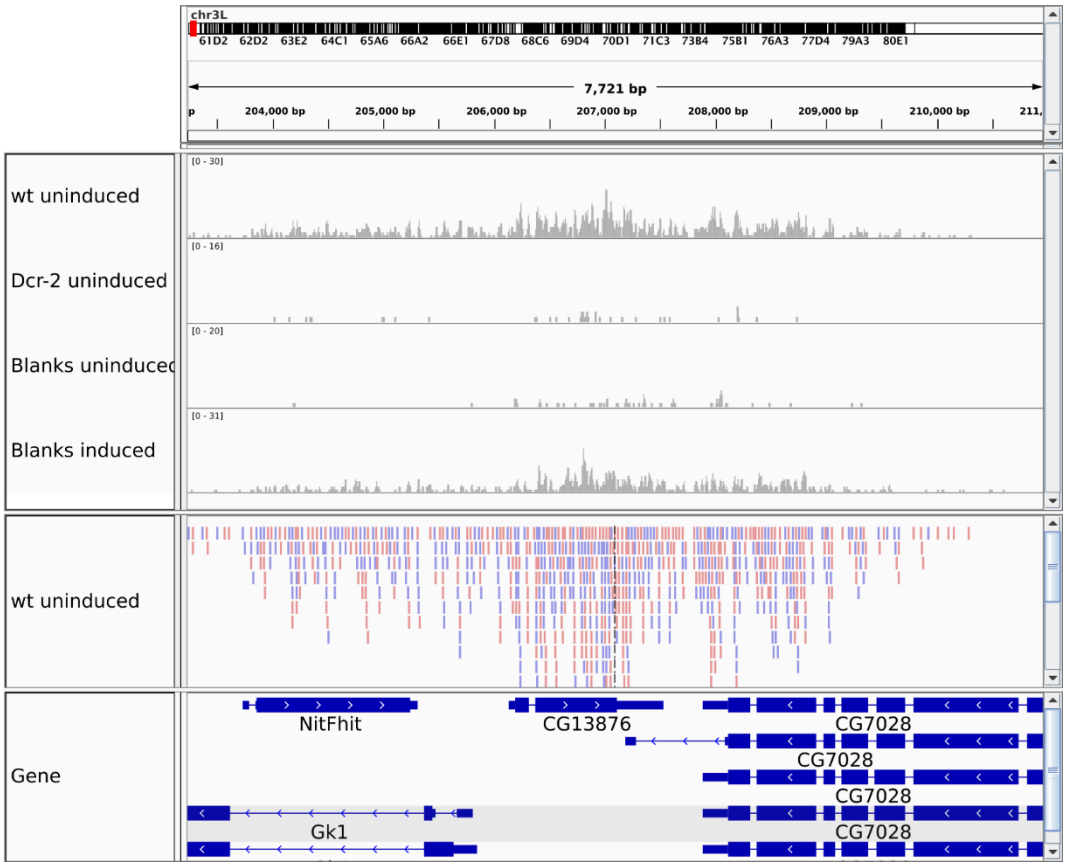

Figure 2.9 IGV browser shot of an exemplary gene region with bepsiRNA loci
The coverage tracks for 21 nt long sRNAs from the indicated cell lines and their treatment are shown in the upper part. Below individual reads from the wildtype library are color coded based on their orientation (red: 5'→3' left to right; blue: 5'→3' right to left). On the bottom the genes giving rise to the sRNAs are annotated.
Figure was adapted from Nitschko et al. (2020).

unaffected by the $CuSO_4$-treatment (Figure 2.8A). In cells without *Dcr-2* expression sRNA levels were overall strongly reduced showing that the observed reads are resulting from siRNAs. In contrast, in the libraries from cells without *blanks* expression most of loci do not show a change in small RNA levels. However, there is a subset of loci which display decreased sRNA levels in comparison to the wildtype control (the grey dots below the diagonal; Figure 2.8B). The reduction in sRNA levels can be reverted by induction of *blanks*, which restores them to wildtype levels (Figure 2.8C). These results indicate that there is a subpopulation of *bona fide* siRNAs whose biogenesis is dependent on *blanks* expression. We named this population of siRNAs bepsiRNAs (for **B**lanks **e**xported **p**recursors siRNAs) as we could link the export of their precursors to *blanks* expression. This will be addressed in more detail later in this manuscript.

In Figure 2.9 the features of bepsiRNA-loci can be seen on two gene pairs with convergent transcription. *Nitfhit* is located in the first intron of *Gk1* but is transcribed in the opposite direction. The transcripts for *CG13876* and *CG7028* show convergent transcription and do not overlap except

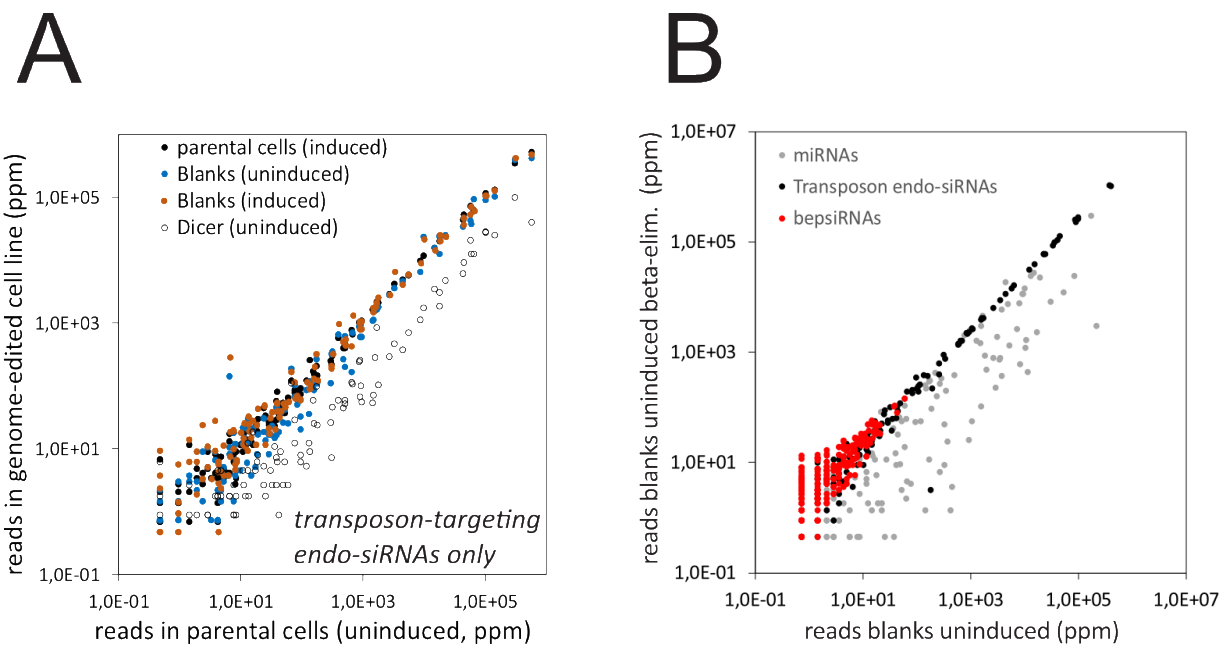

Figure 2.10 Blanks-dependent siRNAs do not target transposon transcripts and are correctly loaded into AGO2
(A) sRNA library reads from the inducible *Dcr-2* and *blanks* cell lines were mapped to the transposon consensus sequences. Without *Dcr-2* induction a reduction of reads is observable due to the resulting processing defect. There is no difference between the induced and uninduced *blanks* samples showing that *blanks*-dependent siRNAs do not target transposons
(B) Comparison of library reads from RNA samples one of which was treated with periodate (β-eliminated: "beta-elim") before library preparation. MiRNAs matching reads show a depletion in the treated sample due to missing modification after loading into AGO1. Transposon targeting and *blanks*-dependent siRNAs do not show a change in the relative number of reads showing correct loading into AGO2 and subsequent 2'-O-methylation.
Figure was adapted from Nitschko et al. (2020).

for a small part of the 3'UTR in case of one of the five *CG7028* transcripts. Under wildtype conditions the locus gives rise to sRNAs in both sense and antisense direction covering exonic, intronic and intergenic regions. Antisense and sense reads are represented in similar numbers as can be seen from the colour coded visualization of single reads in the middle of the figure. When *Dcr-2* expression is turned off barely any sRNA reads can be detected, supporting the hypothesis that these are *bona fide* siRNAs. This severe reduction in sRNA reads can also be observed when *blanks* expression is abolished using the copper-inducible cell line without inducing expression. If the same cell line is treated with 200 µM $CuSO_4$ to restore *blanks* expression to similar levels as in wildtype cells (Kunzelmann (2017)), small RNA reads can be detected again at similar levels as for wildtype cells. This shows that both *Dcr-2* as well as *blanks* are required for theses sRNAs to form.

One of the major population of siRNAs in *Drosophila* target transcripts of transposable elements (TEs). We analysed levels of transposon-targeting siRNAs in the context of our inducible cell lines to assess if *blanks* is involved in the biogenesis of these TE-targeting siRNAs. The library reads were

mapped to the transposon consensus sequences. In contrast to *Dcr-2* expression, Blanks levels had no influence on the number of TE-targeting siRNAs (Figure 2.10A).

The fact that the bepsiRNAs are detectable via sequencing does not mean they are biologically active. To fulfil any regulatory role, they have to be loaded into AGO2 after being processed by Dcr-2. Functional siRNAs are 2'-O-methylated at the 3'-nucleotide when they are successfully loaded into AGO2. This modification can be used to differentiate small RNAs in regard to their AGO2 loading state. Due to the methylated 2'-hydroxy group they can no longer be oxidized by periodate treatment. When treated with periodate they retain their ability to ligate to the adapters in the library preparation due to the accessible 3' hydroxy group. AGO1 loaded RNAs like miRNAs lack the 2'-O-methylation and are thereby depleted from libraries after periodate treatment. When comparing the sRNA reads from a periodate treated and an untreated sample, miRNA mapping reads are depleted as can be expected from AGO1 loaded RNAs. However, transposon mapping siRNA and bepsiRNA levels do not change (Figure 2.10B). They are thereby loaded into AGO2, methylated at their 2' hydroxy group by Hen1 (Ji and Chen (2012); Horwich et al. (2007)) and resistant to the periodate treatment. This shows that bepsiRNAs don't differ from regular siRNAs in regard to their AGO2 loading.

As the next step we wanted to see if the observations made for bepsiRNA in the *in vitro* cell culture setting represent the *in vivo* situation. Additionally, we wanted to verify our observations with another line of analysis. Since testes are the tissue in which *blanks* is almost exclusively expressed we dissected testes for RNA library preparation and deep sequencing. The fly lines we chose were the transgenic lines that were already used in the fertility assay described above. The sRNA reads were first size selected to 21 nt to limit analysis to mature siRNAs. Transposon targeting siRNAs were filtered out by mapping the remaining reads to the transposon consensus sequences and removal of the matches. These preselected reads were then mapped to all annotated gene sequences (*Drosophila melanogaster* genome release 6.02) that were extended with 150 bp of sequence at the 3'- and 5'-ends (details see also 4.1.3). These additional nucleotides were added to account for transcripts that extend the annotated 3'UTR but still might add to the pool of dsRNAs similar as it was done for the *in vitro* library analysis.

We first compared the read numbers from the mutant $blanks^{MI10901}$ library with the wildtype *blanks* rescue line library. A large number of genes can be found that show an increased read count in case of wildtype *blanks* (Figure 2.11A). Since there were differences in sequencing depth between the libraries we imposed an additional cut-off of 10 reads per library for a further semiquantitative analysis of the data. Our hypothesis was that Blanks interacts with the dsRNA precursors of

bepsiRNAs. We therefore wanted to analyse if the change in read numbers between the mutant and wildtype library correlates with the distribution of sense and antisense reads per gene. A roughly equal distribution of sense and antisense reads would be an indication for that. In Figure 2.11B we plotted the fold change in read numbers between the mutant and wildtype library for each gene against the ratio of sense reads mapping to the same gene in the wildtype library. A ratio of 0.5 indicates an equal number of sense and antisense reads for a given gene. As can be seen *blanks* dependent siRNAs indeed cluster at a sense read ration of around 0.5 affirming that Blanks might indeed interact with a dsRNA.

However, we observed that by imposing the cut-off of 10 reads per library in our analysis we also excluded a large number of loci for which none or only small numbers of reads could be found in case of the mutant but which had high read counts in the wildtype library. These loci are potentially the ones with the biggest fold changes but cannot be confidently quantified due to the small read counts in the mutant.

A manual inspection of some of the genes with the highest fold change in reads confirms what we already saw in the *in vitro* data: Many of these genes show convergent transcription with another gene. In Figure 2.11C-E three examples are shown. As can be easily seen in the coverage traces the reads that can be detected in the wildtype library are severely diminished or absent in the mutant library. Moreover, reads do not only cover exonic regions like for *CG8176* in Figure 2.11C but also are present between genes at sites downstream of annotated 3'UTRs (*Mitf*/*Dyrk* in Figure 2.11D) and in introns (*ksr* in *Hcs* intron in Figure 2.11E). This indicates that Blanks already binds to the dsRNA shortly after transcription before introns and processed transcripts downstream of 3'UTRs can be degraded.

2.2.8 BepsiRNA biogenesis is dependent on RNA-binding and shuttling

We further wanted to analyse the influence of the additional NLS and the mutated dsRBD2 we had present in our transgene rescue fly lines. Library preparation and analysis was done as above from dissected testes of these fly lines.

It is very clear that the dsRBD2 plays an important role in the biogenesis of the bepsiRNAs since a mutation leads to the same reduction in sRNA reads that can be observed for the knockout mutant (Figure 2.11C-E). This likely means that Blanks binds the dsRNA with its dsRBD2. These precursors could then be exported in complex with Blanks to the cytoplasm for downstream processing by Dcr-2.

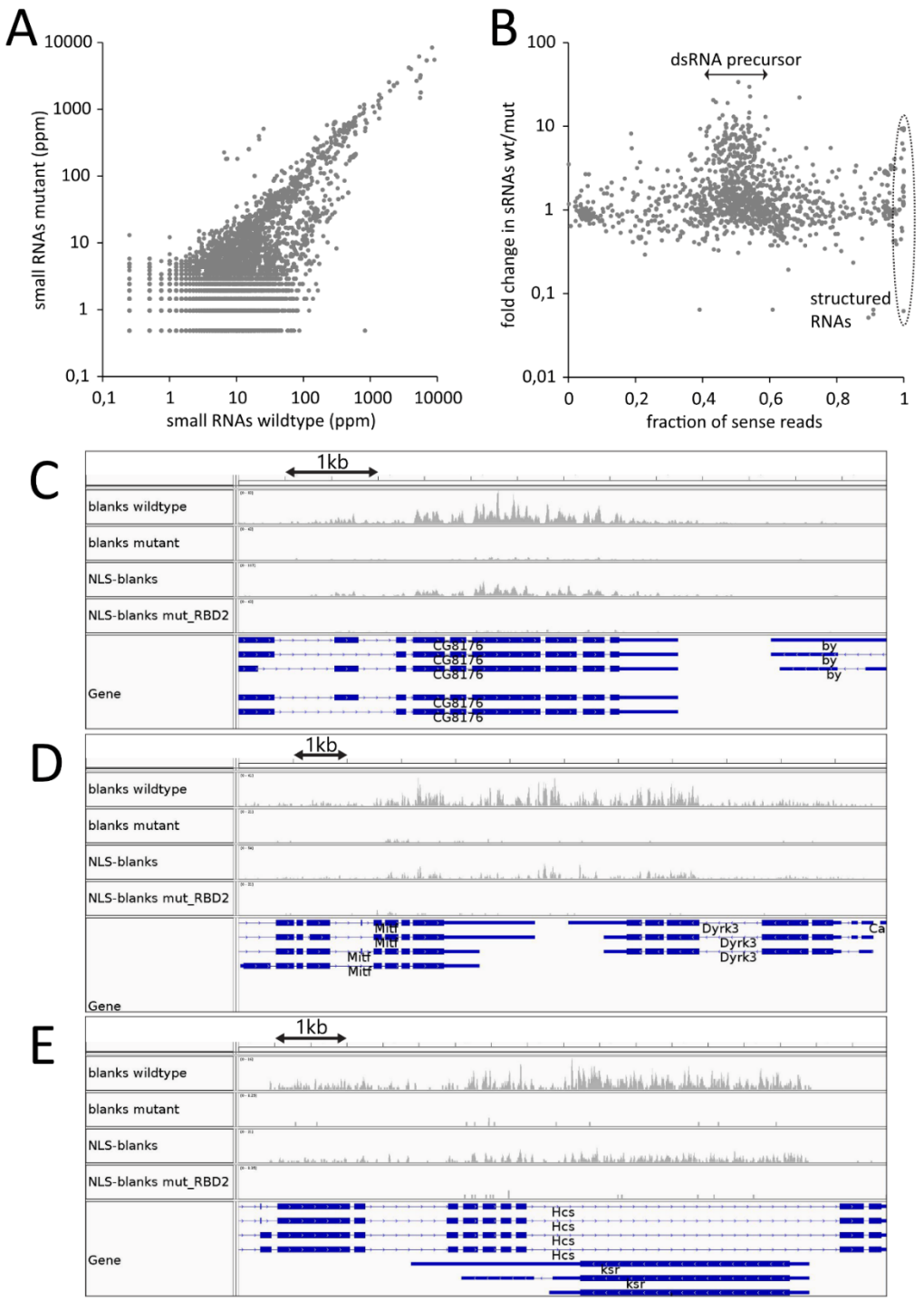

Figure 2.11 BepsiRNAs are present in the *in vivo* sequencing data from testes of the transgenic fly lines

(A) Plot comparing the 21 nt long reads that mapped to all *Drosophila* genes (extended with 150 nt at each end) between the mutant and the wildtype rescue fly line. Reads counts were normalized to genome matching reads.

(B) Comparison of the fold change (wildtype/mutant) of reads mapping to each gene (after cut-off of >10 reads was imposed) with the ratio of sense reads for each given gene.

(C-E) Coverage traces from the 21 nt long reads from the indicated libraries for three exemplary gene loci: *CG8176/by* (C), *Mitf/Dyrk3* (D) and *Hcs/ksr* (E). Track heights were adjusted to sequencing depths to allow for a semiquantitative visual assessment of read counts.

Figure was already published in Nitschko et al. (2020)

The library from the NLS-Blanks flies shows bepsiRNA reads similar to the wildtype library albeit at slightly lower levels. This can be explained by an increased nuclear ratio of Blanks conferred by the NLS in comparison to the wildtype which retains more dsRNAs in the nucleus preventing them from being processed by Dcr-2 in the cytoplasm.

2.2.9 Detection of the dsRNA precursors

The bepsiRNAs we detected via deep sequencing are a strong indicator that there are double-stranded precursors from which they are formed. Since a *Dcr-2* knock-out abolishes bepsiRNA formation these precursors are processed by it in the cytoplasm. The dependency on Blanks leads to our conclusion that it facilitates nuclear export of the precursors. We now wanted to detect the precursors directly via qPCR and designed corresponding primer pairs for the regions for which we got bepsiRNA reads in the deep sequencing experiment. Additionally, we wanted to affirm our hypothesis about the nuclear export via Blanks and decided to analyse the RNA levels separately in cytoplasm and nucleus. For this we established a fractionated lysis protocol to separate cytoplasmic and nucleoplasmic contents.

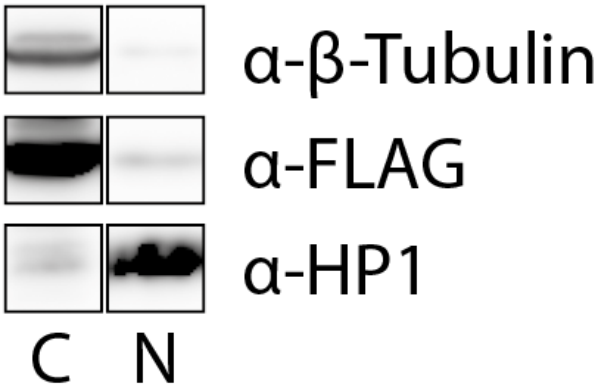

Figure 2.12 Western blot analysis of FLAG-Blanks cell fractionation
Cytoplasmic (C) and nuclear (N) lysates were analysed on a Western blot with antibodies against the nuclear histone protein HP1, the cytoplasmic β-Tubulin and against FLAG to detect Blanks localization.

The Western blot analysis in Figure 2.12 shows that the fractionation worked reasonably well. HP1 as a histone protein is almost exclusively found in our nuclear fraction and β-Tubulin is only in the cytosolic fraction as expected. However, Blanks is normally localized in nucleus as described by others and as can be seen in the microscopy images in Figure 2.2 it appears mainly in the cytosolic fraction. This might be due to an overexpression and resulting mis-localization from our inducible system. Since Blanks can still be found in the nucleus and can potentially still bind to dsRNA there, we decided to continue with our experiment by isolating RNA from both the cytosolic and nuclear fractions. The RNA was reverse transcribed and the cDNA was analysed via qPCR with the primer sets mentioned above. We compared samples from cells with induced expression of *blanks* with

samples without *blanks* expression to determine the effect Blanks has on localization. However, since we are working with RNA levels in both cytoplasm and nucleoplasm there is no easy way of normalizing our data. We decided it might be the most insightful to compare the ratio of the RNAs between both fractions and then see if *blanks* expression has an influence on it. Indeed, we found for the three primer sets for bepsiRNA-loci that the nuclear/cytoplasmic ratio in cells with *blanks* expression was lower by a factor of 0,72 ± 0,06 in comparison to cells without *blanks* expression. This indicates that in presence of Blanks more of the RNA can be exported from the nucleus. However, as mentioned quantification of the RNA level was difficult meaning this number might not be completely accurate but it encouraged us to follow up on the direct detection of the precursors. We decided to examine the RNAs directly bound to Blanks.

During our studies the Kim lab published a protocol for a formaldehyde crosslinking CLIP protocol (fCLIP) which they used to study dsRNAs binding to Drosha (Kim and Kim (2019)). We adopted this protocol to study the RNAs directly bound to Blanks. Since we were still interested in the differences between cytoplasmic and nuclear bepsiRNA precursor levels we incorporated our fractionated lysis protocol into the fCLIP workflow.

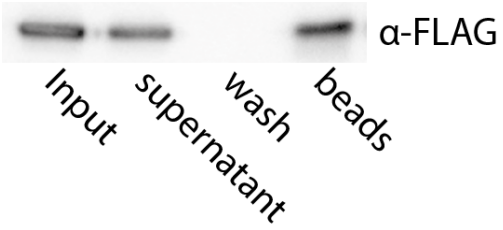

Figure 2.13 Western blot to analyse the fCLIP protocol from FLAG-Blanks cell line
Samples from several steps of the IP were loaded on a gel and detection during the Western blot was done with a monoclonal antibody against FLAG to detect Blanks

The IP worked reasonably well (Figure 2.13) and most of the FLAG-Blanks bound to the beads. The presence of unbound Blanks in the supernatant can be explained by the fact that formaldehyde was used for the crosslink which might have crosslinked some of the lysines in the FLAG-tag and thereby prevents binding to the anti-FLAG antibody. After the promising IP results, we isolated RNA from the bound fractions and compared the RNA profile of the cytoplasmic and nuclear fraction that bound to Blanks on a urea gel to see if we can already detect differences between the two fractions before qPCR analysis. Indeed, the band pattern between the two fractions differed strongly from each other (Figure 2.14) indicating that Blanks has different RNAs bound in the cytoplasm and nucleus respectively.

Figure 2.14 Urea gel with RNA isolated from Blanks-fCLIP from nuclear and cytoplasmic samples
The RNA bound to Blanks differs between the cytoplasm and nucleus. The nucleic acids were stained with SYBR™ Gold.

The subsequent qPCR analysis with primer sets for several bepsiRNA loci did not give any conclusive results. When comparing Ct values from RNA isolated from the FLAG-Blanks cell line they did not differ significantly from Ct values gained from RNA isolated from a cell line lacking a FLAG-tag. This means the RNA detected from the FLAG-Blanks cell lines are mostly unspecifically associated directly to the antibody or beads or to a protein that unspecifically binds to either of them. We suspected bepsiRNA-precursor levels to be low in the first place since they are likely products from unstable transcripts in the nucleus as discussed above. Export might also be fast giving us only a small amount of dsRNAs bound to Blanks before they are being processed by Dcr-2 after which they can no longer be detected via qPCR as the loci to be amplified are around 100 nt long.

2.3 Discussion

In summary, we were able to show that the *Drosophila* dsRBD containing protein Blanks interacts with double-stranded RNA, shuttles between the nucleus and the cytoplasm and that the biogenesis of a subset of endo-siRNAs (which we termed bepsiRNAs) depends on Blanks in both S2 cells as well as flies. The loci of origin of bepsiRNA are sites where two genes are convergently transcribed. The efficient formation of those bepsiRNAs was dependent on the RNA-binding capability and the ability of Blanks to shuttle between nucleus and cytoplasm unperturbed.

In previous studies *blanks* was described as an RNA silencing factor by other labs (Sanders and Smith (2011); Gerbasi et al. (2011); Zhou et al. (2008)) and a genome wide screen performed in our lab (Merk et al. (2017)) found it to be necessary for the efficient siRNA response after a DNA double-

strand break. However, none of these studies could propose a molecular mechanism for Blanks' function. We started the experiments on the molecular function of *blanks* with the hypothesis that due to the overall similar dsRBD domain structure to Dcr-2 interactors Loqs and R2D2, Blanks might act as an alternative factor to them. We, however, found no interaction with Dcr-2 (Nitschko et al. (2020)) and after we could show that Blanks also does not affect cytoplasmic RNAi (s. chapter 2.2.1) our focus switched to a potential role based on its reported nuclear localization (Gerbasi et al. (2011)).

We endogenously tagged Blanks with GFP and could reconfirm that Blanks is almost exclusively localized to the nucleus of S2 cells as has been reported previously. This observed localization is in line with the results of several web based NLS prediction tools like NLStradamus (Nguyen Ba et al. (2009)) showing an amino acid sequence preceding the first dsRBD domain that could serve as a NLS. Using Importazole as an inhibitor of β-importin proteins (Soderholm et al. (2011)) and thereby block most nuclear import pathways, we examined if Blanks can shuttle between nucleus and cytoplasm. Indeed, we observed a switch from a mainly nuclear localization to a cytoplasmic one of Blanks after cells were treated with Importazole (Figure 2.2). In contrast, the HP1 histone protein retained its nuclear localization after the cells were treated with Importazole. This means it is unlikely that the change in localization of Blanks was caused by the accumulation of newly translated proteins in the timeframe of the assay.

These results together with the results of a mass spectrometry experiment performed in the lab (Kunzelmann (2017)) that found nuclear export factors like Bj1 (Rcc), Kap-α3 and Ran to interact with Blanks led us to examine a potential role in nuclear export. We tried to corroborate the interactions with the export factors in Co-IP experiments. However, the results of these Co-IPs were not conclusive. We found an interaction of Blanks with Bj1, Ran and Exp-5 in one replicate which could not be replicated in two other Co-IPs for Bj1 and Exp-5 and in only one of two replicates for Ran. This does however not necessarily mean that these interactions don't exist, but could indicate that these interactions are transient or unstable. Another interpretation we cannot exclude is that Blanks might not directly interact with the aforementioned proteins but interaction might be indirect via another factor that bridges the interaction.

Our hypothesis at this point was that Blanks might serve as a nuclear export factor for double-stranded RNAs. In contrast to the miRNA biogenesis pathway for which it is known that Exp-5 is responsible for export of the hairpin precursors (Kim, Han, and Siomi (2009); Bohnsack, Czaplinski, and Gorlich (2004); Lund et al. (2004)), the nuclear export pathway for double-stranded siRNA precursors is as of yet unknown. For Blanks to act as an export factor for these dsRNAs it should

have the ability to bind dsRNA independently of their sequence via its dsRBDs. In this way it could serve as an adaptor between the dsRNAs and the Ran-GTP/GDP cycle that drives the nuclear export and import pathways. A closer look at the two domains that are annotated as dsRBDs in Blanks reveals that the dsRBD2 is the more likely candidate for RNA-binding. The amino-acid sequence of the annotated dsRBD1 has degenerated most prominently lacking the dual lysine residues which confer binding of dsRNA. This is also known for annotated dsRBD in other proteins like the dsRBD3 in Loqs, TRBP and PACT which act as dimerization or protein binding domains (Jakob et al. (2016)). We wanted to examine if the dsRBD2 of Blanks can bind dsRNA. It resembles the more canonical amino acid sequence of a dsRBD and shows high similarity to other dsRBDs in proteins like Loqs, R2D2, PACT and TRBP that are known to bind dsRNA. We expressed recombinant proteins with either the wildtype sequence or with a mutation in lysines 301 and 302 which according to structural similarities to other dsRBDs should be the amino acids conferring the dsRNA-binding. In an anisotropy measurement with a small fluorescently labelled dsRNA we could show that the wildtype protein indeed showed a fourfold higher affinity to the dsRNA ligand than the protein with the mutant sequence. The wildtype binding affinity (K_D=177 ± 22 nm) is in a comparable range to other proteins like Loqs that show dsRNA-binding via a single dsRBD (Tants et al. (2017)).

In line with the reported predominant expression in testes of *Drosophila*, *blanks* mutants previously showed a defect in male fertility via a spermatogenesis defect (Gerbasi et al. (2011); Sanders and Smith (2011)). This spermatogenesis defect was linked to mutations in the dsRBD1. The second dsRBD however showed no influence on male fertility. We wanted to investigate if this fertility defect is linked to the ability of Blanks to bind dsRNAs. For this we created transgenic rescue fly lines in a *blanks* mutant background. We then analysed fertility of males of these fly lines. As expected from previous studies the *blanks* knockout mutant line was completely infertile. When the mutant was rescued by the expression of wildtype Blanks, fertility was restored (Figure 2.6). In the transgenic lines in which we inserted an NLS in front of Blanks to force localization to the nucleus even more we saw a severe decrease in fertility in comparison to the wildtype protein. We attribute this to unspecific binding of dsRBD2 to structured RNAs and their retention in the nucleus. Structured RNAs have been shown to be influence male fertility in *Drosophila* (Wen et al. (2015)) and the perturbation of the system with our transgene is in line with this observation. The fertility defect caused by the NLS-Blanks could be reverted by mutating the dsRBD2 sequence of Blanks according to the mutations that were shown to cause a loss of dsRNA binding capability in the RNA-binding assay. With this mutation RNAs can no longer be bound by Blanks and the introduced mislocalization of structured RNAs is reverted. The restored fertility in the context of the dsRBD2-mutant also shows that *Drosophila* male fertility does not depend on the ability of Blanks to bind RNA.

As our hypothesis was that Blanks might act as an export factor for nuclear dsRNAs, we decided to analyse the sRNAs from testes of the transgenic fly lines and a cell line with an inducible expression of *blanks* via deep-sequencing. When we were analysing the read counts for mature miRNAs in the sequencing libraries from testes we found that they were strongly decreased in case of the NLS-Blanks line (Figure 2.7). This corroborates the hypothesis outlined in the discussion of the fertility assay. Structured RNAs seem to be held back unspecifically in the nucleus by NLS-Blanks, thereby preventing processing into the mature form via Dcr-1. According to the results of the fertility assay a mutation in the dsRBD2 preventing structured RNA-binding did not show this effect.

To form a dsRNA in the nucleus two transcripts that are antisense to each other have to be transcribed. This is most likely at loci with convergent transcription at which two genes are either in close proximity to each other or even partially or completely overlapping. The resulting convergent transcripts can then base pair with each other and form the dsRNAs. We investigated 21 nt long sRNAs in the sequencing data from cell lines in which we could regulate and even shut-off expression of either *blanks* or *Dcr-2* for the presence of processed siRNAs at loci with potential for convergent transcription. Since transcription does not always end with the annotated 3'-UTR, we added 300 nt to the 3' end of each gene for this analysis and checked for overlaps in opposing directions in the genome coordinates. After removal of redundancies we had a list of 4366 candidate loci for convergent transcription. Of these candidate loci 1748 showed a read count of >5 ppm which we used as a cut-off for further analysis. Most sRNAs at these loci showed dependency on *Dcr-2* expression (Figure 2.8A) indicating that the detected sRNAs are *bona fide* siRNAs. When comparing reads from the wildtype cell line with the cells in which *blanks* expression was shut-off, a population of 528 loci displayed strongly decreased read counts (Figure 2.8B). When *blanks* expression was induced in the same cell clone sRNA levels for these loci were restored close to wildtype levels (Figure 2.8C). This shows that among loci with convergent transcription that form siRNAs in a *Dcr-2*-dependent manner there is a subpopulation that is also dependent on *blanks* expression. We named the sRNAs from these loci bepsiRNAs for **B**lanks **e**xported **p**recursors siRNAs. The coverage traces depicted in Figure 2.9 show the described observations for an exemplary locus. The visualization of single reads also shows that reads can be detected in sense and antisense direction in similar amounts further corroborating that these are indeed siRNAs. We investigated the features of the bepsiRNA loci further and were able to show that transposon-targeting endo-siRNAs do not originate from bepsiRNA loci (Figure 2.10A). Additionally, we could show that like regular siRNAs bepsiRNAs are also loaded into AGO2 thereby likely functional (Figure 2.10B).

An additional line of analysis was performed for the sequencing data from sRNA from testes of wildtype, the transgenic rescue and the knock-out mutant fly lines. The 21 nt long reads were

mapped to all annotated gene loci which were extended by 150 nt at either end with the rationale to capture siRNA reads originating from converging genes that do not overlap but still are in close proximity. When comparing the data from the mutant and wildtype libraries, a population of siRNA generating loci was found that was dependent on Blanks as was seen before in the cell line sequencing data (Figure 2.11A). This shows that the bepsiRNAs not only exist in cell culture but also *in vivo* in the organ of predominant expression of *blanks*. We queried the sRNA data further and saw that the loci with the biggest change in siRNAs levels between the mutant and the wildtype correlated with loci that gave rise to similar amounts of sense and antisense reads (Figure 2.11B). The occurrence of an equal distribution of sense and antisense reads results from the double-stranded character of the siRNA precursors and thereby the mature siRNAs themself. We had to restrict the analysis at this point for loci for which more than 5 reads were detectable per locus (arbitrary cut-off) to allow for a semiquantitative statement about the fold change. It has to be noted that many loci for which close to or no reads could be detected in the *blanks* mutant but gave rise to large amounts in the wildtype case thereby were excluded from the analysis. These loci, however, are likely among the ones most strongly affected by Blanks.

Additionally, we saw that bepsiRNA reads did not only cover exonic regions or 3'-UTRs but also originate from regions downstream of the 3'-UTR and even introns (Figure 2.11C-E). This suggests that the dsRNA precursors that give rise to bepsiRNAs already form shortly after transcription and before the transcript can undergo splicing or 3'-processing. Although we can detect bepsiRNA reads that span the exonic-intronic boundary, we can also not exclude the possibility that dsRNAs can also form from processed RNA fragments after removal from the mRNA and before these fragments can be degraded.

As can be seen in the three examples for bepsiRNA loci in Figure 2.11C-E it is also necessary for Blanks to have an intact dsRBD2 for the successful formation of bepsiRNAs. With a mutated dsRBD2 bepsiRNA levels are close to mutant levels. In contrast, the alteration of Blanks localization via introduction of an additional NLS has only a small effect on bepsiRNA levels which are slightly reduced in this case which could mean that export is perturbed but not completely abolished by addition of the NLS to Blanks.

The deep-sequencing results showed us that siRNAs are generated dependent on Blanks. Since mature siRNA are processed from a double-stranded precursor - which we think is exported from the nucleus by Blanks - we tried to detect and quantify the amounts of longer dsRNAs originating from bepsiRNA loci. Our efforts to quantify these dsRNAs in nuclear and cytoplasmic fractions of cells with and without *blanks* expression hinted at a reduction in the nuclear/cytoplasmic ratio of

longer dsRNAs when *blanks* is expressed. This could cautiously be interpreted as a more efficient export in the presence of Blanks. Since quantification of the data was difficult, this conclusion is not completely resilient to scrutiny. We tried to crosslink and immunoprecipitate the longer dsRNA bound to Blanks directly via fCLIP from both nuclear and cytoplasmic fractions. However, the results were inconclusive when the qPCR signals were compared to those of a control cell line which did not allow for immunoprecipitation of Blanks.

All in all, our model is that Blanks acts a nuclear export factor for double-stranded RNAs originating from genomic loci with convergent transcription (Figure 2.15). For that the -likely nascent-transcripts form dsRNAs which are then bound by the dsRBD2 of Blanks. They are then transported across the nuclear membrane via the NPCs while bound to Blanks. It remains to be seen if Blanks facilitates this nuclear export on its own or in complex to one or more other export factors. In the cytoplasm the dsRNAs are then released by Blanks, processed by Dcr-2 and loaded into AGO2 where they become fully functional parts of RNAi. Due to the fact that we and others could not detect an

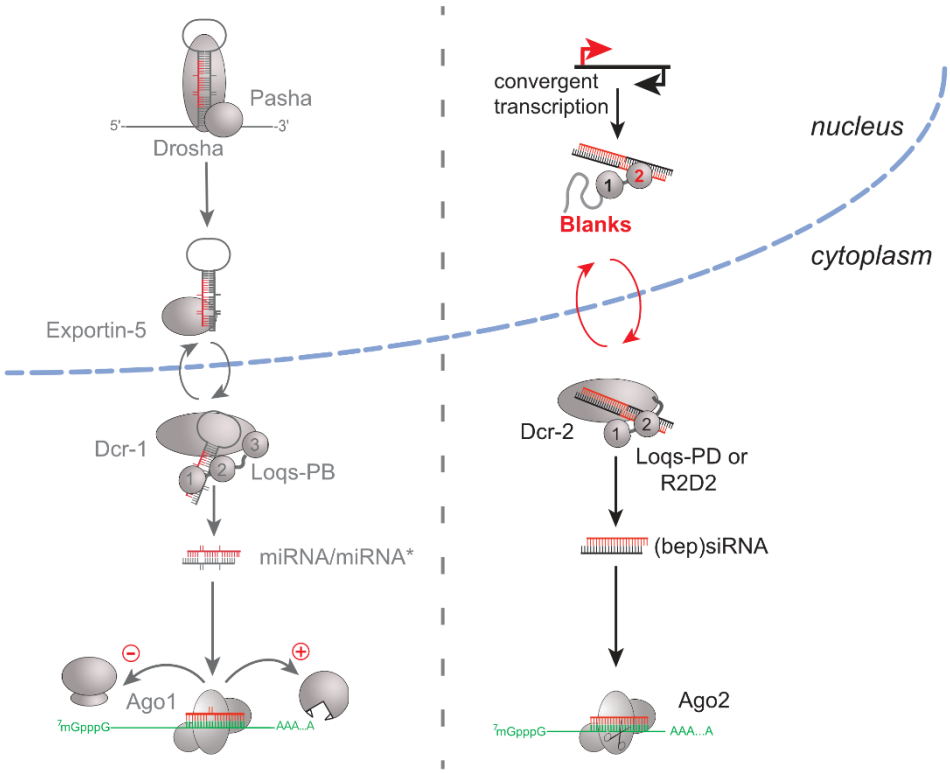

Figure 2.15 Model of nuclear export of dsRNA bepsiRNA precursors by Blanks
The established export pathway for miRNA precursor hairpins and the model for Blanks facilitated export of dsRNAs originating from convergent transcripts are depicted. The dsRBDs of the involved RBPs are represented as spheres and numbered according to their position in the amino acid sequence.
Figure was already published in Nitschko et al. (2020)

interaction between Blanks and Dcr-2, a direct hand-over of the dsRNA between Blanks and Dcr-2 is not likely.

2.4 Outlook

One question that is still not addressed after our studies is if Blanks can export dsRNAs on its own or if interactions with other proteins are needed. The fact that interactions with importin family proteins and Ran could be shown in the MS-experiments that were performed in the past could be an indicator that Blanks needs partners for nuclear export of dsRNAs. However, we were not able to conclusively validate these interactions in the Co-IP experiments that were performed for this study. It is plausible that Blanks might act as an adaptor between the dsRNA and one or more already known nuclear export proteins e.g. Exportin-5. The dsRBD containing protein NF90 that harbours some structural similarity to Blanks has been shown to interact with Exportin-5 in human cells (Gwizdek et al. (2004)). The interaction between the two proteins was facilitated by an RNA and increased the affinity of both proteins to the RNA in the complex. It remains for future studies to investigate these possible interactions further. One possibility would be to use the recombinantly expressed GST-Blanks fusion protein used for the RNA-binding assay and binding it to glutathione beads. After binding a concentrated lysate from cells with endogenously tagged potential binders could be added and incubated with the bound Blanks. If the tagged proteins bind to Blanks they could potentially be detected via Western Blot even if the interaction to Blanks is weak. A different approach would be to endogenously tag Blanks with a biotin ligase. After addition of biotin to the cells proteins that get into close proximity of Blanks will be biotinylated and can be immunoprecipitated with avidin beads.

Additionally, we were not able to directly detect binding of Blanks to bepsiRNAs precursors. It would clearly be of benefit to support our overall model of Blanks as an dsRNA export factor if the interaction between Blanks and the dsRNA could be shown *in vivo*. However, more efforts in this direction exceeding the fCLIP experiments were beyond the scope of the practical work this thesis is based on.

The described structural homology and observed functional similarities with NF90/ILF3 could be the basis for further studies of *blanks*. NF90 has been studied far more extensively than *blanks* and although homology between the two proteins might be limited more functional similarities might exist than reported here. For example, NF90 has been implicated to act in the biogenesis of circular RNAs (Li et al. (2017)). Although the field of circRNA research has gained traction over the last years, our knowledge about many biogenesis steps and the function (especially in *Drosophila*) of these RNAs is still lacking. It is therefore conceivable that Blanks as an RBP might be an actor in circRNA

biogenesis in *Drosophila*. A starting point to address this possible involvement might be the translated circRNA database published by Ashwal-Fluss et al. (2014). It provides sequences that for a start can be used to generate qPCR primers specific for circRNA and mRNA of the same gene. In combination with RNA isolated from the transgenic flies used in this study insights into a possible interaction of Blanks with circRNAs might be addressed.

3 Investigating the role of Rfc-complex factors in siRNA biogenesis at the DSB via ChIP-qPCR

3.1 Introduction and aim of the project

As described in 1.2 a genome-wide screen in *Drosophila* performed in our lab revealed a multitude of factors that influence the diRNA generation near a DSB. One surprising group of proteins that were shown to be a positive regulator for siRNA biogenesis at the DSB in this screen, were all five components (Gnf1, RfC3, RfC4, RfC38 and CG8142) of the replication factor C complex (RFC). The RFC-complex is responsible for loading of PCNA to the replication fork during DNA replication (Burgers and Kunkel (2017)). A modified version of the RFC-complex in which Gnf1 is replaced with Elg1 has been shown to unload PCNA again from the chromatin (Shemesh et al. (2017)). The RFC-complex thereby plays an important role facilitating correct DNA replication and protecting genome stability. However, it is not straightforward to draw a connection between the role of the RFC-complex in replication and the involvement in siRNA biogenesis at the DSB that was observed in the screen. We therefore wanted to further study involvement of the RFC-complex in the siRNA biogenesis after a DSB. This would expedite our understanding of the molecular mechanism of the siRNA biogenesis near a DSB in general, elucidate potential other functions of the RFC-complex and give us further insights into the molecular functions of diRNAs.

3.2 Experimental setup

The experiments described in this thesis are a direct continuation of the ones performed for my Master thesis (Nitschko (2016)). We are using chromatin-immunoprecipitation (ChIP) to detect proteins that are close to a DSB. The general protocol for the ChIP has already been established and tested with success with proteins like RNA polymerase II and III subunits. However, initial experiments with some RFC-complex members showed that there is need to further improve the protocol since the interaction with DNA might be less pronounced or stable and only gave inconsistent results.

In one of the first steps in a ChIP-experiment formaldehyde is added to cells to induce crosslinks between proteins and DNA that are in close proximity to each other. If the RFC-complex is located at or recruited to a DSB it will be crosslinked to the DNA next to the break and thereby covalently bound. The DNA is then sheared during cell lysis to reduce the length of DNA bound to the proteins. Subsequently, an immunoprecipitation (IP) of the proteins is performed.

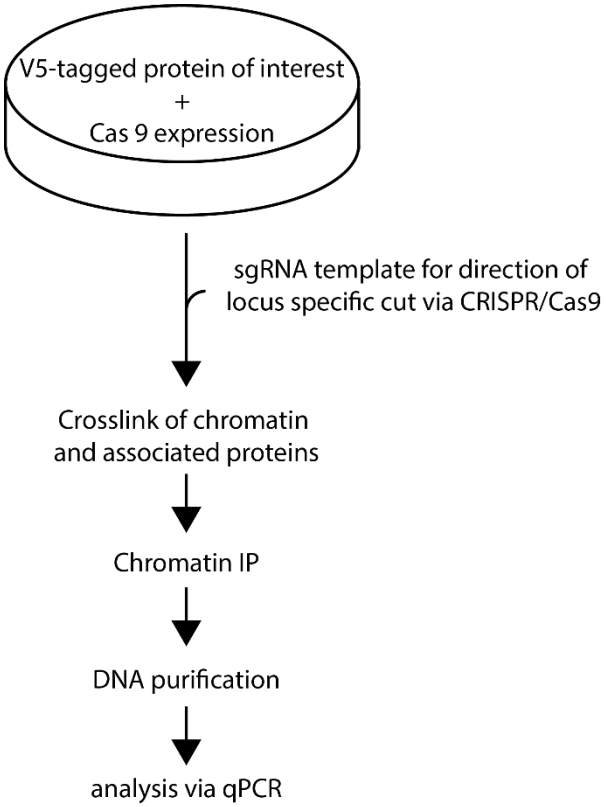

Figure 3.1 Schematic workflow of the ChIP-qPCR experiment
The main steps of the experimental setup are shown as a flow-chart.

For this study the RFC-complex proteins that were to be studied carried a V5-tag. In the experiments performed for the Master thesis the V5-tag performed better in comparison to the also tested FLAG-tag. This is likely due to the fact that it does not contain any lysine residues which could be crosslinked during the formaldehyde treatment and thereby block the interaction between the epitope tag and the antibody. One additional benefit of our V5-fusion proteins is a TEV-cleavage site in between the tag and the protein which can potentially be used for elution from the beads. After the IP the crosslink is reversed and the protein-bound DNA is thereby released. The DNA is purified and analysed via qPCR as described below. A schematic of the experimental workflow is shown in Figure 3.1.

Our hypothesis is that the RFC-complex is recruited to the DSB and then facilitates siRNA biogenesis. To test this via ChIP-qPCR we had two experimental conditions that were compared. Firstly, a sample in which we induce a DSB at a defined locus in the genome via our labs CRISPR/Cas9-protocol (Bottcher et al. (2014); Kunzelmann et al. (2016)). Secondly, a sample in which we did not induce a DSB as a negative control.

The loci for the DSBs were chosen from our previous knowledge from the genome-wide screen (Merk et al. (2017)). The *CG15098* locus was chosen as a locus which produces large amounts of siRNAs after a DSB and the *Tctp* locus as a control which in contrast does not produce siRNAs. The read-out with which we wanted to quantify the amount of binding of RFC-complex members to the DNA was the percentage of DNA recovered after ChIP in comparison to the DNA from an input sample. This is the so called %-Input value. This value increases when a protein is recruited to the DNA, crosslinked and copurifies together with the associated DNA. We quantified the amount of DNA from specific loci in the ChIP and input samples via qPCR. Two primer pairs were located close to the site at which a DSB could be induced; one upstream and one downstream of the DSB site (see Figure 3.2). An additional primer pair was used to quantify the amount of DNA at a control locus that

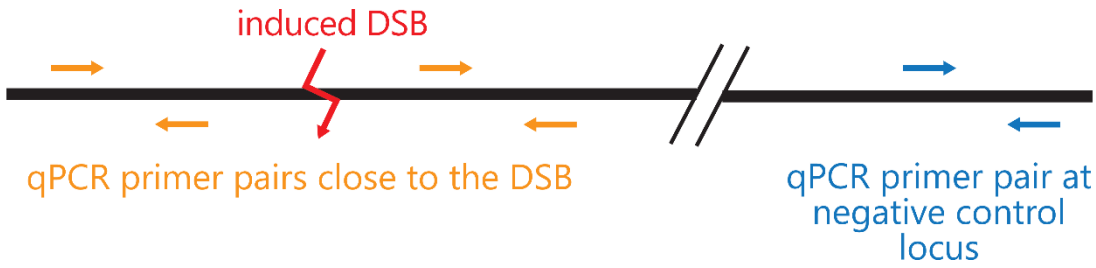

Figure 3.2 qPCR primer locations to detect specific recruitment of proteins to the site of an induced DSB
A schematic representation of the genome is shown. The site of the CRISPR/Cas9-induced DSB is shown in relationship to the specific qPCR primer locations near the DSB (orange). The negative control locus is chosen to be far away from the DSB site or even on another chromosome. Distances and primer lengths are not depicted to scale.

is not in the proximity of the DSB site. If recruitment of the RFC-complex members increases after a DSB is induced the %-Input (also referred to as recovery in the following) value should be increased in these samples in comparison to the samples without DSB. Additionally, the specificity of the ChIP experiment can be evaluated by comparing the %-Input values between the loci close to the DSB and the control locus.

3.3 Evaluation of the performance of the ChIP-protocol from previous work

During the initial experiments performed for my Master thesis (Nitschko (2016)), we validated the performance of the ChIP-protocol using CG17209 – the largest subunit of Pol III – as the bait protein for the IP. Since Pol III transcribes tRNAs we knew that we should be able to detect specific recruitment to a tRNA locus. Indeed, we got a %-Input value of 7,33 % for a tRNA gene in contrast to 0,04 % at the *Act5C* control locus (~183 times specific enrichment). We concluded that we had a solid protocol to detect recruitment of proteins to specific gene regions. Further experiments with proteins of the RFC-complex or PCNA showed that we were able to increase specificity by switching from a FLAG- to a V5-tag which also allowed elution via a TEV-cleavage. This however also reduced

the amount of recovered DNA and lowered the %-Input values by roughly one order of magnitude. Nevertheless, we chose to work with the V5-tag which gave us better specificity and tried to optimize our protocol in regards to recovery for more stable and reliable results.

3.4 Results

3.4.1 Preclearing with agarose beads strongly reduces recovery

As an initial step to evaluate future changes to the ChIP-protocol and their effect on recovery, we performed ChIP with V5-tagged RpII215 and CG17209 – the biggest subunits of Pol II and Pol III respectively. Since the sites of transcription of these two polymerases are generally well known, we chose them as a model system to improve our ChIP protocol. A tRNA- (for CG17209) and the *Act5C*-locus (for RpII215) were chosen as specific recruitment sites for the two polymerases. Since the two polymerases do not transcribe the respective other locus, it can serve as a control in the experiment. For RpII215 we had a %-Input value of 5,7 at the *Act5C*-locus (~45-fold enrichment when compared to the tRNA locus) and for CG17209 the %-Input value was 11,0 (~45-fold enrichment when compared to *Act5C*). These values on one hand showed that the previous results could be replicated and gave us a solid base line with reasonably good recovery and enrichment values.

Many ChIP protocols include a step that is known as preclearing which was not part of our protocol. For preclearing the samples are incubated with beads (we chose Pierce® Biotin Agarose beads (ThermoScientific)) that lack a binding site for the epitope-tagged protein before the immunoprecipitation step. The rationale behind this step is that proteins or DNA that unspecifically associate with the beads, and thereby would bind to them during the immunoprecipitation, are already removed before the IP is performed.

The agarose beads were equilibrated by washing them twice with the RIPA buffer (see methods) before the lysate was incubated with them for 1 hour at 4 °C under rolling. After the incubation the beads were removed by centrifugation and the precleared lysate was used for the IP as described in 4.1.9.

In all three replicates that were performed after the lysate had undergone preclearing, the %-Input values decreased by up to three orders of magnitude showing that the recovery was severely affected by the preclearing step. Additionally, the enrichment at the specific locus did not differ from the experiments that were previously performed without preclearing.

3.4.2 Elution by TEV-cleavage is inefficient

One question we had not addressed at this point was the cleavage efficiency of the TEV protease which we used for elution of the bait protein from the beads after IP. The TEV-cleavage site in our

system is located between the epitope tag and the protein. If this site would be inaccessible for the protease or the cleavage itself is not efficient this would decrease the recovery in our experiments. To evaluate how much of the bait protein (and crosslinked DNA) stays on the beads after the TEV-treatment, we incubated them with Proteinase K after the TEV-cleavage was performed in two replicates of the preclearing experiments. DNA purification was performed as described in 4.1.9. When we now calculated the %-Input values of the Proteinase K treated beads we saw that they were higher by a factor of 7 to 10 (CG17209: 0,37 % vs. 0,03 %; RpII215: 0,041 % vs. 0,006 %) at the loci specific for the investigated polymerase subunit. This increase in recovery was striking especially since beads were already eluted by TEV-cleavage and still apparently had most of the protein and associated DNA bound that then was released by Proteinase K digestion. In contrast, the enrichment to the specific gene locus was not differing much between the two conditions. We validated these results by looking at a second specific locus for both polymerase subunits (*snRNA:7SK* for CG17209 and *RpL32* for RpII215) which gave comparable results. We decided remove the TEV-cleavage for all future ChIP experiments due to these results and the protocol changed from what was described in Nitschko (2016) to the protocol in this thesis that can be found in 4.1.9. This had the additional benefit of removing one overnight incubation step potentially reducing degradation of the proteins by residual proteases.

3.4.3 Effects of high molar concentrations of urea on the ChIP experiments

It has been reported that addition of 6 M urea to the lysate and subsequent dialysis before the IP increases the signal after a ChIP (Park et al. (2001); Chromatin immunoprecipitation assay with urea denaturation. Available at: http://www.personal.psu.edu/faculty/d/s/dsg11/labmanual/Chromatin_structure/ChIP_for_Drosophila_cells_preferred.html (accessed on 22.2.2017)). We wanted to test if we see the same improvement in our experiment. After lysis and centrifugation, we added an equal amount of 6 M urea to the lysate and transferred it to a Spectra/Por® 4 (Roth) dialysis membrane. Dialysis was done overnight at 4 °C against 500 mL RIPA buffer that was exchanged twice after 2 and 4 hours with fresh buffer. If compared to the non-dialysed samples we saw that recovery after the CG17209 IP at the specific locus was slightly lower after urea treatment and dialysis (2,5 vs. 4,7 %). The specific enrichment was significantly higher after dialysis (300-fold vs. 8-fold). However, we have seen in our previous experiments that recovery for the RFC-complex factors was quite low. Since there was no improvement to recovery, we chose not to include the urea treatment and dialysis to the ChIP protocol. However, if we succeed in increasing the recovery by other means we still might benefit from the enrichment increase by the treatment. Additionally, by omission of the dialysis the protocol requires one less overnight incubation.

We also tried to simply add increasing amounts of urea to the lysates without dialysing the samples. Our observation was that recovery was slightly decreased (5,4 % vs. 10,7 %) up to concentration of 2 M urea and enrichment to the specific locus was barely affected (8-fold vs. 11-fold). A concentration of 3 M urea decreased both recovery and enrichment more.

3.4.4 The CRISPR/Cas9 induced does not allow for detection of DSB recruited factors

After we could improve the recovery in the ChIP protocol via elution with Proteinase K instead of TEV-cleavage, we wanted to investigate if this change is sufficient to detect the recruitment of the RFC-complex to the DSB. We transfected sgRNA templates to induce a cut at either *CG15098* or *Tctp* in cells with V5-tagged RfC4. An additional sample without the induction of a cut was prepared as a control. For the qPCR analysis after the ChIP we used five primer pairs: The primer pairs were located either 5' or 3' of either *CG15098* or *Tctp* (~50 nt distance to the site at which the DSB is introduced). The two primer sets near the location of one of the intended DSB sites can be used as a negative control when the DSB is introduced in the other gene and vice versa. The last primer pair was located in *RpL32* and was used for normalization.

The %-Input values received from the loci near the *CG15098*- and *Tctp*-cut sites were normalized to the corresponding RpL32 values (transfection or omission thereof of the same sgRNA template). The second round of normalization was performed with the values from the samples in which no DSB was induced.

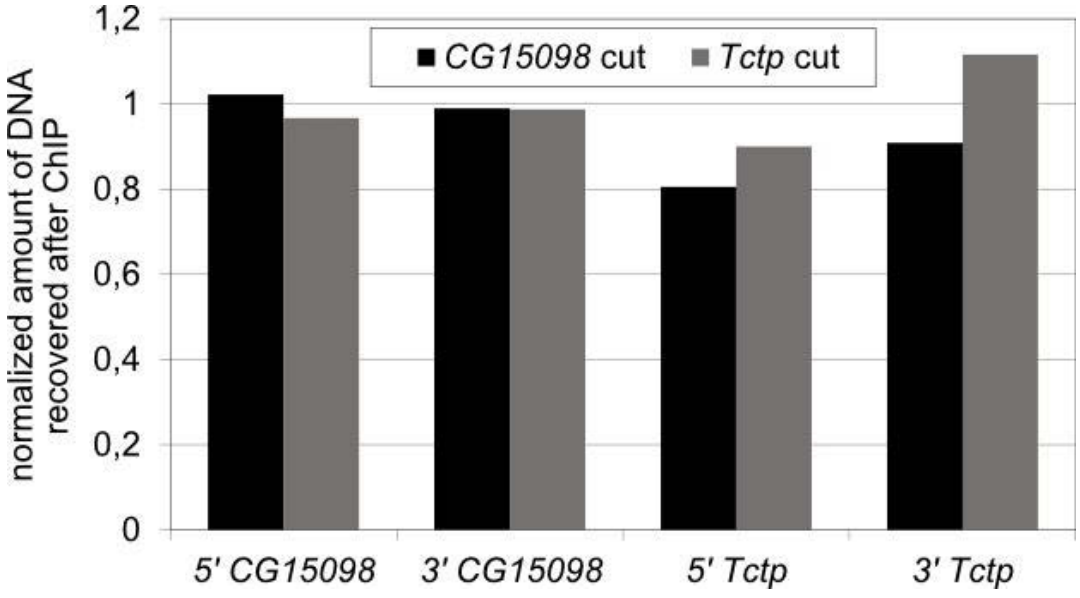

Figure 3.3 Induction of a DSB via CRISPR/Cas9 does not lead to recruitment of RfC4
A DSB was induced either in *CG15098* (black) or *Tctp* (grey). The %-Input values for the four loci near *CG15098* or *Tctp* were normalized to the RpL32 values and the control without a DSB. If RfC4 would be recruited to the DSB a noticeable increase in signal near the cut site would have been expected. This graph is exemplary for multiple experiments with different RfC-complex factors and IrbP (Ku70).

As can been seen in Figure 3.3 the signal from each of the four investigated genomic sites is similar in both the *CG15098* and the *Tctp* cut condition. The data thereby suggests that an induced DSB does not lead to the specific recruitment of RfC4. We performed the same experiment with Gnf1 and Irbp (Ku70) – a known DSB associated protein – in two replicates with similar results. We know from previous work that the induction of the DSB does only work in a small percentage of cells (Bottcher et al. (2014)) which might not be enough to result in a specific signal in this experimental setup. Our conclusion was that the ratio of cells in which a DSB is successfully induced is not big enough to give a signal that raises over background levels.

3.4.5 Specific enrichment at linearized plasmids is inconsistent

Since we suspected that the amount of DSB created by the CRISPR/Cas9 system was not enough to raise the signal above background levels, we wanted to investigate Irbp and RFC-complex factor recruitment to a linearized plasmid. A linearized plasmid is a good DSB model system since the cell cannot distinguish it from a regular DSB, and large amounts can be transfected to have a high level of DSBs to which the protein can potentially be recruited. We linearized pKF63 near the 5'-end of the GFP coding sequence by restriction with BamHI and transfected it into S2 cells. From these cells we performed both Irbp- and Gnf1-ChIPs. For qPCR-analysis we used a GFP primer pair near the linearization site and a RpL32 primer pair as a control.

For Irbp we performed four replicates of the ChIP experiment and saw more recovery of DNA from the GFP gene in comparison to the control locus in two of them. For the other two replicates the control locus showed a considerably higher signal. Additionally, in three out of the four replicates we got recovery values of GFP DNA of up to 80 % in cells which were not transfected with linearized plasmids and thereby should not contain any GFP sequence. These results show that even in the background of higher DSB level it is not possible to reliably detect specific enrichment of known DSB associated factors via our ChIP method. We tried one replicate with Gnf1 as the bait protein with similar results as for Irbp.

3.5 Discussion

The main two goals of the experiments were to improve the amount of specific recovered DNA after ChIP by altering the protocol and then to use the improved version to investigate the RFC-complex recruitment to a DSB site.

We succeeded to recovery overall by changing from a TEV-elution to digestion of all proteins via Proteinase K without affecting specificity. The TEV-cleavage was inefficient which might have several reasons. The reaction conditions might have been needed to be improved for a higher activity of the protease. However, it is also conceivable that the cleavage site might not have been accessible to

the protease when the fusion protein was bound to the beads. The fact that parts of the fusion protein were crosslinked to other proteins or DNA might have amplified this effect. Since we got reasonably good recovery via the Proteinase K digestion, we did not pursue an improvement of the TEV-cleavage activity. The other alterations to the ChIP protocol that were tested did not result in a increase in recovery or even decreased it. For the attempts with preclearing of the lysate with agarose beads recovery decreased severely. The experiments with the addition of urea (and subsequent dialysis) showed that recovery was slightly reduced in comparison to the standard conditions. We chose not to include urea addition and dialysis into the protocol since improvement of the low recovery was our main priority. However, it has to be said that the dialysis experiment showed a significantly better specific enrichment.

With the slightly altered ChIP protocol we now tried to detect Irbp and RFC-complex recruitment to the DSB. In the experiments with the CRISPR/Cas9 induced DSB we saw that recruitment was not altered by the DSB. This was likely caused by the fact that the signal did not exceed background levels. We knew from previous experience with our CRISPR/Cas9 system that a DSB is only induced in a subset of cells. If the proteins are recruited to the DSB this signal might be masked by the large amount of protein from cells in which no DSB was induced. We tried to increase the amount of available DSBs by transfecting a linearized plasmid. However, even under these conditions we were not able to reliably detect recruitment to the DSB, even of a known DSB associated factor like Irbp. One other caveat of both the CRISPR/Cas9 induced DSB and the linearized plasmid systems is that they are transfection based. Although we usually get reasonably high transfection rates ranging from 30 to 70 % in other experiments, fluctuations in transfection rates between experiments and the signal from untransfected cells might cause the inconsistent results or weak signals. We chose to pause experiments at this point to think about ways to either improve the ChIP protocol significantly or to assess other methods of investigating RFC-complex recruitment to the sites of DSBs.

3.6 Outlook

Due to the low and varying transfection efficiencies a large number of cells in the experiment do not receive the sgRNA template that is necessary for the induction of the DSB. This means that most cells can give no signal for technical reasons which in turn decreases that amount of specific signal in the qPCR read-out. The lower the transfection efficiency is in the experiment the more likely it is that the signal is covered by unspecifically bound DNA or DNA crosslinked to unspecifically bound proteins. Additionally, differences in transfection efficiencies between replicates and experiment make it hard to normalize the data and make comparisons between experiments. One way to make the induction of DSBs more reproducible would be to create a (ideally clonal) Cas9 expressing cell line with a transgenic inducible sgRNA template. This way no transfection is necessary and DSBs

should be induced at a more consistent and higher rate potentially increasing the signal to noise ratio in the experiment. However, any cell line in which both the Cas9 protein and the programming sgRNA are stably integrated are not ideal experimental systems from a biological safety perspective, especially if the promotors are not tightly regulated and leaky.

Another way to explore the hypothesis that the RFC-complex is recruited to DSBs would be by using a proximity labelling approach. This way interactions to other proteins and not only to DNA could be explored if the RFC-complex is not directly recruited to the DNA but to proteins in the vicinity of the DSB. We now have the ability to used our CRISPR/Cas9 system to add a biotin ligase to a protein of choice instead of an epitope tag. When the cells are then treated with biotin the proteins carrying the biotin ligases can biotinylate proteins in their vicinity. Proteins that are known to be located at the DSB like Irbp (Ku70), Ku80 or spn-A are ideal targets for the addition of the biotin ligase. When this tagging is done in cells that already have epitope tagged proteins of interest (like the V5-tagged RFC-complex proteins) an immunoprecipitation with Streptavidin or Streptactin beads will be performed. The epitope tagged protein of interest can then be detected via a Western blot. A change in the amount of biotinylated and thereby isolated protein might be visible between samples with and without an induction of DSBs. Since this proximity-based approach is also not limited by the necessity to have a defined site of the DSB, DSBs can be chemically induced on a broader scale by treatment with zeocin or camptothecin. The increased amount of DSBs could allow for the detection of even weak and transient interactions between the biotin ligase-labelled protein and its interactors.

One conceivable caveat of the qPCR analysis after the ChIP might be the primer locations. We chose primer pairs to be roughly 50 nt up- or downstream of the DSB site. This seemed to be a reasonable compromise between being close to and not directly at the DSB. However, if the proteins bind further away than roughly 500 nt (the average length of the DNA fragments after shearing) from the DSB, we would not be able to detect them. This could be overcome by analysing the DNA recovered after ChIP by deep sequencing. Although this might help to detect binding of the RFC-complex members to sites not directly next to the DSB, there could be the same problem with signals not exceeding the background levels as with the qPCR.

Additionally, one could experiment with ways to select for cells with a DSB. A fluorescence DSB-reporter system for which fluorescence is either induced or eliminated after a repair event could be used to select for those cells via flow cytometry and cell sorting. The sorted subpopulation of cells should then contain only cells with a DSB and background levels in the ChIP could decrease since cells without a DSB are depleted and are thereby not giving signals.

4 Material and Methods

4.1 Molecular Biology

4.1.1 Generation of Blanks plasmid variants

The generation of the plasmids used for recombinant expression of Blanks with a wildtype and mutant dsRBD2 used for the RNA-binding assay and the plasmids used for generation of the transgenic flies were described in Nitschko et al. (2020) as follows:

"The plasmid backbone for the expression of the *blanks* variant transgenes in flies was pKF63 (Forstemann et al. 2007) that was modified by adding an attB site by oligonucleotide annealing and cloning the product into the *Nde*I site. The blasticidin resistance from plasmid pMH3 (Bottcher et al. 2014) was excised and cloned into the *Nde*I site in front of the attB-site. The resulting plasmid was digested with *Bam*HI and *Not*I and the GFP-insert was replaced with the Blanks CDS generated by PCR with *Drosophila* cDNA as a template. The cDNA sequence contains polymorphisms and corresponds to the sequence described with GenBank ID: AY119201.1. In comparison to the reference sequence, this sequence variant harbors an 8 aa deletion of D^{69}-R^{76}, a 4 aa deletion of V^{90}-D^{93} followed by an aa exchange (D^{94}→N), a 2 aa insertion of KE after K^{104}, three single aa exchanges (R^{150} → H, L^{190} → M, G^{194} → E) and two silent point mutations (A^{435} → G, A^{816} → G). […] The FLAG-NLS sequence and the mutations in the dsRBD2 were introduced via the PCR primers into the inserts of the corresponding plasmids.

The inserts for the recombinant expression vectors were created by PCR from the vectors created above with the '5' wildtype' and either the '3' wt-dsRBD2' or '3' mut-dsRBD2' primer. The primers used for the PCR were '5' wildtype' and '3' wt-dsRBD2'. The PCR products were cloned into the pGEX-6P-1 backbone using BamHI and NotI."

In addition to the plasmids described in the paper, recombinant expression vectors for expression of a 6xHis-TEV-*blanks*_mutdsRBD2 and an eXact-tag-*blanks*_mutdsRBD2 fusion protein were created using the same strategy as described above. The target plasmids were pET28a for the 6xHis-TEV vector (pVN7) and pPAL7 for the eXact-tag vector (pVN8).

For the FLAG-*blanks* plasmids that were used for injection to generate the transgenic flies for the paper additional vectors were created for this study. The plasmids with the wildtype dsRBD2 containing a FLAG-tag and a NLS-FLAG N-terminus were previously cloned in the lab by generating the inserts via PCR with the primers 5' FLAG and 5' FLAG-NES and 3' wt-dsRBD2 (s. primer table in 4.5) and cloning them into the BamHI/NotI digested plasmid backbone described above (pRB10) after digestion with the same restriction enzymes. For this study the versions with the mutant

dsRBD2 were generated using the same 5'-primers but using the 3' mut-dsRBD2 antisense primer (pVN3-5).

For the *in vivo* CRISPR tagging experiments the sgRNA template for Actin5C targeting was generated via PCR as described in Bottcher et al. (2014) and Kunzelmann et al. (2016). The PCR product was inserted into pJET1.2 via blunt-end cloning with the CloneJET™ PCR cloning kit (ThermoFisher Scientific) according to the manufacturer's protocol.

4.1.2 Recombinant expression and purification of Blanks protein

The recombinant expression of the Blanks proteins with a wildtype or mutant dsRBD2 were described in Nitschko et al. (2020) as follows:

"The expression plasmids were transformed in BL21 (DE3) pLysS cells. The expression culture was inoculated at OD600 = 0.1 in 1 l with ampicillin and chloramphenicol that was supplemented with 0.5 % glucose. The culture was grown at 25◦C until OD600 = 0.6 and then induced with 1 mM IPTG for 2 h. Cells were harvested, washed once in PBS and frozen in 8 ml lysis buffer (50mMTris pH 7.0, 150mM NaCl, 5 mM DTT, 10 µg/ml lysozyme, 0.1 U/ml DNase I, 1 % Triton and 1 tablet of protease inhibitor on 10 ml (complete mini, Roche). After thawing lysis was facilitated using a Bioruptor (Diagenode, 30 cyles: 30 s ON, 30 s OFF). The lysate was cleared by centrifugation and by passing it through a syringe filter before loading on a GSTrap HP column (GE Healthcare). The column was washed with 10 ml lysis buffer, 5 ml high salt buffer (50mMTris pH 7.0, 1,5M KAc pH 7.0, 5mM DTT) and again with 10 ml lysis buffer. Elution was done in 5 ml 50 mM Tris pH 8.0, 1,5 M, 5 mM DTT and 20mM reduced glutathione. The elution fractions containing the protein were pooled and incubated with 2 U of PreScission protease (GE Healthcare) over night at 4 °C to cleave off the GST-tag. The cleaved protein was diluted in 3 volumes of 50 mM HEPES pH 7.0 and loaded on a HiTrap SP HP ion exchange column (GE Healthcare). The column was washed with 10 ml 50 mM HEPES, pH 7.0 and eluted in 50 mM HEPES pH 7.0 with a gradient from 0 to 1000 mM NaCl. Fractions containing the protein were pooled and diluted in 3 volumes of 100 mM KAc pH 7.4, 10mM HEPES pH 7.4, 2mM MgAc, 5mM DTT and then concentrated using Amicon Ultra Centrifugal Filters with a 10 kDa cutoff."

4.1.3 Generation and analysis of sequencing libraries

The preparation of sequencing libraries and data analysis were already published in Nitschko et al. (2020) as follows:

"The […] deep sequencing library preparation were performed as previously described (Elmer et al. 2014) with the exception of using the ZR small RNA PAGE Recovery Kit (Zymo Research) for small RNA purification after the PAGE-steps. Sequencing was performed on an Illumina HiSeq instrument

at LAFUGA (Gene Center, LMU Munich, Germany). Sequencing reads were demultiplexed and 3'-adapter trimmed with custom scripts (available on request). The reads for the analysis of the miRNA-, transposon and endo-siRNA-mapping sequences were size selected for 20–23 nt long reads and mapped with bowtie.

To create the list of loci with the potential for convergent transcription, we first extracted all gene coordinates from the *Drosophila* genome annotation file (version 6.02 .gff, downloaded from Flybase) with linux command line tools (grep -w 'FlyBase' dmel-all-no-analysisr6.02. gff | grep -w 'gene' | grep -v 'parent type' |cut - f1,4,5,7,9 >gene coordinates r6 02.bed), then created a list of overlapping genes with opposite orientation extended by 300 nt on the 3'-end using bedtools window (Quinlan (2014)) (bedtools window -l 0 -r 300 -sw -Sm -a gene coordinates r6 02.bed -b gene coordinates r6 02.bed > overlapping 3p300 extended genes r6 02.bed). This list contained two entries for every potential overlap (one from the sense and one from the antisense-running gene), we thus generated a non-redundant set by restricting the orientation of the first gene to sense only. Finally, we simplified the name field to only the FBgn number with a custom Perl script.

The sequencing libraries were first size-selected to 21-mers and then filtered by mapping to the *Drosophila* transposon consensus sequences (no mismatch allowed), retaining only the non-matching reads. This dataset was then mapped to the *Drosophila* genome (version 6.02) with no mismatch allowed, only reporting reads that map uniquely. The overlap of this analysis with the regions of convergent transcription (see above) was determined by applying bedtools intersect with the –c option. We normalized differences in sequencing depth by calculating the ppm values relative to all genome matching reads in the filtered dataset.

The analysis of the bepsiRNAs in the testes samples was performed by first selecting only 21 nt long reads and removing the transposon-matching reads as described above. The remaining reads were mapped to all extended gene regions (2 kb extended precomputed set based on release 6.02, downloaded from Flybase and reduced to 150 nt on each side) using bowtie. The total number of *Drosophila* genome matching reads was used to normalize for differences in sequencing depth between the libraries in all cases. The sequencing data has been deposited at the European Nucleotide Archive (ENA) under accession number PRJEB32123."

4.1.4 RNA binding assay

The 23 nt long dsRNA (21 bp and a 2 nt overhang at the 3' ends) annealing out of synthetic oligos and the RNA binding assay were previously described in Tants et al. (2017).

4.1.5 Protein extract from cell culture cells for Western blots

Cells (500 µL) were harvested via centrifugation for 6 minutes at 3500 g before washing them twice in PBS. The cell pellet was then resuspended in 50 µL PBS + 8M urea and cooked at 95 °C for 10 minutes. Insoluble fragments were removed after a centrifugation of 10 minutes at 13000 g. Protein concentration of the extract was measured via a Bradford assay and 20 µg protein was loaded on the polyacrylamide gel for the Western blots.

4.1.6 Co-immunoprecipitation from double-tagged cell lines

10 mL cells were diluted to 1,5 million cells/mL and grown in regular cell culture medium supplemented with 200 µM $CuSO_4$ for expression of the N-terminally tagged proteins for four days. Cells were harvested via centrifugation and washed once with PBS. For mild crosslinking cells were resuspended in 10 mL PBS with 0,1 % formaldehyde for five minutes at room temperature while rotating the sample. The reaction was stopped by addition of 1 mL 1,25 M glycine before incubating the tube on ice for an additional five minutes. After pelleting via centrifugation cells were lysed in 1 mL lysis buffer (150 mM KAc pH 7.4, 30 mM HEPES pH 7.4, 5 mM MgAc, 1 mM DTT, 15 % glycerine, 1 % Tergitol, protease inhibitor tablet (cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail; Roche). Lysis was facilitated by sonification in a Bioruptor (20 cycles, 30 s on, 30 s off). Insoluble debris was removed by centrifugation for 5 mines at maximum speed at 4 °C. 25 µL (2,5 %) of the lysate/input fraction were kept for Western blot analysis.

The protein G beads (20 µL per IP) were prepared by washing them three times with lysis buffer before incubating them with 2 µL of the appropriate antibody for 1 h at 4 °C under constant rolling. After the incubation beads were washed three times with lysis buffer before the lysate was added and incubated with the beads for 1 h at 4 °C under rolling. 2,5 % of the supernatant was taken as a sample for Western Blot analysis. The beads were washed twice each with 750 µL of wash buffer 1 (150 mM KAc pH 7.4, 30 mM HEPES pH 7.4, 5 mM MgAc, 0,1 % Tergitol) and wash buffer 2 (150 mM KAc pH 7.4, 30 mM HEPES pH 7.4, 5 mM MgAc). A sample (3,3 %) from the first wash was taken for Western Blot analysis. After the removal of the last washing buffer 25 µL 1x SDS loading buffer was added to the beads and the samples were cooked at 95 °C for five minutes. The liquid portion was separated from the beads via centrifugation and loaded in its entirety on the acrylamide gel for Western blot analysis.

4.1.7 Fractionated lysis

Cells (6 mL) were harvested by centrifugation and washed once with PBS before resuspending them in 1 mL cytosolic lysis buffer (10 mM HEPES pH 7.4, 1 mM $MgCl_2$, 10 mM KCl, 1 mM DTT, 1 % Triton X-100, protease inhibitor tablet (cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail; Roche),

phosphatase inhibitor (PhosSTOP™, Roche)). The samples were incubated for 30 min on ice before being snap-frozen in liquid nitrogen. After thawing on ice, the samples were centrifuged at maximum speed at 4 °C for 15 minutes and the cytosolic lysate in the supernatant was kept. The pellet was resuspended in 1 mL nuclear lysis buffer (10 mM HEPES pH 7.4, 1,5 mM $MgCl_2$, 400 mM KCl, 1 mM DTT, 1 % Tergitol, 10 % glycerol, protease inhibitor tablet (cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail; Roche), phosphatase inhibitor (PhosSTOP™, Roche)). Lysis was facilitated using a Bioruptor (25 cycles, 30 s on, 30 s off). The samples were centrifuged at maximum speed at 4 °C for 15 minutes after which the nuclear lysate was in the supernatant. 15 µL of the cytosolic and nuclear lysates were used for Western blot analysis.

4.1.8 Adapted fCLIP protocol

Cells (10 ml) were harvested by centrifugation and washed once with PBS. The cell pellets were then resuspended in 5 mL PBS + 0,1 % formaldehyde and incubated under rolling for 5 minutes at room temperature. The crosslinking reaction was stopped by the addition of 555 µL of 1,5 M glycine and the samples were kept on ice for 5 minutes. The cell pellets were washed once with PBS before they were lysed according to the protocol in 4.1.7.

For the IP 20 µL Protein G Dynabead (Invitrogen) slurry was washed three times with 500 µL IP buffer (10 mM HEPES pH 7.4, 150 mM KCl, 0,1 % SDS, 0,5 % Na-DOC, 0,5 % Tergitol, protease inhibitor tablet (cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail; Roche)). These preequilibrated beads were incubated with 2 µL of the monoclonal anti-FLAG M2 antibody in 500 µL IP buffer for at least two hours at 4 °C. The beads were washed again three times with 500 µL IP buffer before the lysate samples were incubated with the beads at 4 °C for two hours. A 15 µL sample was kept before and after incubation with the beads to serve as input and supernatant samples for the Western blot analysis. The beads were washed six times with 10 mL IP buffer. With the last wash the beads were transferred to a new tube to get rid of unspecific binders to the tube surface. A sample of 15 µL from the third wash was kept for Western blot analysis. The beads were resuspended in 60 µL IP buffer and 10 µL were cooked in 1x SDS loading dye to serve as an IP sample for the Western blot.

The remaining 50 µL of resuspended beads were used for the Proteinase K digest. For this 500 µL Proteinase K buffer were incubated with 1 µL Proteinase K at 37 °C for 30 minutes. The preincubated buffer was added to the resuspended beads and incubated over night at 65 °C under shaking. For RNA isolation 300 µL Phenol/Chloroform/IAA (25:24:1) pH 8.0 was added before the samples were centrifuged for two minutes at 16000 rpm at room temperature. The supernatant was transferred to the same volume of isopropanol with 2,5 µL GlycoBlue™ (Thermo Fisher Scientific) and incubated for

10 min. The RNA was pelleted by centrifugation for 20 min at 16000 rpm at room temperature. After removal of the supernatant the pellet was washed with 150 µL 70 % EtOH which was removed after centrifugation (15 min, 16000 rpm, room temperature). The pellet was then incubated with 150 µL 70 % EtOH over night at -20 °C before the supernatant was removed after centrifugation (15 min, 16000 rpm, room temperature). The pellet was air-dried and resuspended in 15 µL RNase free water. The RNA was DNAse treated by adding 1 µL of DNase I, 1 µL of Ribolock, 5 µL of DNase buffer with magnesium and 28 µL of water before incubating at 37 °C for 30 minutes. The samples were then purified using the RNA Clean and Concentrator kit (Zymo Research) according to the manufacturer's protocol. The purified RNA was used as a whole for cDNA synthesis with SuperScript™ III (Invitrogen) according to the manufacturer's protocol. For qPCR the cDNA was diluted 1:10 and 1 µL cDNA was used per reaction.

4.1.9 Chromatin immunoprecipitation (ChIP)

Cells (10 mL) were diluted to 1,5 x 10^6 cells/mL, transfected with 7,5 µg sgRNA template using 80 µL Fugene HD (Promega) und then grown for three days at 25 °C. Cells were harvested by centrifugation at room temperature with 170 g for 6 minutes. The cell pellet was resuspended in 40 mL PBS and then added to 1,11 mL 37 % formaldehyde solution (f.c. 1 %). The sample was incubated under constant rolling for five minutes and the crosslinking reaction then stopped by adding 5,73 mL 2 M glycine (pH 7.0), mixed and incubated on ice for five minutes. The cells were pelleted via centrifugation at 1500 g for 4 minutes at 4 °C and resuspended in 4 mL ChIP wash A buffer (10 mM HEPES pH 7.6, 10 mM EDTA pH 8.0, 0,5 mM EGTA pH 8.0, 0,25 % Trition X-100, protease inhibitor tablet (cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail; Roche)). The sample was transferred to a 5 mL tube and rolled at 4 °C for 10 minutes. The cell pellet was then washed with ChIP wash B buffer (10 mM HEPES pH 7.6, 100 mM NaCl, 1 mM EDTA pH 8.0, 0,5 mM EGTA pH 8.0, 0,01 % Trition X-100, protease inhibitor tablet (cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail; Roche)) in the same way as done for wash buffer A. The chromatin was pelleted by centrifugation at 1500 g for four minutes at 4 °C. The pellet can then be frozen in liquid nitrogen and stored at -80 °C until used.

The Protein G beads for the IP were prepared as follows: 30 µL bead slurry (Dynabeads™ Protein G (Invitrogen)) were used per reaction and first washed with 2x 500 µL RIPA buffer (140 mM NaCl, 10 mM Tris-Cl pH 7.3, 1 mM EDTA pH 8.0, 1 % Triton X-100, 0,1 % SDS, 0,1 % sodiumdeoxycholate). 3 µL of the appropriate antibody was added together with 50 µL RIPA and incubated with the beads for 3 hours at 4 °C under rolling. The beads were then washed with 2x 500 µL RIPA before use in the IP.

The crosslinked chromatin was thawed on ice and resuspended in 1 mL TE buffer (10 mM Tris-Cl pH 7.3, 1 mM EDTA pH 8.0, protease inhibitor tablet (cOmplete™, Mini, EDTA-free Protease Inhibitor

Cocktail; Roche)). The supernatant was removed after centrifugation for five minutes at 1500 g and 4 °C. The pellet was resuspended in 1 mL TE + 0,1 % SDS. Lysis was facilitated using a Bioruptor (30 cycles, 30 s on, 30 s off). The buffer was adjusted to RIPA conditions by adding 60 µL 20 % Triton X-100, 12,2 µL 10 % sodiumdeoxycholate and 34,2 µL 5 M NaCl. The sample was rotated for 10 min at 4 °C. The lysate was harvested by centrifugation at maximum speed at 4 °C for 20 minutes before being transferred to a new tube and centrifuged again with the same conditions. 100 µL were taken as an input sample. The rest of the lysate was added to the beads and incubated under rolling for three hours at 4 °C under rolling. The supernatant was removed and the beads were washed 5x with 500 µL RIPA + protease inhibitor tablet (cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail; Roche). The beads were resuspended in 200 µL TE buffer (without protease inhibitor). In parallel 100 µL TE was added to the input sample.

To purify the DNA from the input and IP samples 4 µL RNAse (10 mg/mL) were added and incubated at 37 °C for 30 minutes. Proteins were degraded by addition of 10 µL 10 % SDS and 20 µL Proteinase K (10 mg/mL) and incubation for two hours at 56 °C under shaking. To revert the crosslink the samples were incubated at 65 °C overnight. The DNA was then purified using the Wizard® SV Gel and PCR Clean-Up System according to the manufacturer's protocol except that elution was performed with 60 µL water.

qPCR reaction mix:

2 µL purified DNA
5 µL SYBR™ Green PCR Master Mix (ThermoFisher Scientific)
0,5 µL 5 µM sense primer
0,5 µL 5 µM antisense primer
0,1 µL 0,03 % xylenecyanol
1,9 µL H_2O

qPCR program:

50 °C 10 s
95 °C 3 min

95 °C 30 s
59 °C 30 s
72 °C 42 s
} 40 cycles

Melting curve 59 °C to 95 °C in 1 °C steps, 6 s each

Data from the qPCR was gathered on a TOptical thermocycler (analytikjena) and Ct values were calculated using the qPCRsoft 3.4 (analytikjena). The remaining data analysis and calculation were performed in Microsoft Excel. The difference in volume used as an input sample and for the IP was accounted for in the calculations.

4.2 Cell culture

4.2.1 General cell culture and treatments

The *Drosophila melanogaster* S2 cells were grown in Schneider's medium (Bio&Sell) supplemented with 10 % FBS (Sigma) and Penicillin/Streptomycin (ThermoFisher Scientific). Cells were split 1:10 once a week by dilution in new medium.

Endogenous C- and N-terminal epitope tags were introduced via a PCR-based CRISPR/Cas9 protocol as previously described in Bottcher et al. (2014) and Kunzelmann et al. (2016).

RNAi knockdowns of genes was done by soaking cells in medium with 500 ng/ml of the corresponding dsRNA for four days. Details on the generation of the dsRNA can be found in Bottcher et al. (2014) and Kunzelmann et al. (2016).

To induce expression for the copper-inducible cell lines 200 µM of CuSO4 was added directly to the cell culture medium. Cells were then grown for 4 days before harvesting if not specified differently in the text.

4.2.2 Importazole assay and fluorescence microscopy

The Importazole assay was performed as published in Nitschko et al. (2020). In short, cells (Blanks-GFP or H2Av-GFP) were diluted to a concentration of $2x10^6$ cells/mL and Importazole (Sigma) was added to a concentration of 200 µM. For the control an equal volume of DMSO was used. Distribution of the fluorescent GFP-fusion protein was analysed by visual inspection. Fluorescence microscopy was performed with a Zeiss LSM710 confocal microscope. Nuclear DNA was counterstained with Hoechst33342 (1 µL of a 10 µg/ml solution was added to 10 µL cells).

4.3 *Drosophila melanogaster in vivo* methods

4.3.1 Fertility assay

One (transgene) male fly was crossed with two *yw* virgins. Both the virgins and males to be tested were aged for two days after hatching to make sure that they are fully matured. Egg laying was allowed for 4 days before all adult flies were removed from the vial. Pupae were counted 9 to 10 days after that.

4.3.2 Generation of transgenic fly lines

The injection procedure and crosses that were performed for generation of the *blanks*-transgenic flies and the *blanks* knockout mutant rescue fly lines were already described in Nitschko et al. (2020):

"For generation of the transgenic lines the plasmids with a mini-white marker were injected into

embryos and inserted into an attp2 site via the ϕC31 integrase method (Groth et al. (2004)). The injected flies were crossed with a *w[1118]* stock and the resulting red-eyed flies (from the transgene) were used for crosses to generate the homozygous transgene flies.

For the generation of the rescue line, mutant virgins (*y*^{1} *w***; +; Mi[MIC]blanks*^{MI10901}) were crossed with males from the homozygous *blanks* transgene lines. F1 virgins were crossed with males from the balancer stock *yw; +; D/TM3,Sb*. Through meiotic recombination both the *blanks* mutant and the *blanks* transgene can end up on the same chromosome. Flies with red eyes were backcrossed with the balancer stock to isolate the chromosome carrying the transgene with the mini-white marker. Brothers and sisters from the backcross were used to generate a homozygous stock and the presence of the *blanks*^{MI10901} mutant was confirmed by PCR."

The plasmids for injection were created as described in 4.1.1. For this study transgenic fly lines were created with all available *blanks* construct variants (pHF1-3, pSK42 and pVN3-5; s. also 4.4).

Rescue lines carrying both a *blanks* transgene and the *Mi{MIC}blanks*^{MI10901} mutation were created from these transgenic fly lines as described above with the exception of pHF1 for which no flies with the correct genotype was found after the crossing process.

To confirm the presence of the *Mi{MIC}blanks*^{MI10901} background mutation in the generated rescue lines a PCR was performed with the primers "*blanks* s us MI site" and "Minos as" (sequences can be found in 4.5).

4.3.3 RNA isolation from dissected testes

Testes from roughly 150 flies were dissected by hand and crushed with a pestle in 500 µL Trizole reagent (Ambion). RNA was isolated following the manufacturer's protocol and used for library preparation.

4.4 Plasmid list

name	**source**	**insert**	**resistance**
pHF1	AG Förstemann	*ubi*-3xFLAG-*blanks*-cDNA-polyA, Blasti-R, *attB*	Amp
pHF2	AG Förstemann	*ubi*-3xFLAG-NLS-*blanks*-cDNA-polyA, Blasti-R, *attB*	Amp
pHF3	AG Förstemann	*ubi*-3xFLAG-NES-*blanks*-cDNA-polyA, Blasti-R, *attB*	Amp
pHZ1	AG Förstemann	GST-PreScission_site-blanks-cDNA	Amp
pKF63	AG Förstemann	*ubi*-GFP-polyA, Blasti-R	Amp
pRB10	AG Förstemann	*ubi*-GFP-polyA, Blasti-R, *attB*	Amp
pSK42	AG Förstemann	*ubi-blanks*-cDNA-polyA, Blasti-R, *attB*	Amp

pVN1	this study	sgRNA template Actin 5C (specific primer #570)	Amp
pVN3	this study	*ubi*-3xFLAG- *blanks*-cDNA_mutdsRBD2-polyA, Blasti-R, *attB*	Amp
pVN4	this study	*ubi*-3xFLAG-NLS-*blanks*-cDNA_mutdsRBD2-polyA, Blasti-R, *attB*	Amp
pVN5	this study	*ubi*-3xFLAG-NES-*blanks*-cDNA_mutdsRBD2-polyA, Blasti-R, *attB*	Amp
pVN6	this study	GST-PreScission_site-*blanks*-cDNA_mutdsRBD2	Amp
pVN7	this study	6xHis-TEV_site-*blanks*-cDNA_mutdsRBD2	Kan
pVN8	this study	eXact_tag-*blanks*-cDNA_mutdsRBD2	Kan

4.5 Primer list

name	internal number	sequence
attP sense	87	TATGGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTA
attP antisense	88	TATACGCGCCCGGGGAGCCCAAGGGCACGCCCTGGCACCCCA
5’ wildtype	528	AAAGGATCCATGGAAGCAAAGCAATTGTTG
5’ FLAG	1343	TTGGATCCATGGACTACAAGGACCACGACGGCGACTACAAGGACCACG ACATCGACTACAAGGACGACGACAAGGAAGCAAAGCAATTGTTG
5’ FLAG-NLS	1344	TTGGATCCATGGACTACAAGGACCACGACGGCGACTACAAGGACCACG ACATCGACTACAAGGACGACGACAAGCCAGCCGCAAAAAGGGTAAAAC TCGACGAAGCAAAGCAATTGTTG
5’ FLAG-NES	1345	TTGGATCCATGGACTACAAGGACCACGACGGCGACTACAAGGACCACG ACATCGACTACAAGGACGACGACAAGATGCAGTTCGGTTTGCTGAGTCT GACGGAAGCAAAGCAATTGTTG
3’ wt-dsRBD2	930	TTGCGGCCGCTTATTTTTGTGGATAGTCGGTTCCA
3’ mut-dsRBD2	1346	AAAGCGGCCGCTTATTTTTGTGGATAGTCGGTTCCAAATAGTTTGTTACA AACTAAAGCGGATAGTTTGTAACGGGCTGCGGCTGCGGATGGCCCTTCA GC
Act5C CRISPR	570	CCTATTTTCAATTTAACGTCGACCGCAAGTGCTTCTAAGAGTTTAAGAGC TATGCTG
blanks s us MI site	1453	GATATTGATGTCGCAACCCT
Minos as	1452	TGTTAAACATTGCGCACTGC
tRNA qPCR s	975	GACCCCGACGTGATTTGAAC
tRNA qPCR as	976	ATCCCTCCCCTCATCTGACT
Act5C qPCR s	973	ACCGGTATCGTTCTGGACTC
Act5C qPCR s	974	CGGTCAGGATCTTCATCAGG

snRNA:7SK qPCR s	1288	ACCCTCCGTCACACCTTTG
snRNA:7SK qPCR as	1289	TAACCCGTCGTCATCCAGTG
RpL32 qPCR s	1360	ATCGGTTACGGATCGAACA
RpL32 qPCR as	1361	ACAATCTCCTTGCGCTTCTT
CG15098 qPCR 5' s	1047	TTGGCCTGCTCTGTAAGTGA
CG15098 qPCR 5' as	1048	GATAAGCGCAATGGGGATCC
CG15098 qPCR 3' s	996	ACCCAGCAGACCAACTCCTA
CG15098 qPCR 3' as	1001	ATGATGAAATAAGAGCTGGCCA
Tctp qPCR 5' s	1218	GGCGTGGATGTTGTGCTTAA
Tctp qPCR 5' as	848	GTCGCAGTCCATAGATTCGC
Tctp qPCR 3' s	1161	CCAAACAAAACGGAACCTATACTCATGCAT
Tctp qPCR 3' as	1160	CGCACTGATCGAAGACAGGCATT
GFP qPCR s	581	ACGTAAACGGCCACAAGTTC
GFP qPCR as	582	AAGTCGTGCTGCTTCATGTG
CG15098 CRISPR primer	991	CCTATTTTCAATTTAACGTCGGCTCGTTTTCAGTGCTGACCGTTTAAGAGCTATGCTG
Tctp CRISPR primer	748	CCTATTTTCAATTTAACGTCGTTCAAGCACGGTCTGGAGGGTTTAAGAGCTATGCTG

4.6 Fly lines

phenotype	**source**
y, w	AG Förstemann
w^{1118}	AG Förstemann
y^{1}, w^{*}; +; $Mi\{MIC\}blanks^{MI10901}$	Bloomington
y, w; +; D/TM3, Sb	AG Förstemann
y,w; nos-phiC31 (NLS); P{CaryP}attP2	AG Gaul
w; nos-phiC31 (NLS); P{CaryP}attP2:pHF1	this study
w; +; P{CaryP}attP2:pHF2	this study
w; +; P{CaryP}attP2:pHF3	this study
w; +; P{CaryP}attP2:pSK42	this study
y, w; nos-phiC31 (NLS); P{CaryP}attP2:pVN3	this study

y, w; nos-phiC31 (NLS); P{CaryP}attP2:pVN4	this study
y, w; nos-phiC31 (NLS); P{CaryP}attP2:pVN5	this study
w; +; Mi{MIC}blanks^{MI10901}, P{CaryP}attP2:pHF2	this study
w; +; Mi{MIC}blanks^{MI10901}, P{CaryP}attP2:pHF3	this study
w; +; Mi{MIC}blanks^{MI10901}, P{CaryP}attP2:pSK42	this study
w; +; Mi{MIC}blanks^{MI10901}, P{CaryP}attP2:pVN3	this study
w; +; Mi{MIC}blanks^{MI10901}, P{CaryP}attP2:pVN4	this study
w; +; Mi{MIC}blanks^{MI10901}, P{CaryP}attP2:pVN5	this study

5 Literature

Aravin, A. A., N. M. Naumova, A. V. Tulin, V. V. Vagin, Y. M. Rozovsky, and V. A. Gvozdev. 2001. 'Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline', *Curr Biol*, 11: 1017-27.

Ashwal-Fluss, Reut, Markus Meyer, Nagarjuna Reddy Pamudurti, Andranik Ivanov, Osnat Bartok, Mor Hanan, Naveh Evantal, Sebastian Memczak, Nikolaus Rajewsky, and Sebastian Kadener. 2014. 'circRNA Biogenesis Competes with Pre-mRNA Splicing', *Molecular Cell*, 56: 55-66.

Azlan, A., N. Dzaki, and G. Azzam. 2016. 'Argonaute: The executor of small RNA function', *Journal of genetics and genomics = Yi chuan xue bao*, 43: 481-94.

Bartel, D. P. 2004. 'MicroRNAs: genomics, biogenesis, mechanism, and function', *Cell*, 116: 281-97.

———. 2009. 'MicroRNAs: target recognition and regulatory functions', *Cell*, 136: 215-33.

Bernstein, E., A. A. Caudy, S. M. Hammond, and G. J. Hannon. 2001. 'Role for a bidentate ribonuclease in the initiation step of RNA interference', *Nature*, 409: 363-6.

Bohnsack, M. T., K. Czaplinski, and D. Gorlich. 2004. 'Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs', *Rna*, 10: 185-91.

Bonath, F., J. Domingo-Prim, M. Tarbier, M. R. Friedländer, and N. Visa. 2018. 'Next-generation sequencing reveals two populations of damage-induced small RNAs at endogenous DNA double-strand breaks', *Nucleic Acids Res*, 46: 11869-82.

Bottcher, R., M. Hollmann, K. Merk, V. Nitschko, C. Obermaier, J. Philippou-Massier, I. Wieland, U. Gaul, and K. Forstemann. 2014. 'Efficient chromosomal gene modification with CRISPR/cas9 and PCR-based homologous recombination donors in cultured Drosophila cells', *Nucleic acids research*.

Brennecke, J., A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, and G. J. Hannon. 2007. 'Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila', *Cell*, 128: 1089-103.

Brennecke, J., A. Stark, R. B. Russell, and S. M. Cohen. 2005. 'Principles of microRNA-target recognition', *PLoS Biol*, 3: e85.

Bronkhorst, A. W., and R. P. van Rij. 2014. 'The long and short of antiviral defense: small RNA-based immunity in insects', *Curr Opin Virol*, 7: 19-28.

Burgers, P. M. J., and T. A. Kunkel. 2017. 'Eukaryotic DNA Replication Fork', *Annual review of biochemistry*, 86: 417-38.

Bussing, I., J. S. Yang, E. C. Lai, and H. Grosshans. 2010. 'The nuclear export receptor XPO-1 supports primary miRNA processing in C. elegans and Drosophila', *Embo J*, 29: 1830-9.

Ceccaldi, R., B. Rondinelli, and A. D. D'Andrea. 2016. 'Repair Pathway Choices and Consequences at the Double-Strand Break', *Trends Cell Biol*, 26: 52-64.

Chung, W. J., K. Okamura, R. Martin, and E. C. Lai. 2008. 'Endogenous RNA interference provides a somatic defense against Drosophila transposons', *Curr Biol*, 18: 795-802.

Denli, A. M., B. B. Tops, R. H. Plasterk, R. F. Ketting, and G. J. Hannon. 2004. 'Processing of primary microRNAs by the Microprocessor complex', *Nature*, 432: 231-5.

Elbashir, S. M., W. Lendeckel, and T. Tuschl. 2001. 'RNA interference is mediated by 21- and 22-nucleotide RNAs', *Genes Dev*, 15: 188-200.

Elbashir, S. M., J. Martinez, A. Patkaniowska, W. Lendeckel, and T. Tuschl. 2001. 'Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate', *Embo J*, 20: 6877-88.

Elmer, K., S. Helfer, M. Mirkovic-Hosle, and K. Forstemann. 2014. 'Analysis of endo-siRNAs in Drosophila', *Methods in molecular biology*, 1173: 33-49.

Eulalio, A., F. Tritschler, and E. Izaurralde. 2009. 'The GW182 protein family in animal cells: new insights into domains required for miRNA-mediated gene silencing', *Rna*, 15: 1433-42.

Fire, A., S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, and C. C. Mello. 1998. 'Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans', *Nature*, 391: 806-11.

Forstemann, K., M. D. Horwich, L. Wee, Y. Tomari, and P. D. Zamore. 2007. 'Drosophila microRNAs Are Sorted into Functionally Distinct Argonaute Complexes after Production by Dicer-1', *Cell*, 130: 287-97.

Forstemann, K., Y. Tomari, T. Du, V. V. Vagin, A. M. Denli, D. P. Bratu, C. Klattenhoff, W. E. Theurkauf, and P. D. Zamore. 2005. 'Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein', *PLoS Biol*, 3: e236.

Francia, Sofia, Flavia Michelini, Alka Saxena, Dave Tang, Michiel de Hoon, Viviana Anelli, Marina Mione, Piero Carninci, and Fabrizio d 'Adda di Fagagna. 2012. 'Site-specific DICER and DROSHA RNA products control the DNA-damage response', *Nature*, advance online publication.

Fukaya, T., and Y. Tomari. 2012. 'MicroRNAs mediate gene silencing via multiple different pathways in drosophila', *Molecular Cell*, 48: 825-36.

Gerbasi, V. R., J. B. Preall, D. E. Golden, D. W. Powell, T. D. Cummins, and E. J. Sontheimer. 2011. 'Blanks, a nuclear siRNA/dsRNA-binding complex component, is required for Drosophila spermiogenesis', *Proceedings of the National Academy of Sciences of the United States of America*, 108: 3204-9.

Ghildiyal, M., H. Seitz, M. D. Horwich, C. Li, T. Du, S. Lee, J. Xu, E. L. Kittler, M. L. Zapp, Z. Weng, and P. D. Zamore. 2008. 'Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells', *Science*, 320: 1077-81.

Ghildiyal, M., and P. D. Zamore. 2009. 'Small silencing RNAs: an expanding universe', *Nature reviews. Genetics*, 10: 94-108.

Gregory, R. I., K. P. Yan, G. Amuthan, T. Chendrimada, B. Doratotaj, N. Cooch, and R. Shiekhattar. 2004. 'The Microprocessor complex mediates the genesis of microRNAs', *Nature*, 432: 235-40.

Groth, A. C., M. Fish, R. Nusse, and M. P. Calos. 2004. 'Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31', *Genetics*, 166: 1775-82.

Gunawardane, L. S., K. Saito, K. M. Nishida, K. Miyoshi, Y. Kawamura, T. Nagami, H. Siomi, and M. C. Siomi. 2007. 'A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila', *Science*, 315: 1587-90.

Gwizdek, C., B. Ossareh-Nazari, A. M. Brownawell, S. Evers, I. G. Macara, and C. Dargemont. 2004. 'Minihelix-containing RNAs mediate exportin-5-dependent nuclear export of the double-stranded RNA-binding protein ILF3', *J Biol Chem*, 279: 884-91.

Hammond, S. M., E. Bernstein, D. Beach, and G. J. Hannon. 2000. 'An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells', *Nature*, 404: 293-6.

Hammond, S. M., S. Boettcher, A. A. Caudy, R. Kobayashi, and G. J. Hannon. 2001. 'Argonaute2, a link between genetic and biochemical analyses of RNAi', *Science*, 293: 1146-50.

Han, J., Y. Lee, K. H. Yeom, Y. K. Kim, H. Jin, and V. N. Kim. 2004. 'The Drosha-DGCR8 complex in primary microRNA processing', *Genes & development*, 18: 3016-27.

Hartig, J. V., S. Esslinger, R. Bottcher, K. Saito, and K. Forstemann. 2009. 'Endo-siRNAs depend on a new isoform of loquacious and target artificially introduced, high-copy sequences', *Embo J*, 28: 2932-44.

Hartig, J. V., and K. Forstemann. 2011. 'Loqs-PD and R2D2 define independent pathways for RISC generation in Drosophila', *Nucleic acids research*, 39: 3836-51.

Hartig, J. V., Y. Tomari, and K. Forstemann. 2007. 'piRNAs--the ancient hunters of genome invaders', *Genes Dev*, 21: 1707-13.

Horwich, M. D., C. Li, C. Matranga, V. Vagin, G. Farley, P. Wang, and P. D. Zamore. 2007. 'The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC', *Curr Biol*, 17: 1265-72.

Jakob, L., T. Treiber, N. Treiber, A. Gust, K. Kramm, K. Hansen, M. Stotz, L. Wankerl, F. Herzog, S. Hannus, D. Grohmann, and G. Meister. 2016. 'Structural and functional insights into the fly microRNA biogenesis factor Loquacious', *Rna*, 22: 383-96.

Ji, L., and X. Chen. 2012. 'Regulation of small RNA stability: methylation and beyond', *Cell Res*, 22: 624-36.

Jiang, F., X. Ye, X. Liu, L. Fincher, D. McKearin, and Q. Liu. 2005. 'Dicer-1 and R3D1-L catalyze microRNA maturation in Drosophila', *Genes Dev*, 19: 1674-9.

Khurana, J. S., and W. Theurkauf. 2010. 'piRNAs, transposon silencing, and Drosophila germline development', *The Journal of cell biology*, 191: 905-13.

Khvorova, A., A. Reynolds, and S. D. Jayasena. 2003. 'Functional siRNAs and miRNAs exhibit strand bias', *Cell*, 115: 209-16.

Kim, B., and V. N. Kim. 2019. 'fCLIP-seq for transcriptomic footprinting of dsRNA-binding proteins: Lessons from DROSHA', *Methods*, 152: 3-11.

Kim, K., Y. S. Lee, and R. W. Carthew. 2007. 'Conversion of pre-RISC to holo-RISC by Ago2 during assembly of RNAi complexes', *Rna*, 13: 22-9.

Kim, V. N., J. Han, and M. C. Siomi. 2009. 'Biogenesis of small RNAs in animals', *Nature reviews. Molecular cell biology*, 10: 126-39.

Kunzelmann, Stefan Bernhard. 2017. 'Characterization of cis-elements and trans-factors that are involved in RNAi-mediated genome defense in Drosophila melanogaster', Dissertation, LMU München.

Kunzelmann, Stefan, Romy Bottcher, Ines Schmidts, and Klaus Forstemann. 2016. 'A Comprehensive Toolbox for Genome Editing in Cultured Drosophila melanogaster Cells', *G3 (Bethesda, Md)*, 6: 1777-85.

Lee, Heng-Chi, Shwu-Shin Chang, Swati Choudhary, Antti P. Aalto, Mekhala Maiti, Dennis H. Bamford, and Yi Liu. 2009. 'qiRNA is a new type of small interfering RNA induced by DNA damage', *Nature*, 459: 274-77.

Lee, Y., C. Ahn, J. Han, H. Choi, J. Kim, J. Yim, J. Lee, P. Provost, O. Rådmark, S. Kim, and V. N. Kim. 2003. 'The nuclear RNase III Drosha initiates microRNA processing', *Nature*, 425: 415-9.

Lee, Y., K. Jeon, J. T. Lee, S. Kim, and V. N. Kim. 2002. 'MicroRNA maturation: stepwise processing and subcellular localization', *Embo J*, 21: 4663-70.

Lee, Y., M. Kim, J. Han, K. H. Yeom, S. Lee, S. H. Baek, and V. N. Kim. 2004. 'MicroRNA genes are transcribed by RNA polymerase II', *Embo J*, 23: 4051-60.

Li, X., C. X. Liu, W. Xue, Y. Zhang, S. Jiang, Q. F. Yin, J. Wei, R. W. Yao, L. Yang, and L. L. Chen. 2017. 'Coordinated circRNA Biogenesis and Function with NF90/NF110 in Viral Infection', *Mol Cell*, 67: 214-27.e7.

Liao, S. E., Y. Ai, and R. Fukunaga. 2018. 'An RNA-binding protein Blanks plays important roles in defining small RNA and mRNA profiles in Drosophila testes', *Heliyon*, 4: e00706.

Liu, J., M. A. Carmell, F. V. Rivas, C. G. Marsden, J. M. Thomson, J. J. Song, S. M. Hammond, L. Joshua-Tor, and G. J. Hannon. 2004. 'Argonaute2 is the catalytic engine of mammalian RNAi', *Science*, 305: 1437-41.

Liu, Q., and Z. Paroo. 2010. 'Biochemical principles of small RNA pathways', *Annual review of biochemistry*, 79: 295-319.

Liu, Q., T. A. Rand, S. Kalidas, F. Du, H. E. Kim, D. P. Smith, and X. Wang. 2003. 'R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway', *Science*, 301: 1921-5.

Lund, E., and J. E. Dahlberg. 2006. 'Substrate selectivity of exportin 5 and Dicer in the biogenesis of microRNAs', *Cold Spring Harb Symp Quant Biol*, 71: 59-66.

Lund, E., S. Guttinger, A. Calado, J. E. Dahlberg, and U. Kutay. 2004. 'Nuclear export of microRNA precursors', *Science*, 303: 95-8.

Matranga, C., Y. Tomari, C. Shin, D. P. Bartel, and P. D. Zamore. 2005. 'Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes', *Cell*, 123: 607-20.

Meister, G. 2013. 'Argonaute proteins: functional insights and emerging roles', *Nat Rev Genet*, 14: 447-59.

Meister, G., M. Landthaler, A. Patkaniowska, Y. Dorsett, G. Teng, and T. Tuschl. 2004. 'Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs', *Mol Cell*, 15: 185-97.

Merk, K., M. Breinig, R. Bottcher, S. Krebs, H. Blum, M. Boutros, and K. Forstemann. 2017. 'Splicing stimulates siRNA formation at Drosophila DNA double-strand breaks', *PLoS genetics*, 13: e1006861.

Michalik, K. M., R. Bottcher, and K. Forstemann. 2012. 'A small RNA response at DNA ends in Drosophila', *Nucleic acids research*, 40: 9596-603.

Mirkovic-Hosle, M., and K. Forstemann. 2014. 'Transposon defense by endo-siRNAs, piRNAs and somatic pilRNAs in Drosophila: contributions of Loqs-PD and R2D2', *PLoS ONE*, 9: e84994.

Miyoshi, K., H. Tsukumo, T. Nagami, H. Siomi, and M. C. Siomi. 2005. 'Slicer function of Drosophila Argonautes and its involvement in RISC formation', *Genes Dev*, 19: 2837-48.

Napoli, C., C. Lemieux, and R. Jorgensen. 1990. 'Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans', *Plant Cell*, 2: 279-89.

Nguyen Ba, A. N., A. Pogoutse, N. Provart, and A. M. Moses. 2009. 'NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction', *BMC Bioinformatics*, 10: 202.

Nitschko, V., S. Kunzelmann, T. Frohlich, G. J. Arnold, and K. Forstemann. 2020. 'Trafficking of siRNA precursors by the dsRBD protein Blanks in Drosophila', *Nucleic Acids Res*.

Nitschko, Volker Fabian. 2016. 'ChIP of factors at the DNA double-strand break and experiments with genetic alterations in Drosophila melanogaster embryos', Master Thesis, LMU Munich.

Okamura, K., S. Balla, R. Martin, N. Liu, and E. C. Lai. 2008. 'Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in Drosophila melanogaster', *Nat Struct Mol Biol*, 15: 581-90.

Okamura, K., W. J. Chung, J. G. Ruby, H. Guo, D. P. Bartel, and E. C. Lai. 2008. 'The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs', *Nature*, 453: 803-6.

Okamura, K., J. W. Hagen, H. Duan, D. M. Tyler, and E. C. Lai. 2007. 'The Mirtron Pathway Generates microRNA-Class Regulatory RNAs in Drosophila', *Cell*.

Okamura, K., and E. C. Lai. 2008. 'Endogenous small interfering RNAs in animals', *Nat Rev Mol Cell Biol*, 9: 673-8.

Park, J. M., J. Werner, J. M. Kim, J. T. Lis, and Y. J. Kim. 2001. 'Mediator, not holoenzyme, is directly recruited to the heat shock promoter by HSF upon heat shock', *Mol Cell*, 8: 9-19.

Quinlan, A. R. 2014. 'BEDTools: The Swiss-Army Tool for Genome Feature Analysis', *Curr Protoc Bioinformatics*, 47: 11 12 1-34.

Rand, T. A., S. Petersen, F. Du, and X. Wang. 2005. 'Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation', *Cell*, 123: 621-9.

Ruby, J. G., C. H. Jan, and D. P. Bartel. 2007. 'Intronic microRNA precursors that bypass Drosha processing', *Nature*, 448: 83-6.

Sabin, L. R., S. L. Hanna, and S. Cherry. 2010. 'Innate antiviral immunity in Drosophila', *Curr Opin Immunol*, 22: 4-9.

Sabin, L. R., Q. Zheng, P. Thekkat, J. Yang, G. J. Hannon, B. D. Gregory, M. Tudor, and S. Cherry. 2013. 'Dicer-2 processes diverse viral RNA species', *PLoS ONE*, 8: e55458.

Saito, K., A. Ishizuka, H. Siomi, and M. C. Siomi. 2005. 'Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells', *PLoS Biol*, 3: e235.

Sanders, C., and D. P. Smith. 2011. 'LUMP is a putative double-stranded RNA binding protein required for male fertility in Drosophila melanogaster', *PLoS ONE*, 6: e24151.

Schmidts, I., R. Bottcher, M. Mirkovic-Hosle, and K. Forstemann. 2016. 'Homology directed repair is unaffected by the absence of siRNAs in Drosophila melanogaster', *Nucleic acids research*, 44: 8261-71.

Schneiderman, J. I., S. Goldstein, and K. Ahmad. 2010. 'Perturbation analysis of heterochromatin-mediated gene silencing and somatic inheritance', *PLoS genetics*, 6: e1001095.

Schuster, S., P. Miesen, and R. P. van Rij. 2019. 'Antiviral RNAi in Insects and Mammals: Parallels and Differences', *Viruses*, 11.

Schwarz, D. S., G. Hutvagner, T. Du, Z. Xu, N. Aronin, and P. D. Zamore. 2003. 'Asymmetry in the assembly of the RNAi enzyme complex', *Cell*, 115: 199-208.

Shemesh, K., M. Sebesta, M. Pacesa, S. Sau, A. Bronstein, O. Parnas, B. Liefshitz, C. Venclovas, L. Krejci, and M. Kupiec. 2017. 'A structure-function analysis of the yeast Elg1 protein reveals the importance of PCNA unloading in genome stability maintenance', *Nucleic Acids Res*, 45: 3189-203.

Siomi, M. C., T. Miyoshi, and H. Siomi. 2010. 'piRNA-mediated silencing in Drosophila germlines', *Seminars in cell & developmental biology*, 21: 754-9.

Siomi, M. C., K. Sato, D. Pezic, and A. A. Aravin. 2011. 'PIWI-interacting small RNAs: the vanguard of genome defence', *Nature reviews. Molecular cell biology*, 12: 246-58.

Soderholm, J. F., S. L. Bird, P. Kalab, Y. Sampathkumar, K. Hasegawa, M. Uehara-Bingen, K. Weis, and R. Heald. 2011. 'Importazole, a small molecule inhibitor of the transport receptor importin-beta', *ACS chemical biology*, 6: 700-8.

Swenson, J. M., S. U. Colmenares, A. R. Strom, S. V. Costes, and G. H. Karpen. 2016. 'The composition and organization of Drosophila heterochromatin are heterogeneous and dynamic', *eLife*, 5.

Tants, J. N., S. Fesser, T. Kern, R. Stehle, A. Geerlof, C. Wunderlich, M. Juen, C. Hartlmuller, R. Bottcher, S. Kunzelmann, O. Lange, C. Kreutz, K. Forstemann, and M. Sattler. 2017. 'Molecular basis for asymmetry sensing of siRNAs by the Drosophila Loqs-PD/Dcr-2 complex in RNA interference', *Nucleic Acids Res*, 45: 12536-50.

Tomari, Y., C. Matranga, B. Haley, N. Martinez, and P. D. Zamore. 2004. 'A protein sensor for siRNA asymmetry', *Science*, 306: 1377-80.

van Rij, R. P., and E. Berezikov. 2009. 'Small RNAs and the control of transposons and viruses in Drosophila', *Trends in microbiology*, 17: 163-71.

Wang, W., B. W. Han, C. Tipping, D. T. Ge, Z. Zhang, Z. Weng, and P. D. Zamore. 2015. 'Slicing and Binding by Ago3 or Aub Trigger Piwi-Bound piRNA Production by Distinct Mechanisms', *Mol Cell*, 59: 819-30.

Weber, F., V. Wagner, S. B. Rasmussen, R. Hartmann, and S. R. Paludan. 2006. 'Double-stranded RNA is produced by positive-strand RNA viruses and DNA viruses but not in detectable amounts by negative-strand RNA viruses', *J Virol*, 80: 5059-64.

Wei, W., Z. Ba, M. Gao, Y. Wu, Y. Ma, S. Amiard, C. I. White, J. M. Danielsen, Y. G. Yang, and Y. Qi. 2012. 'A Role for Small RNAs in DNA Double-Strand Break Repair', *Cell*, 149: 101-12.

Wen, J., H. Duan, F. Bejarano, K. Okamura, L. Fabian, J. A. Brill, D. Bortolamiol-Becet, R. Martin, J. G. Ruby, and E. C. Lai. 2015. 'Adaptive regulation of testis gene expression and control of male fertility by the Drosophila hairpin RNA pathway. [Corrected]', *Mol Cell*, 57: 165-78.

Wilson, R. C., and J. A. Doudna. 2013. 'Molecular mechanisms of RNA interference', *Annu Rev Biophys*, 42: 217-39.

Yi, R., Y. Qin, I. G. Macara, and B. R. Cullen. 2003. 'Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs', *Genes Dev*, 17: 3011-6.

Zhou, R., B. Czech, J. Brennecke, R. Sachidanandam, J. A. Wohlschlegel, N. Perrimon, and G. J. Hannon. 2009. 'Processing of Drosophila endo-siRNAs depends on a specific Loquacious isoform', *Rna*, 15: 1886-95.

Zhou, R., I. Hotta, A. M. Denli, P. Hong, N. Perrimon, and G. J. Hannon. 2008. 'Comparative analysis of argonaute-dependent small RNA pathways in Drosophila', *Molecular Cell*, 32: 592-9.

6 Acknowledgements

First of all, my special thanks go to my Ph.D. (and M.Sc. and B.Sc.) supervisor Prof. Dr. Klaus Förstemann. Thank you for all the support, discussions and motivation you were always happy to give over the times in which I had the pleasure of being part of your lab. I guess we both did not expect into what my request to join you team for my B.Sc. thesis would eventually evolve - much like the at this time still a little bit obscure CRISPR/Cas9 technology I started my work on. Thank you for taking initiative at several crucial milestones of my Ph.D. student career. I might have continued "ChIP'ing" the RFC-complex for far longer if you would have not recognized that we were banging our metaphorical head against the wall. This allowed us to continue the "Blanks story" and bring it to a satisfying conclusion with the publication which you also captained into a save harbour after some rough experimental weather in the revisions phase. No matter where the future will steer me I will always happily think back on the time I spend in your lab and under your mentorship.

I would also like to thank Dietmar Martin and the other members of my dissertation committee for taking the time to read and evaluate this thesis and my performance in the defence. My thanks also go out to Katja Lammens and Nicolas Gompel for being part of my TAC and starting me of running with the Blanks project.

Now to the people in the lab: Of course, I have to thank Romy before all others. You always like to keep reminding me how much shyer/different/weirder I was when you met my 23-year-old self. I don't think I can (or want to) argue with that. Even though our first experiment together much more resembled modern black and blue art than a Western blot, you still did not object when Klaus told you I would like to come back or even stay for the Master and Ph.D. thesis. I hope you never regretted not vetoing these plans. You even had to go to the French Mediterranean Sea with me shortly after the start of my Ph.D. Thank you for the friendship that evolved over time and the fact that we both learned that we could rely on each other no matter what the lab day to day life threw at us - especially when it was basically just the two of us in the lab for several months. Thank you for always having an open ear. I will never forget that. Additionally, I think I can now admit it that you might not have been completely wrong with your (hair) fashion tips.

Petar, my Serbian friend: Although we are different in many things working with you has always been a pleasure over the last two and a half years. You were thrown into a difficult situation when you started in the lab. Not only were you starting to work in a lab in which the other members still had to get used to the fact that they could no longer keep on talking German for the whole day, but you were also thrown into a group that knew each other for quite a while. That also meant that you were thrown into "our" way of working which did not always align with yours and therefore certainly

was not easy to adapt to and to find your place. Now a few years later you speak perfektes Hochdeutsch, understand my apparently difficult Swabian dialect and are a member of the team I would not like to miss. It is/was a pleasure working with you!

Selina, first of all thank you for not being scared away by one of my first sentences to you. You understood that it was laced with an overplayed local patriotism. We continued on a friendly path to strengthen the Swabian-Badener relationship from then on. Although we had limited time during your interview Petar, Klaus and I knew that we could dare to offer you a position even though Romy was in Australia. You never gave us any reason to regret this decision and quickly became a valued member of the lab family. Thank you also for encouraging me to pick up the tennis racket again (at least for the one time) and the times we went to explore some of the culinary highlights of southwestern Munich.

Ines, Stefan and Karin I think you can agree that we had a nice and especially fun time in the 1 to 1,5 years we spend together in the lab until you left us for a higher calling. Your help, encouragement and expertise allowed me to hit the ground running for transitioning from being the least experienced to being the most experienced Ph.D. student in the lab within a matter of months. Thank you also, Ines, for proof-reading and your feedback on large parts of this thesis.

Thank you to the numerous students I had the pleasure of supervising over the last years in both the lab and the practical courses. I hope I could spread some of my knowledge and enthusiasm for biochemistry to you all. Thank you to Hannah especially, since you were “my” first student and for the cloning work you did that we at this time did not suspect to play such an important role for one of my projects.

In addition to the people in our lab I also want to thank the people from all the other labs I had the pleasure to work with over the time. I especially have to thank the people from the Beckmann, Herzog and Stingele lab with which I had many nice conversations in the kitchens, labs and cafeterias in between experiments and who never hesitated to lend a hand if their special expertise was required or even adopted what was left of the AG Förstemann to their Christmas celebration. I also have to thank Bettina for showing me how to successfully inject fly embryos and Christophe for a proper introduction into fluorescence microscopy.

I also want to thank many people in the Gene Center administration and infrastructure departments that are not directly involved in the research, but still are essential to keep the institute running. My special thanks go out to Michael Englschall, Gabrielle Bittner and Aleksandra Sarman-Grilc to name just a few.

Anna, Tiana and Cati, I want to thank you for the friendship that followed your time in the lab even though you could or did not want to stay, but found you luck in other labs/cities. However, I could not let the fun and conversations stop after you left and now I have the pleasure to “torture” you Inga and Ines once a month by throwing imaginary dragons, elves and “Buletten” your way. I hope we can keep up this tradition for a long time. While on the subject I also want to thank the other D&D group (Kadda, Elli, Dieter, Michelle, Wibke and Mihaela) that was build up during my time on the 3rd floor for the adventures we have together.

While talking about D&D and good friends I have to mention Inga in particular who started off as a B.Sc. student in that lab and that did then not want to leave anymore (or did we not want to let her go?). You became one of the best friends I ever had and I treasure all the times, talks and text conversations we had together which we will hopefully have many more of in the future.

In the end I would like to thank my parents and my sister for supporting me with your love from the distant Swabian homeland. You were there for me and supported me especially when times occasionally got difficult during the studies and the Ph.D. years and pushed me or lifted me up when it was necessary. You then had to suffer through my explanations about small RNAs, flies and other things for which it might have been hard to understand why I find them exciting and fascinating.