

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

Biogenesis and biological function of  
DNA double-strand break triggered  
small interfering RNAs in *Drosophila melanogaster*

---



---

Ines Schmidts  
aus  
Zeiden, Rumänien

2017

## 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, 31.07.2017

.....

(Ines Schmidts)

Dissertation eingereicht am 03.08.2017

1. Gutachterin / 1. Gutachter: Prof. Dr. Klaus Förstemann

2. Gutachterin / 2. Gutachter: PD Dr. Dietmar Martin

Mündliche Prüfung am 24.10.2017

“The beauty of science is that it does not claim to know the answers before it asks the questions. There is nothing wrong with not knowing. It means there is more to learn[.]”

Lawrence Krauss

## Table of contents

<b>I Summary</b> .....	<b>1</b>
<b>II Introduction</b> .....	<b>3</b>
II.1 A DNA double-strand break.....	3
II.2 The protein response to a DSB.....	4
II.2.1 DSB recognition and signaling.....	4
II.2.2 DSB repair .....	4
II.2.2.1 HR repair .....	5
II.2.2.2 MMEJ repair .....	6
II.2.2.3 NHEJ repair.....	6
II.3 An RNA response to a DSB .....	7
II.4 Reporter systems to investigate the role of RNA in DSB repair .....	9
II.4.1 SSA reporter .....	9
II.4.2 Reporters to assess Rad51 homology directed repair.....	12
II.5 Transcription and the break .....	14
II.6 Objectives of the present study.....	15
<b>III Results</b> .....	<b>19</b>
III.1 Antisense transcription at the break.....	19
III.1.1 Establishing a NET-Seq protocol for <i>Drosophila</i> S2 cells.....	20
III.1.1.1 Tagging of RNA polymerase II subunits.....	20
III.1.1.1.1 Tagging of <i>RpII33</i> .....	21
III.1.1.1.2 Tagging of <i>RpII215</i> .....	23
III.1.1.2 Optimization of RNA polymerase II extraction from <i>Drosophila</i> S2 cells.....	24
III.1.1.3 Optimization of RNA polymerase II immunoprecipitation and specific elution using RpII33-Flag <sub>2</sub> tagged cell lines.....	32
III.1.1.4 Optimization of RNA polymerase II immunoprecipitation and specific elution using RpII33-Flag <sub>2</sub> RpII215-Strep <sub>2</sub> tagged cell lines .....	35

III.1.1.5 Evaluation of co-purified RNA from RpII215-Strep <sub>2</sub> IP by deep sequencing and NET-RT-PCR.....	36
III.1.2 Preliminary results on transcription at induced DNA double-strand breaks.....	37
III.1.2.1 Quality of NET-Seq libraries.....	40
III.1.2.2 Transcriptional response to a DSB.....	43
III.2 Biological role of DSB-triggered small RNAs .....	51
III.2.1 Single-strand annealing repair is unaffected by the absence of DSB-induced siRNAs .....	51
III.2.1.1 Assessing the putative link between DSB-triggered siRNAs and SSA in <i>Drosophila</i> S2 cells.....	52
III.2.1.2 Assessing the putative link between DSB-triggered siRNAs and SSA <i>in vivo</i> .....	58
III.2.2 Rad51-dependent HDR repair is unaffected by the absence of DSB-induced siRNAs .....	61
III.2.3 The DSB stress response is unaffected by the absence of DSB-triggered siRNAs ....	64
<b>IV Discussion.....</b>	<b>71</b>
IV.1 Further optimization is needed to improve specific enrichment of reads mapping to RNAPII transcribed loci in NET-Seq libraries.....	71
IV.2 Further optimization of the NET-Seq protocol is needed to obtain a pattern of DSB-triggered antisense transcription .....	72
IV.3 Homology directed repair and the general DSB stress response are unaffected by the absence of DSB-triggered siRNAs .....	74
<b>V Materials and Methods .....</b>	<b>79</b>
V.1 Materials.....	79
V.1.1 Generation of sgRNA and sgRNA expression templates .....	79
V.1.2 Generation of HR donor templates for C-terminal tagging.....	80
V.1.3 dsRNA constructs.....	82
V.1.4 Cell lines used in this study.....	83
V.2 Methods.....	84
V.2.1 Cell culture .....	84
V.2.1.1 General culturing conditions and transfection.....	84
V.2.1.2 SSA assay .....	84
V.2.1.3 C-terminal tagging.....	85
V.2.1.3.1 C-terminal tagging of RNAPII subunits.....	85
V.2.1.3.2 C-terminal tagging of <i>act5C</i> for the HDR assay.....	85

V.2.1.4	Introduction of locus-specific DSBs exploiting the CRISPR/ <i>cas9</i> system for analysis of the antisense transcription at a break via NET-Seq.....	86
V.2.2	<i>In vitro</i> experimental procedures .....	86
V.2.2.1	T7-Endonuclease Assay .....	86
V.2.2.2	Cell lysis .....	87
V.2.2.2.1	Cytoplasmic lysis protocol .....	87
V.2.2.2.2	Cellular fractionation protocol (modified from Abcam).....	88
V.2.2.2.3	Combined fractionation and sonication protocol .....	88
V.2.2.2.4	Optimized lysis protocol.....	89
V.2.2.2.5	Cell lysis by cryogenic grinding .....	89
V.2.2.3	Immunoprecipitation .....	89
V.2.2.4	Elution .....	90
V.2.2.5	Western blotting .....	90
V.2.2.6	Recovery of co-immunoprecipitated RNA .....	91
V.2.2.7	NET-RT-PCR.....	91
V.2.2.8	Generation of small RNA libraries from co-immunoprecipitated RNA .....	92
V.2.2.9	Deep sequencing data analysis.....	92
V.2.3	Fly strains and husbandry .....	92
V.2.3.1.1	Fly strains.....	92
V.2.3.1.2	<i>In vivo</i> SSA assay.....	93
V.2.3.1.2.1	Fluorescence microscopy.....	94
V.2.3.1.2.2	Preparation of fly extracts for Western blotting.....	94
V.2.3.1.3	<i>In vivo</i> camptothecin hypersensitivity assay.....	94
	<b>References .....</b>	<b>97</b>
	<b>List of Figures .....</b>	<b>109</b>
	<b>List of tables.....</b>	<b>111</b>
	<b>Acknowledgments.....</b>	<b>113</b>

## I Summary

Small interfering RNAs (siRNAs) play a crucial role in genome surveillance and protect the organism against harmful transcripts from exogenous (e.g. viral) or endogenous (e.g. transposons) sources. The discovery of siRNAs induced by DNA double-strand breaks (DSB) in *Neurospora crassa*, *Arabidopsis thaliana*, *Drosophila melanogaster* and human cells underlines that there might be yet another facet to the genome surveillance properties of siRNAs. siRNA production depends on a long double-stranded RNA precursor, with the strands showing perfect complementarity. While *Neurospora crassa* and *Arabidopsis thaliana* encode an RNA-dependent RNA polymerase, this enzyme is not known to be present in *Drosophila* and mammalia.

Since production of these siRNAs depends on local transcription, this could imply that the break acts as a bidirectional promoter. Recruitment of an RNA polymerase could then result in the required antisense transcription. To this end, we tried to establish nascent elongating transcript (NET-) Sequencing. This technique allows to map actively transcribing RNA polymerase with nucleotide resolution and thus to gain insight into ongoing transcription at a certain time point, e.g. upon a DNA double-strand break. Following lysis of the cells, RNA polymerase II (RNAPII) is immunoprecipitated. The co-purified RNA is incorporated into a cDNA library and subjected to deep sequencing. During the thesis, we were able to generate RNAPII tagged cell lines and we could optimize the lysis and immunopurification protocol. Using this protocol, we generated deep sequencing libraries from untreated cells and cells with a CRISPR/*cas9* induced double-strand break at three different loci (*CG15098*, *Tctp*, *7SK*). Due to a very poor signal-to-noise ratio, the initial goal to monitor a putative antisense transcription profile at the break to elucidate biogenesis of the long dsRNA precursor could not be achieved. Nevertheless, the preliminary data shows antisense reads for *CG15098*, a locus that is transcribed by RNA polymerase II and contains introns. For *Tctp*, an intronless gene, no antisense reads were observed. While we detected antisense reads for the *7SK* locus that only occurred upon a break at the locus, none of the antisense RNAs was longer than 21 nt. This argues for a contamination of the NET-Seq library with siRNAs rather than for the occurrence of RNAPII transcripts.

Since the small RNA response to a DSB is well-conserved and has been observed in fungi, plants, the fruit fly as well as human cells, we speculated that there also might be a conserved biological role. A current hypothesis in the field is that DSB-triggered small RNAs ensure efficient homologous recombination (HR). However, in many organisms biogenesis of siRNAs is intertwined with other small RNA species (e.g. microRNAs). The possible indirect effects substantially complicate dissection of the putative function of break derived siRNAs. In contrast to this, small RNA pathways in *Drosophila* are well separated. siRNAs are produced by Dcr-2, while Dcr-1 is the dedicated biogenesis factor for microRNAs (miRNAs). This allowed us to investigate the function of DSB-induced siRNAs without miRNA deregulation. This thesis examines the putative role of break-derived siRNAs in cultured *Drosophila* cells as well as mutant flies. We made use of reporters to assess efficiencies of the single-strand annealing pathway, homology directed repair and the systemic DSB stress response upon treatment of flies with the DSB-inducing drug camptothecin. We did not observe any DSB repair impairments for any of the tested scenarios. We therefore conclude that the function of break-induced siRNAs resides in transcript surveillance rather than in DSB repair.

## **II Introduction**

### **II.1 A DNA double-strand break**

Genome integrity is constantly challenged by extrinsic and intrinsic agents, stresses and processes. Faithful repair is therefore essential. The DNA double-strand break (DSB), in which the phosphate backbones of both DNA strands are severed, is considered the most genotoxic DNA lesion, since there is no complementary strand that could be used as template for repair (as is the case for e.g. nucleotide excision repair or DNA mismatch repair) (van Gent et al., 2001).

Among the exogenous sources for DSBs are chemical compounds contained in our food, ionizing radiation from natural radiation, ultraviolet light from the sun and chemotherapeutics (Mehta and Haber, 2014). Endogenously, DNA damage is a byproduct of replication (Mehta and Haber, 2014), since the collision of the replication fork with atypical DNA or chromatin topology, RNA polymerases (Aguilera and Gaillard, 2014; Prado and Aguilera, 2005) or DNA binding proteins (Mirkin and Mirkin, 2007) can lead to stalled replication forks and may result in fork collapse. On the other hand, DSBs are deliberately introduced to resolve supercoiled chromatin structures and during the essential cross-overs in the meiotic cell cycle. Moreover, the rearrangement of immunoglobulin and T-cell receptor variable regions relies on DSB induction. Since DSBs are inevitable, cells have evolved a well-orchestrated network for recognition of DSBs and their repair, while cell cycle checkpoints inhibit progression of the cell cycle until the DNA damage is repaired.

## II.2 The protein response to a DSB

### II.2.1 DSB recognition and signaling

The MRN complex (comprised of Rad50, Mre11 and Nibrin (in humans Nbs1)) is acting at the forefront of DSB recognition and subsequent DSB signaling as well as repair. In its function as a DNA damage sensor, the MRN complex interacts with the signaling kinases ataxia-telangiectasia mutated (ATM) and ataxia telangiectasia and Rad3-related protein (ATR) and triggers their activation (Cerosaletti et al., 2006; Lee and Paull, 2004; Uziel et al., 2003). The recruitment of ATM/ATR to the site of a DSB is described as primary recruitment and paves the way for signal amplification and secondary recruitment (Celeste et al., 2003; Lukas et al., 2004). Among the substrates for ATM and ATR is the histone variant H2Ax, which gets phosphorylated at Serine 139. The formation of  $\gamma$ H2Ax is key to the downstream signaling response since it acts as platform for the DNA damage response proteins MDC1 and 53BP1 (Bekker-Jensen et al., 2005; Chapman and Jackson, 2008; Melander et al., 2008), which in turn recruit more MRN/ATM/ATR proteins to the break (Lou et al., 2006). The signal is hereby amplified to an extent that allows cytological visualization of sites of ongoing repair as repair foci. Mammalian ATM and ATR also activate cell cycle checkpoints in response to a break. Both kinases are conserved and are encoded by the *tefu* gene (ATM orthologue) and *mei-41* gene (ATR orthologue) in *D. melanogaster*. In contrast to the mammalian system, where ATM is the key factor in cell cycle checkpoint activation, this function is fulfilled by Mei-41 (dATR) in *Drosophila*. Tefu (dATM) on the other hand plays important roles in telomere stabilization and is a regulator of p53-dependent apoptosis (LaRocque et al., 2007). Like in the mammalian system, *Drosophila* possesses a H2A variant, called H2Av, that is phosphorylated upon DNA damage to enable secondary recruitment of the repair machinery (Madigan et al., 2002).

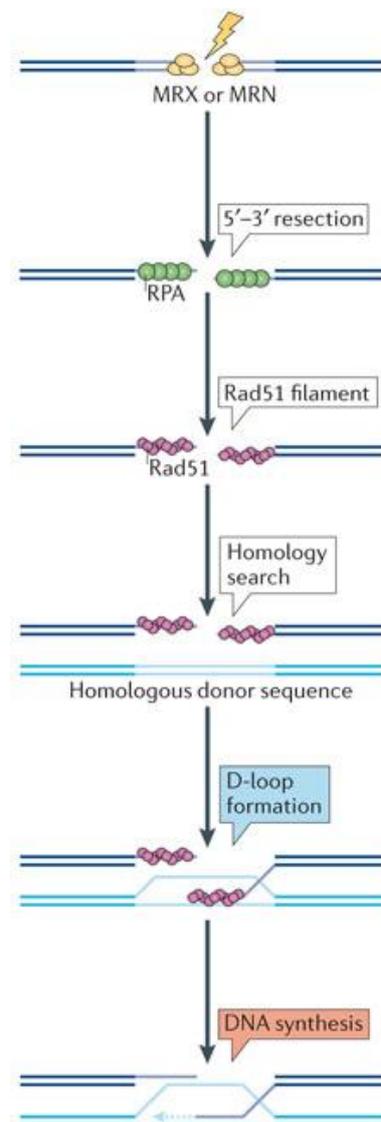
### II.2.2 DSB repair

Broken ends of DNA can be mended by three different pathways for DSB repair: NHEJ (non-homologous end-joining), MMEJ (microhomology-mediated end-joining) and homologous recombination (HR). Several subpathways mediate HR. Among them are gene conversion (GC), single-strand annealing (SSA), break-induced replication (BIR), synthesis-dependent strand annealing (SDSA) and double Holliday junctions (dHJ) (Mehta and Haber, 2014).

### II.2.2.1 HR repair

Of the three major repair pathways, homologous recombination (HR) is often considered to be the most faithful repair mechanism, since the DSB can be repaired using the sister chromatid as template (Renkawitz et al., 2014; Symington and Gautier, 2011). HR is therefore limited to the late S and G2-phase of the cell cycle (Renkawitz et al., 2014; Symington and Gautier, 2011). Homology directed repair (HDR) is always initiated by 5' to 3' end resection by the MRN complex together with the nuclease CtIP (Mehta and Haber, 2014; Renkawitz et al., 2014; Symington and Gautier, 2011). Extensive 5' to 3' resection is mediated by exonuclease-1 (Exo-1) (Mehta and Haber, 2014). ssDNA ends are protected by replication protein A (RPA), which is subsequently exchanged for Rad51 filaments that coat the ssDNA ends (Heyer et al., 2010; Mehta and Haber, 2014; Renkawitz et al., 2014). These filaments are visible as Rad51 foci, and can be used as an indicator of ongoing DSB repair. Rad51 finally mediates homology search and strand invasion. A displacement (D)-loop structure is formed and DNA synthesis can occur to copy the intact template sequence from the sister chromatid (Heyer et al., 2010; Mehta and Haber, 2014; Renkawitz et al., 2014). Depending on how the D-loop structure is resolved, the literature describes different HR subtypes, such as SDSA, dHJ and BIR (Mehta and Haber, 2014; Renkawitz et al., 2014).

In case the DSB occurs between directly repeated sequences, the break can be repaired by the single-strand annealing pathway (SSA). During resection, the homologous regions are exposed and can then anneal. This abrogates the need for homology search and strand invasion. SSA is thus a Rad51-independent repair mechanism (Mehta and Haber, 2014). The single-stranded flaps that are not complementary are eventually removed by a complex comprised of Rad1, Rad10, Msh2, Msh3, Slx4 and Saw1 (Mehta and Haber, 2014). This



**Fig. 1: Common steps in homologous recombination.**

Key steps in HR are initial resection by the MRN complex, which is followed by extensive resection mediated by action of Exo-1, the formation of Rad51 filaments that engage in homology search and strand invasion that results in formation of a D-loop. Figure adapted by permission from Macmillan Publishers Ltd: [Nature Reviews Molecular Cell Biology], (Renkawitz et al., 2014), copyright (2014).

results in the deletion of one copy of the direct repeat, thus SSA is a highly mutagenic subtype of HR.

### II.2.2.2 MMEJ repair

Most interestingly, the alternative end-joining pathway MMEJ also depends on initial resection via the MRN complex. Comparable to SSA, ends are subsequently annealed at microhomologies and the flaps trimmed by Rad1 and Rad10 endonucleases (McVey and Lee, 2008). While homologies in SSA are longer than 30 nt, MMEJ acts on microhomologies of 5-25 nt homology length (McVey and Lee, 2008). The MMEJ pathway is highly error-prone and accompanied by deletions of variable length. MMEJ acts independently from the ‘classical’ end-joining by the NHEJ pathway factors Ku70/Ku80 and DNA ligase 4 (Lig4), but requires DNA polymerase theta (pol theta) (Chan et al., 2010; Kent et al., 2015; Yu and McVey, 2010). This polymerase directly binds the two resected ends (Kent et al., 2015). Upon binding of pol theta, annealing of microhomologies within the overhangs can occur (Kent, Chandramouly et al. 2015). In *Drosophila* pol theta is encoded by the *mus308* gene (Chan et al., 2010). Binding of pol theta to the resected DNA ends also inhibits Rad51-mediated HR repair (Ceccaldi et al., 2015), suggesting crosstalk between the different DSB repair pathway choices.

### II.2.2.3 NHEJ repair

In contrast to HR, which is limited to late S- and G2-phase, NHEJ can act throughout the whole cell cycle and is considered the predominant pathway in the G1-phase. The NHEJ pathway acts independently from Rad51, but requires the Ku-proteins Ku70 and Ku80 and DNA ligase 4. Although NHEJ is often considered as an error-prone mechanism, one possible NHEJ repair outcome is the precise re-joining of short overhanging, complementary ends (e.g. restriction enzyme introduced DSB) (Mehta and Haber, 2014). Another NHEJ subpathway that shows a dependence on the MRN/MRX complex is ‘misalignment’ of partially cohesive ends (Xie et al., 2009). Gaps are filled in by a specialized DNA Polymerase (Polymerase IV), which leads to short insertions at the break site (Wilson and Lieber, 1999). Finally, deletions ranging from a few bp to a few kb can occur if ends are annealed at microhomologies that can be as little as 1 bp (McVey et al., 2004).

### II.3 An RNA response to a DSB

The notion that the cellular response to a DSB is mediated only by proteins has been overcome with the discovery of the DSB-triggered small RNA species in the fungus *Neurospora crassa* in 2009 (Lee et al., 2009). Replication stress, induced by chemical compounds, led to the generation of small RNAs targeting solely the ribosomal DNA locus, the only repetitive locus in the *N. crassa* genome (Lee et al., 2009). The break-triggered small RNAs were named QDE-2-interacting (qi) RNAs (Lee et al., 2009). Although qiRNAs are a few nucleotides shorter (20-21 nt) than other *N. crassa* small interfering RNAs (siRNAs) (~ 25 nt), they share common biogenesis factors. siRNAs in *N. crassa* arise via a phenomenon termed quelling, which was first described in post-transcriptional gene silencing of transgenes (Romano and Macino, 1992) and later in transposon defense (Nolan et al., 2005). Biogenesis factors were retrieved by performing a screen for quelling deficient (QDE) mutants and are named accordingly (Cogoni and Macino, 1997).

The biogenesis of DSB-triggered qiRNAs depends on the Argonaute protein QDE-2, Dicers, as well as the Werner and Bloom RecQ helicase homologue QDE-3 and the factor QDE-1, an enzyme that displays both DNA- and RNA-dependent RNA polymerase activity (Cogoni and Macino, 1999; Lee et al., 2010). Replicative stress leads to the generation of aberrant RNAs from the rDNA locus. QDE-1 acts on those aberrant transcripts via its RNA-dependent RNA polymerase activity, which is strongly promoted in the presence of RPA (Lee et al., 2010). The dsRNA precursor, which is generated by QDE-1 upon DNA damage, is processed into 20-21 nt long qiRNAs. Since insertion of transgenes into *N. crassa* results in a robust quelling response, dissection of the requirements for qiRNA production is possible. Transformation with a truncated albino-1 fragment encoding plasmid, led to qiRNA production, if the transgenes inserted in tandem repeat arrays and a DSB was introduced by either replicative stress or by a site-specific nuclease. DNA damage thus appears to give cells the opportunity to detect repetitive loci, which also can arise from transposon replication (Yang et al., 2015). In contrast to other model organisms, the crosstalk between homologous recombination and damage-induced qiRNAs has been well established. The H3K56 methyltransferase RTT109 facilitates recruitment of Rad51 to DSBs and is required for quelling (Zhang et al., 2014). In addition, a screen for impairment of the replicative stress induced accumulation of QDE-2, showed involvement of the Rad52 protein. Similarly, qiRNA production is abolished in mutants of Rad51 and Rad54 (Zhang et al., 2013). In turn, the qiRNA response promotes HR and helps to maintain tandem repeats within the genome (Yang et al., 2015; Zhang et al., 2013).

Subsequent to the discovery of break-derived small RNAs in *N. crassa*, this phenomenon was described also in the plants *Arabidopsis thaliana* and rice (Chen et al., 2013; Wei

et al., 2012), *Drosophila melanogaster* and human cells (Francia et al., 2012; Michalik et al., 2012; Wei et al., 2012). Given the conservation of a small RNA response to a DSB from fungi to man, one might hypothesize whether their function shows a similar conservation. The role of qiRNAs in *N. crassa* is dual: a DSB in repetitive arrays enables detection of transposable elements and mediates protection from aberrant transcripts (Yang et al., 2015). On the other hand, there is a strong evidence for interplay of HR and break-derived qiRNAs (Lee et al., 2010; Zhang et al., 2013; Zhang et al., 2014).

Similar crosstalk between DSB-triggered small RNAs and homology directed repair has been postulated for *A. thaliana* and the mammalian system. A SSA reporter assay in *A. thaliana* indicated that SSA repair rates were decreased upon depletion of small RNA biogenesis factors or components of the RNA-dependent DNA methylation pathway (Wei et al., 2012). For cultured human cells there are reports that repair kinetics are impaired in the absence of break-derived small RNAs, as indicated by the fact that repair-associated  $\gamma$ -H2Ax foci persist for a longer time (Francia, 2012). A HR reporter was used to study the effect of break-derived small RNAs on repair efficiencies. Knock-down of the factors DICER, DROSHA and AGO2 have resulted in reduced HR repair rates (Gao et al., 2014; Wei et al., 2012). How break-derived small RNAs might contribute to HR is under ongoing investigation. Recently, a protein-protein interaction between AGO2 and Rad51 has been reported (Gao et al., 2014). Another study (Wang and Goldstein, 2016) rather argues that loaded AGO2 recruits the chromatin remodelers Tip60 and MMSET to the break, which in turn, leads to accumulation of Rad51 (Wang and Goldstein, 2016).

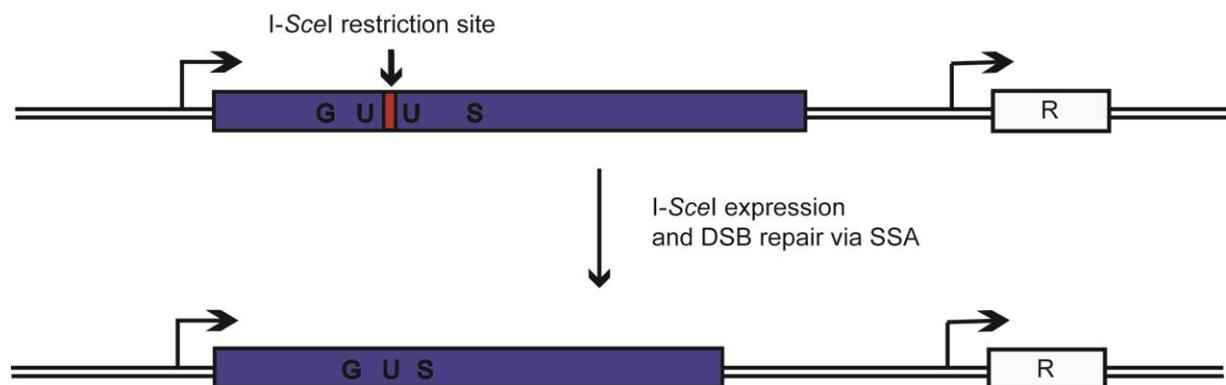
However, there are several caveats to the studies conducted in *A. thaliana* and mammalian cells. The sheer multitude of both small RNA biogenesis factors and small RNAs in *A. thaliana* makes it difficult to dissect the roles of individual factors (Bologna and Voinnet, 2014). The fact that the Dicer like proteins (DCL) 2 - 4 show an effect in the SSA reporter argues for functional redundancy (Wei et al., 2012). In contrast to this, there are two small RNA species described in mammalian somatic cells, microRNAs (miRNAs) and siRNAs. Their biogenesis depends on the same DICER protein. Furthermore, though the mammalian genome encodes four Argonaute proteins, only AGO2 is catalytically active and the major acceptor of both small RNA species. Therefore, it is very difficult to find an experimental setup for depletion of break-derived small RNAs that does not interfere with miRNA biogenesis. In line with this concern, a recent study reports that the observed DSB repair impairment in cultured human cells after depletion of DICER is due to indirect effects (Liu et al., 2015). Depletion of DICER inhibits cell cycle progression and therefore affects HR repair (Liu et al., 2015). However, transfection of a let-7 miRNA mimic restored cell cycle progression and efficient HR repair (Liu et al., 2015).

While *Drosophila* somatic cells also generate miRNAs and siRNAs, the respective biogenesis pathways are well separated. Dicer-1 and Ago1 are designated factors of the microRNA pathway. Dicer-1 processes the pre-miRNA into the miRNA/miRNA\* intermediate by removal of the pre-miRNA loop (Förstemann et al., 2005; Lee et al., 2004). miRNAs are then loaded into Ago1 (Förstemann et al., 2007). Biogenesis of siRNAs on the other hand requires Dicer-2, which dices the long double-stranded RNA precursor that shows perfect complementarity into siRNA duplexes (Förstemann et al., 2005; Lee et al., 2004). The processed 21 nt long siRNAs are finally loaded into the effector protein Ago2 (Liu et al., 2003; Matranga et al., 2005; Tomari et al., 2004a; Tomari et al., 2004b). In accordance with their function, the phenotypes of small RNA factor mutants vary. Flies that carry miRNA factor mutant alleles are not viable (Lee et al., 2004; Okamura et al., 2004). In contrast to this, siRNA factor mutant flies show only a mild phenotype (Lee et al., 2004; Okamura et al., 2004). However, it should be noted that the biogenesis of long hairpin derived endo-siRNAs is also dependent on Dicer-2 (Czech et al., 2008; Okamura et al., 2008). Taken together, we conclude that *Drosophila* is an advantageous model organism to dissect the role of break-derived siRNAs without interfering with the miRNA pathway.

## II.4 Reporter systems to investigate the role of RNA in DSB repair

### II.4.1 SSA reporter

For the *A. thaliana* study, a single-strand annealing reporter was used (Wei et al., 2012). Single-strand annealing is a subtype of HDR that can occur *in cis* between two repeats. Since the resected strands can form a duplex at the complementary region, this type of HDR results in the loss of one repeat, and is thus highly mutagenic (compare also II.2.2.1).

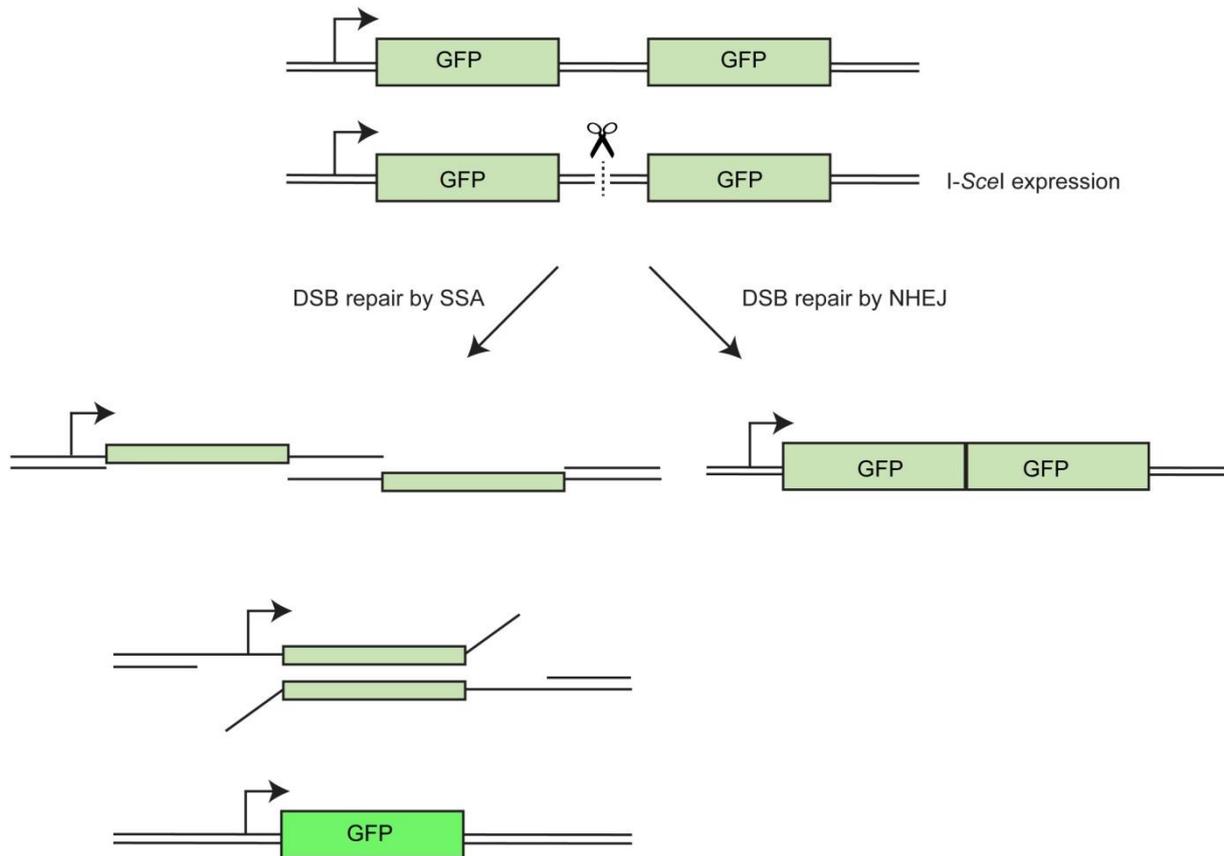


**Fig. 2: The DGUS.US-1 reporter for SSA.** The reporter construct comprises a mutated  $\beta$ -glucuronidase gene in which the internal region is repeated. The repetitive regions are separated by an I-SceI restriction site. The reporter construct in the uncut state is not able to metabolize the substrate X-Gluc. However, if cells undergo SSA after

introduction of a break via the I-*SceI* nuclease, one repeat is lost, thus restoring the  $\beta$ -glucuronidase gene. The substrate X-Gluc can now be metabolized into a blue dye by cells that underwent SSA, allowing for assessment of SSA efficiencies in different genetic backgrounds.

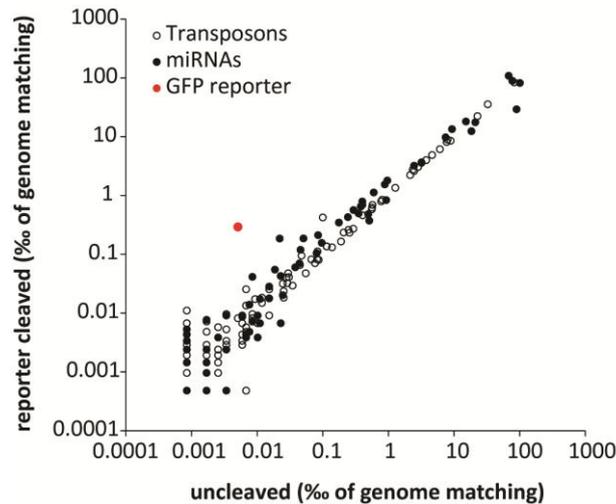
Wei *et al.* deployed an inducible  $\beta$ -glucuronidase reporter system. The DGUS.US-1 reporter is comprised of two halves of the GUS gene, separated by an I-*SceI* restriction site (see Fig. 2). The two halves share a sequence overlap of 557 bp (Mannuss *et al.*, 2010; Orel *et al.*, 2003). The reporter in its initial state does not produce functional  $\beta$ -glucuronidase and X-Gluc cannot be hydrolyzed into glucuronic acid and 5-Brom-4-chlor-indoxyl, which upon oxidation becomes the deeply blue dye 5,5'-Dibrom-4,4'-dichlor-indigo. I-*SceI* is introduced by crossing reporter plants with trigger plants that express I-*SceI*. If cells repair via SSA, the repetitive overlap is removed, which results in expression of functional enzyme (Orel *et al.*, 2003). After staining with X-Gluc, the blue sectors can be counted and serve as readout for the plant's ability to undergo SSA. Using this reporter, Wei *et al.* showed that SSA was severely inhibited in ATR mutants as well as in mutants of DCL 2-4, AGO2 and the RNA polymerases IV and V (Wei *et al.*, 2012). The latter are components of the RNA-dependent DNA methylation pathway. RNA Pol IV is a biogenesis factor for heterochromatin-derived siRNAs. Transcription of RNA Pol V provides a scaffold for binding of heterochromatin-derived siRNAs by sequence complementarity. This in turn is the trigger for recruitment of the RNA-dependent DNA methylation machinery (Matzke and Mosher, 2014).

In order to relate to the study conducted in *A. thaliana*, a SSA reporter was designed and introduced into cultured *Drosophila* cells and transgenic flies prior to this thesis. The reporter is comprised of two copies of GFP that are separated by an I-*SceI* restriction site. Since there is no poly(A) signal after the first copy, the construct in its original state gives rise to transcripts that are targeted by the nonsense-mediated decay (NMD) pathway (Schmidts *et al.*, 2016). NMD in *Drosophila* is triggered by long 3' untranslated regions (UTRs) (Behm-Ansmant *et al.*, 2007; Giuliani *et al.*, 2014). Introduction of a DSB by crossing in I-*SceI* (*in vivo*) or transient transfection of cells with an I-*SceI* encoding plasmid, can either lead to NHEJ or SSA (Fig. 3). NHEJ products still have two copies of GFP present and display, like the uncut reporter, only weak GFP fluorescence. Rearrangement by the SSA pathway, however, results in the loss of one GFP copy. The 3' UTR is thus considerably shortened and the construct is no longer degraded, leading to a substantial increase in GFP fluorescence, which can be quantified by flow cytometry (*in vitro*) or fluorescence microscopy (*in vivo*).



**Fig. 3: A GFP based SSA reporter for *Drosophila melanogaster*.** Analogous to the GUS based SSA reporter in *A. thaliana*, a SSA reporter for *Drosophila* was introduced both *in vitro* and *in vivo*. Two copies of GFP are separated by the restriction site for I-SceI. A poly(A) signal is found downstream of the second copy of GFP. Due to the long 3' UTR, NMD is triggered and results in low fluorescence signal in the uncut reporter. Same holds true for NHEJ repair products. Only upon repair by SSA one copy of GFP is lost, the 3' UTR shortened and the construct is no longer substrate for degradation via NMD, thus resulting in high GFP fluorescence signal.

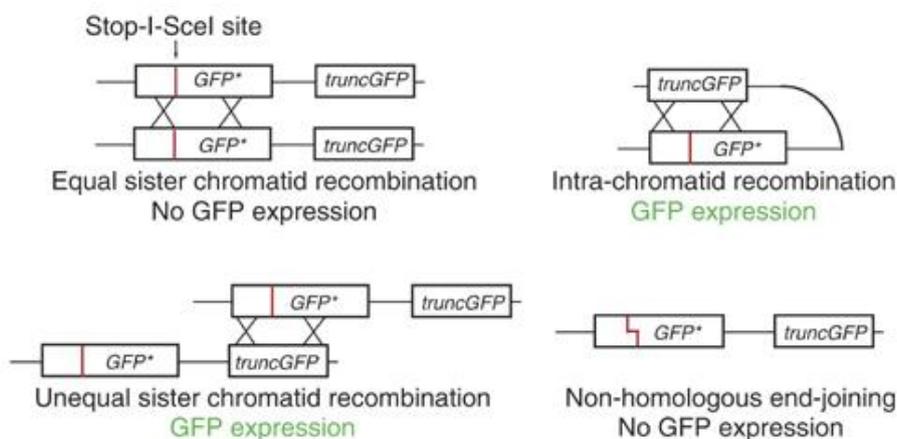
Introduction of a DSB into the SSA reporter is accompanied by the generation of siRNAs against the reporter (see Fig. 4) via the canonical siRNA pathway and loading into Ago2 (Schmidts et al., 2016).



**Fig. 4:** The comparison of small RNA libraries after control treatment or transfection of the *I-SceI* nuclease shows that introduction of a break in the reporter leads to generation of siRNAs against the break. Generation of microRNAs and siRNAs against transposons is unaffected by the break. Figure adapted by permission from Oxford University Press (Schmidts et al., 2016), copyright (2016).

#### II.4.2 Reporters to assess Rad51 homology directed repair

A hallmark of SSA repair is its independence from homology search by Rad51. Wei and colleagues also investigated the role of DSB-triggered small RNAs in the context of Rad51-dependent gene conversion (Francia, 2012; Gao et al., 2014; Wei et al., 2012). The used DR-GFP reporter was introduced by Jasin *et al.* and is comprised of two non-functional GFP copies (Pierce et al., 1999). The GFP coding sequence within the first copy of GFP is disrupted by an *I-SceI* recognition site, while the second copy is truncated and only encodes the internal part of GFP (Pierce et al., 1999). The uncut reporter construct does not yield GFP fluorescence. Upon introduction of the break by *I-SceI*, several repair mechanisms might be deployed for repair (see Fig. 5).

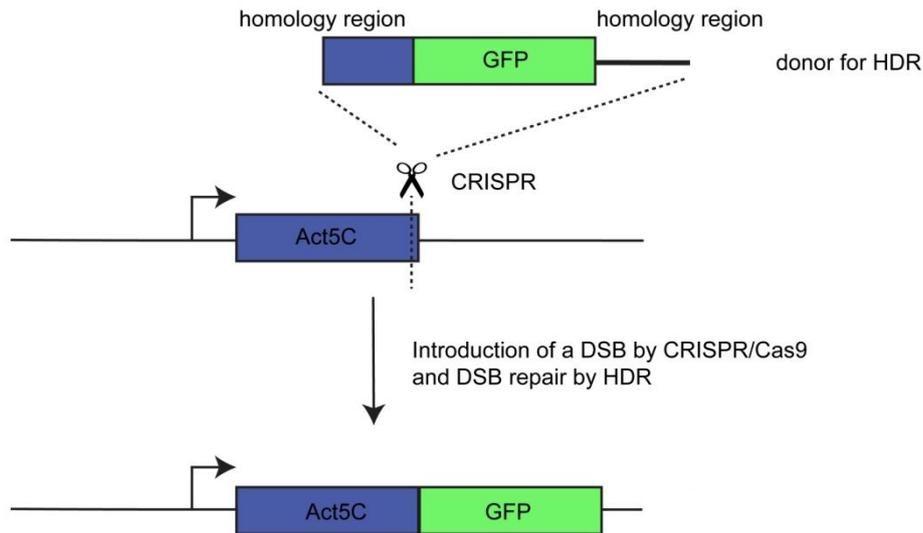


**Fig. 5:** Schematic overview of the short tract gene conversion reporter that was established by the Jasin lab (Pierce et al., 1999) and its potential repair products. The ORF of the first copy of GFP is disrupted by an *I-SceI*

recognition site, while the second copy of GFP is truncated. The uncut construct does not yield GFP fluorescence. Upon introduction of I-SceI, cells can undergo NHEJ (lower, right) which is not able to restore expression of functional GFP. Of the three possible homologous recombination pathways, only intra-chromatid recombination (upper, right) as well as unequal sister chromatid recombination (lower, left) lead to GFP expression. Figure taken from (Potts et al., 2006) with permission from the European Molecular Biology Organization, Copyright (2006).

If cells undergo NHEJ, this might lead to the generation of small indels or result in *bona fide* ligation of the ends. In the latter case, the I-SceI recognition site is restored and the nuclease can cut again. There are three possible scenarios for homology directed repair: equal sister chromatid recombination, unequal sister chromatid recombination and intra-chromatid recombination. Only the latter two result in restoration of a functional GFP copy and measurable GFP fluorescence (Fig. 5). Wei and colleagues reported that depletion of the small RNA biogenesis factors AGO2 and DICER resulted in decreased repair rates of this short-tract gene conversion reporter (Gao et al., 2014; Wei et al., 2012).

In the present thesis, we relate to the study of Rad51-dependent homology directed repair by exploiting our approach to tag proteins at their endogenous locus by CRISPR/*cas9* (see Fig. 6). Cells are transfected with a single guide RNA (sgRNA) expression template that programs Cas9 to introduce a cut in the target gene. We provide also a template for repair, where the tag is flanked by a homology stretch that is homologous to the site upstream and downstream of the DSB. Repair of the DSB via homology directed repair is indicated by successful introduction of the tag at the target site. Introduction of a GFP moiety again allows for readout by flow cytometry (Fig. 6).



**Fig. 6: Exploiting our C-terminal tagging approach to assess Rad51-dependent HDR repair.** A DSB in the highly expressed *act5C* gene is introduced by transfection of an sgRNA expression template into Cas9 expressing *Drosophila* S2 cells. A donor for homology directed repair is co-transfected. A GFP moiety is framed by homology regions up- and downstream of the break. Successful integration of the tag will thus lead to green fluorescence and allows readout by flow cytometry.

## II.5 Transcription and the break

The link between transcription and DNA damage is two-sided. It has become increasingly evident that transcription itself poses a great threat to genome integrity. Mainly by inducing replicative stress, since the progressing transcription bubble invokes R-loop structures. Those are defined by a RNA:DNA hybrid that forms between the nascent transcript and the coding strand while the non-template strand is displaced and forms a loop. Usually those structures are dissolved with progression of transcription, but stalled or paused transcription increases the likelihood of collision between the transcription and replication machinery, leading to replication fork stalling and possibly even collapse. R-Loop formation is promoted at sites with a high GC content, a hallmark of promoter-proximal regions as well as 3' untranslated regions. At those sites, R-loop formation is even required to facilitate initiation or termination, respectively (Tresini et al., 2016). Although DNA lesions are considered to be an undesired byproduct of transcription, a recent publication suggests that DNA breaks are deliberately introduced into mammalian, inducible genes and are required for processive elongation by RNA polymerase II (RNAPII) (Bunch et al., 2015). Splicing, on the other hand, acts inhibitory on R-loop formation by the steric hindrance of the splicing machinery and the reduced complementarity of spliced RNA to its DNA template. R-loops are additionally destabilized by all RNA processing steps,

since the transcript is incorporated into ribonucleotide particles. Furthermore, RNase H enzymes are specialized to resolve DNA:RNA hybrids (Tresini et al., 2016).

When a DSB occurs, aberrant transcripts originating from the damaged locus might pose a threat to the cells. Accordingly, cells have evolved mechanisms to inhibit transcription from damaged loci. In response to DNA damage, e.g. by UV radiation, Rpb1 (the biggest subunit of RNA polymerase II) is ubiquitinated in human cells (Bregman et al., 1996) and degraded by the proteasome (Anindya et al., 2007). Likewise, if DNA damage occurs in RNAPI transcribed genes, RNAPI-mediated rDNA transcription is inhibited (Kruhlak et al., 2007). Transcriptional silencing was found to be dependent on DSB signaling factors, such as ATM or DNA-PKcs, both for RNAPI and RNAPII transcribed loci (Iannelli et al., 2017; Kruhlak et al., 2007; Pankotai et al., 2012; Shanbhag et al., 2010). Depending on the reports and the respective experimental setup, transcriptional silencing is restricted to the damaged loci (Pankotai et al., 2012) (Iannelli et al., 2017; Kim et al., 2016) or spreads to several kilobases (Shanbhag et al., 2010).

Although it seems contradicting given the reports on DNA damage-induced transcriptional silencing, the break itself might act as a bidirectional promoter. Since *Drosophila* and mammalian cells lack an RNA-dependent RNA polymerase (RdRP), it can only be concluded that there is active transcription to generate the antisense transcript that finally forms the dsRNA precursor required for siRNA biogenesis (Michalik et al., 2012). This is in accordance with reports about RNAPII displaying a high affinity for damaged DNA ends *in vitro* (Dynan and Burgess, 1979; Dynan and Burgess, 1981) and its recruitment to break sites in *S.pombe* (Ohle et al., 2016). This and the observation that DSB-derived siRNAs only occurred if the break was located within a transcribed region in *Drosophila* (Michalik et al., 2012), makes it very likely that RNAPII plays a pivotal role in the biogenesis of long double-stranded RNA, the precursor molecule for siRNA production. So far, there is no report on the mechanism of this early step in DSB-triggered small RNA production.

## II.6 Objectives of the present study

The occurrence of DSB-derived siRNAs has been well established and argues unmistakably for the biogenesis of an antisense transcript from the break site to form the dsRNA progenitor. Although the genetic requirement for enzymes with RdRP activity in *N. crassa* and *A. thaliana* gives insight into how the dsRNA progenitor is generated in those organisms, the antisense transcripts themselves have never been directly detected in any of the published studies. The question of how and by which enzyme those antisense transcripts are generated is even more intriguing in model systems that lack a canonical RdRP, such as insects and mammalia. Some

reports describe that both, human RNAPII and human telomerase reverse transcriptase (TERT), might have residual RdRP activity (Lehmann et al., 2007; Maida et al., 2009; Wagner et al., 2013), but whether this holds true in the context of DNA damage is still an unresolved issue. In addition to that, human TERT is solely expressed in germ and stem/progenitor cells at a very low level of 1-5 mRNA copies per cell (Cong et al., 2002; Yi et al., 1999). Therefore it is inconceivable that human TERT would be sufficient to mediate a small RNA response to a DSB. It is very likely that antisense transcription from the break is a rather rare event and/or antisense transcripts are very rapidly processed into siRNAs, thus their detection by a classical RNA-Seq approach might prove rather difficult.

To gain insight into the RNAPII recruitment to the break, we want to use nascent elongating transcript sequencing (NET-Seq). This technique captures RNA polymerase II by immunoprecipitation (IP) and enables for specific enrichment of nascent transcripts, which are extracted and used for library preparation. Mapping of the reads allows for recapitulation of transcription events. We wanted to compare samples, where a DSB was introduced by CRISPR/*cas9* to uncut samples to learn about transcription events at the break.

So far, NET-Seq has been established in yeast and human cells (Churchman and Weissman, 2011; Churchman and Weissman, 2012; Mayer and Churchman, 2016; Mayer et al., 2015; Nojima et al., 2016; Nojima et al., 2015), but not yet in insect cells, although there are modified approaches that enrich for nuclear transcripts via fractionation and thereby circumvent immunoprecipitation of RNA polymerase II (Ferrari et al., 2013; Weber et al., 2014). In addition, NET-Seq has not been used to monitor transcription at the break, as of yet.

In a first step, we generated tagged cell lines to facilitate immunoprecipitation of RNA polymerase II. Since our 'standard' laboratory lysis protocol failed to retrieve nuclear RNA polymerase II, we next optimized the lysis protocol. Further focus of this study was the comparison of different tags and elution methods with the aim to find conditions where the isolated RNA shows a good signal-to-noise ratio. Eventually, we generated NET-Seq libraries after inducing a DSB in either the *CG15098*, *Tctp* or the *7SK* locus. NET-Seq from untreated cells served as control. Although we observe an increased fraction of antisense reads to all gene matching reads upon induction of a DSB in the respective gene (with exception of *Tctp*), the preliminary data is not sufficient to obtain an antisense transcription profile.

In a second project, we tested the current hypothesis that siRNAs function in HR repair in *D. melanogaster*. *Drosophila* is the ideal model organism to investigate the role of break-derived siRNAs, since depletion of the biogenesis factors Dcr-2 and Ago2 does not lead to a deregulation of miRNA biogenesis. This is a major advantage, especially compared to the mammalian system, where only one Dicer and one functional Argonaute protein mediate

biogenesis of both siRNAs and miRNAs. Thus, depletion of those factors results in side effects. In order to relate to the studies conducted in *A. thaliana*, we established a SSA reporter *in vitro* and *in vivo*. Since SSA is a very special HR subtype, we also investigated Rad51-dependent homology directed repair by knock-in of GFP exploiting our CRISPR/*cas9* approach to tag proteins at their endogenous locus. Finally, we wanted to test whether depletion of RNAi factors would render fruit flies more susceptible to DSB stress induced by the topoisomerase inhibitor camptothecin.

---

---

### III Results

#### III.1 Antisense transcription at the break

Studies in several model organisms have shown that there is a small RNA response to a DSB. This is the case from fungi to human cells (Francia, 2012; Lee et al., 2009; Michalik et al., 2012; Wei et al., 2012). Yet the biogenesis, especially its early steps, is elusive. For the fungus *N. crassa* and for the plant *A. thaliana* DSB-triggered siRNA production depends on the presence of an RNA-dependent RNA polymerase. This polymerase can synthesize a dsRNA molecule from aberrant transcripts, which then can be processed into siRNAs by the action of Dicer proteins. Mammalian and *Drosophila* cells do not encode a recognizable RNA-dependent RNA polymerase in their genome, thus the biogenesis of DSB-triggered siRNAs is at this point enigmatic. *In vitro* studies showed that RNAPII has a high affinity for DSB ends (Dyner and Burgess, 1979; Dyner and Burgess, 1981). In line with this observation, RNAPII was found to be recruited to DSB ends in the model organism *S. pombe* (Ohle et al., 2016). Furthermore, active transcription is a prerequisite for the formation of DSB-induced siRNA in *Drosophila* (Michalik et al., 2012). Therefore, we hypothesized that the break might function as a bidirectional promoter (Michalik et al., 2012).

To test this hypothesis, we aimed to establish nascent elongating transcript (NET)-Sequencing (Churchman and Weissman, 2011; Churchman and Weissman, 2012). This technique gives insight into active RNAPII transcription events at a genome wide level. In case of a recruitment of RNAPII to a DSB we would observe antisense transcription specifically upon induction of a break. In addition, the immunoprecipitation step of the NET-Seq protocol can be adapted for other RNA polymerases and thereby allows investigating whether DSB-induced antisense transcripts are generated by a dedicated polymerase or whether the respective locus-associated polymerase fulfills this function. To this end, RNAPIII NET-Seq was established in the laboratory in parallel to this study (Gatz, 2017).

### III.1.1 Establishing a NET-Seq protocol for *Drosophila* S2 cells

The NET-seq protocol was originally established for *S. cerevisiae*, but has very recently been adapted to monitor RNAPII transcription in cultured human cells (Churchman and Weissman, 2011; Churchman and Weissman, 2012; Mayer and Churchman, 2016; Mayer et al., 2015; Nojima et al., 2016; Nojima et al., 2015). So far, there is no report of a NET-Seq protocol for *Drosophila*. We thus decided to adopt and optimize all the essential steps of the different published NET-Seq protocols for *Drosophila* S2 cells.

The first step is the lysis of cells with special focus on the release of actively engaged RNAPII from chromatin. In a second step immunoprecipitation is performed. In lack of a suitable antibody that allows for immunoprecipitation of the endogenous RNAPII complex, we exploited our CRISPR/*cas9* approach to tag proteins at the C-terminus (Böttcher et al., 2014; Kunzelmann et al., 2016)(see III.1.1.1.1 and III.1.1.1.2). Introduction of a tag at RNAPII subunits also offers the possibility for specific elution of the complex (see III.1.1.3 and III.1.1.4), which should result in an increased signal-to-noise ratio in the downstream analysis. Following immunoprecipitation, co-purified RNA is released from protein by Proteinase K digestion, further purified, size-selected and incorporated into an RNA library. We decided to make use of the previously established small RNA library preparation protocol (Elmer et al., 2014) with some minor adaptations, e.g. in size-selection. The protocol was initially introduced for deep sequencing of microRNAs (Elmer et al., 2014) and information of the strand orientation is conserved during library preparation, which allows for downstream analysis of sense and antisense transcription.

#### III.1.1.1 Tagging of RNA polymerase II subunits

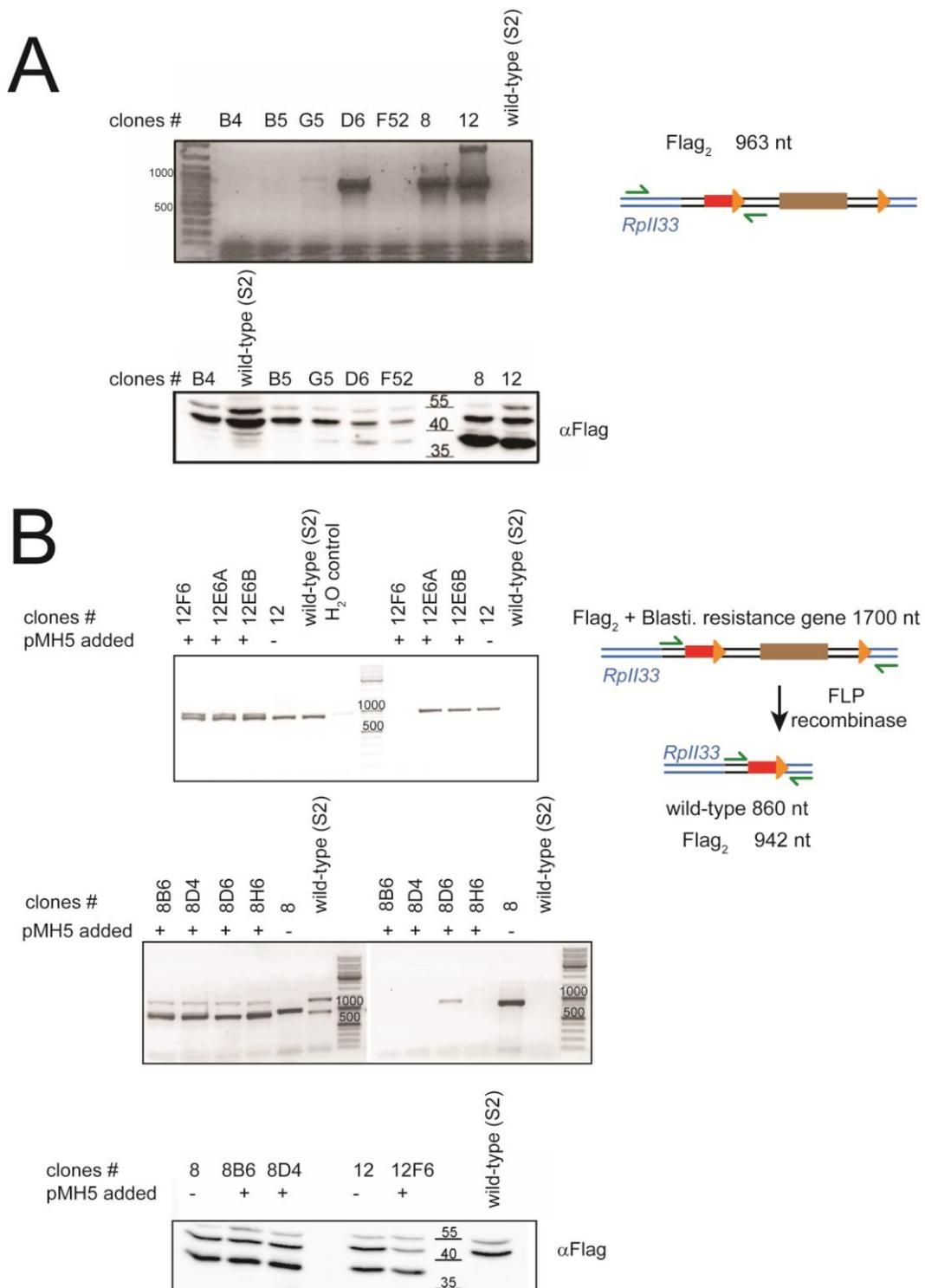
The C-terminal domain (CTD) of RNA polymerase II is a central hub for transcriptional regulation by phosphorylation of the repetitive heptapeptide YSPTSPS (Eick and Geyer, 2013). The CTD of *S. cerevisiae* RNAPII has 26 heptad repeats, 19 of which show the consensus heptad sequence. The CTD of mammalian RNAPII comprises 52 heptad repeats, of which 31 deviate from the consensus sequence (Chapman et al., 2005; Eick and Geyer, 2013). In contrast to this, only two of the 45 heptad repeats of *Drosophila* RNAPII show conservation of the consensus sequence (Mayfield et al., 2017). In many cases variation of the consensus heptad sequence does not disrupt CTD function since e.g. serin is mutated to threonin, an acidic mimic such as glutamic acid or to the non-phospho acceptor alanine (Eick and Geyer, 2013; Mayfield et al., 2017).

---

Since most available antibodies against endogenous RNAPII are directed against the consensus heptapeptide, detection and immunoprecipitation of *Drosophila* RNAPII is assumed to be difficult. Use of the Pol 3-3 antibody, provided by the Eick laboratory, allowed us to detect endogenous RNAPII, but this antibody is not suited for immunoprecipitation. Using the commercial antibody 8WG16 for immunoprecipitation yielded ambiguous results. To facilitate immunoprecipitation, we added C-terminal tags to the RNAPII subunits RpII33 and RpII215.

#### III.1.1.1.1 Tagging of *RpII33*

It has been well established that immunoprecipitation of the third biggest of the 12 subunits of the RNAPII complex (Rpb3 in human cells, RpII33 in *Drosophila*) retrieves the whole, intact complex (Churchman and Weissman, 2011; Churchman and Weissman, 2012). We thus decided to add a Flag<sub>2</sub>-tag to the C-terminus of RpII33 using our CRISPR/*cas9* tagging protocol (Böttcher et al., 2014). In parallel to experiments to optimize lysis conditions, we conducted a serial cell dilution and retrieved a number of clones that showed integration of the tag at the correct site, as well as a strong protein signal in the Western blots (see Fig. 7 A). To enrich for cells that showed integration of the tag, our HR donor also carries a selection cassette, in this case the blasticidin resistance gene under control of the *copia* promoter. Successful integration thus disrupts the endogenous 3' UTR. Since the selection cassette is flanked by FRT sites, the endogenous 3' UTR can be restored by transient transfection of an FLP recombinase expression plasmid. After transfection of pMH5, a plasmid encoding the FLP recombinase, cells were again cloned. From the serial cell dilution we finally retrieved RpII33-Flag<sub>2</sub> clones 8B6 and 12D4 (see also Fig. 7 B).



**Fig. 7: C-terminal tagging of *Rpl133*.** A C-terminal Flag<sub>2</sub>-tag was introduced to *Rpl133*, which encodes the third-largest subunit of RNA polymerase II. By single cell cloning, *Rpl133*-Flag<sub>2</sub> clones were retrieved and subjected to screening by PCR and Western blotting. **(A)** A PCR was conducted on purified gDNA of the indicated cell lines using a sense primers that lies within the *Rpl133* ORF in combination with an antisense primer that binds to the *copII* promoter. A band of 963 nt is indicative of successful integration of the tag. Western blotting (lower panel) confirmed the PCR results, and strongest Western blot signals were observed for clones 8 and 12. We thus selected those clonal cell lines for transient transfection with the FLP-recombinase encoding plasmid pMH5 and subjected

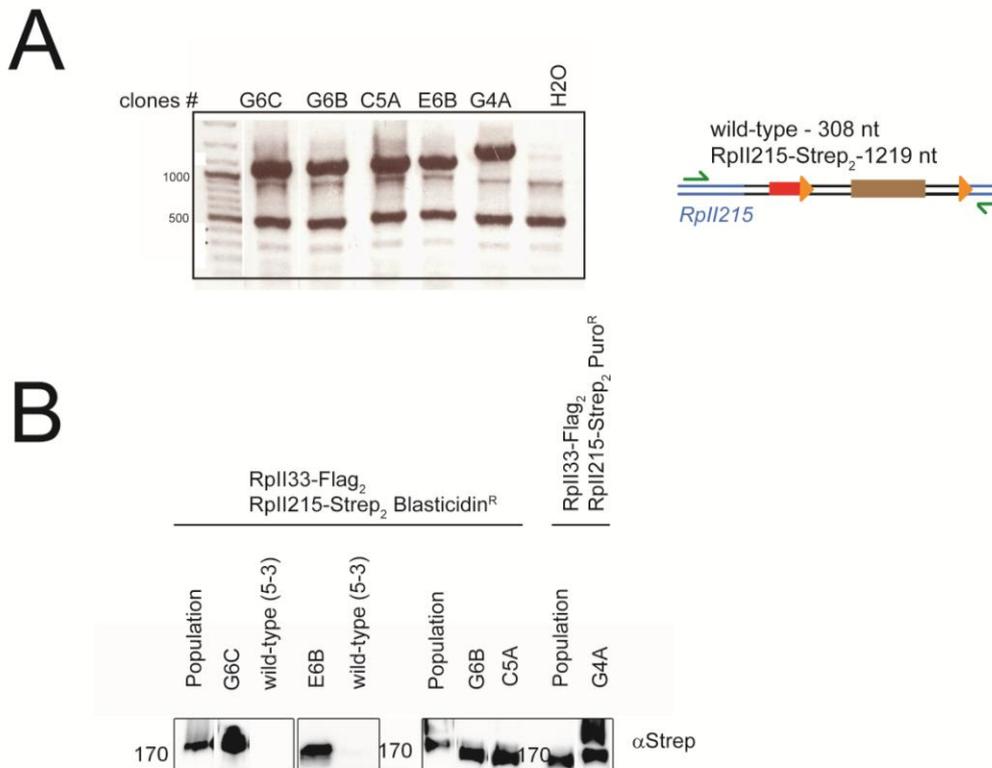
---

cells to another round of clonal dilution. We screened the retrieved clones again by PCR. **(B)** Left side of the respective PCR gels: A PCR using primers within the *RpII33*-ORF and the *RpII33* 3' UTR allows us to discern untagged alleles (860 nt) from tagged alleles (942 nt). A second PCR was performed with one primer binding within the *RpII33* ORF and the other within the *copia* promoter. The latter PCR (right side of the respective PCR gels) allows screening for the absence of the *copia* promoter and identified clones 8B6, 8D4 and 12F6 to have undergone recombination of all tagged alleles. Western blot analysis confirmed the presence of Flag-tagged RpII33 in the selected resistance marker free clonal cell lines.

### III.1.1.1.2 Tagging of *RpII215*

Modifications that disturb the C-terminal domain of RNAPII might interfere with the process of transcription. However, unpublished experiments indicate that transgenic flies with a tag at the end of the CTD are viable and do not show phenotypic impairment. Furthermore, this may not alter the structure of the CTD of the fusion protein (Portz et al., 2017). We thus decided to modify also the *RpII215* gene.

The cell line RpII33-Flag<sub>2</sub> clone 8B6 was used as parental cell line for tagging of the biggest subunit (215 kDa) RpII215 with a Twin-Strep-tag. Since we removed the blasticidin resistance cassette by the Flp-FRT system to generate RpII33-Flag<sub>2</sub> clone 8B6 (Fig. 8), we used in parallel HR donors containing a blasticidin resistance cassette (pIW1 as template for the HR donor PCR) or a puromycin resistance cassette (pSK24 as template for the HR donor PCR). After a serial dilution, we eventually retrieved the puromycin resistant clone G4A and the blasticidin resistant clones G6C, G4B, C5A and E6B (see also Fig. 8). The Twin-Strep tag allows for specific elution by competition with biotin for binding to Strep-Tactin beads. We optimized (see III.1.1.4) and used this approach to prepare NET-Seq libraries after introduction of DSBs at different sites in the *Drosophila* genome.



**Fig. 8: C-terminal tagging of *RpII215*.** The clonal cell line RpII33-Flag<sub>2</sub> clone 8B6 was used as a parental cell line for tagging of the C-terminus of *RpII215* with a Twin-Strep tag to allow for specific elution. Tagging was performed in parallel with a HR donor comprising the blasticidin resistance cassette (pIW1 as donor) and a HR donor comprising the puromycin resistance cassette. After single cell dilution, RpII33-Flag<sub>2</sub> RpII215-Strep<sub>2</sub> clones were retrieved and **(A)** subjected to PCR (upper panel). A band of 308 nt would be indicative of a wild-type allele, but was not detected. For *RpII215-Strep<sub>2</sub>* alleles a band of 1219 nt is expected and is detected for all the depicted clones. We thus conclude that all *RpII215* alleles were modified in four clones that carry a blasticidin resistance, and the one clone (G4A) that carries the puromycin resistance gene. **(B)** We also performed Western blotting to validate the presence of functional tagged protein. Shown above are only clonal cell lines that screened positive in both, PCR and Western blotting, as well as corresponding controls.

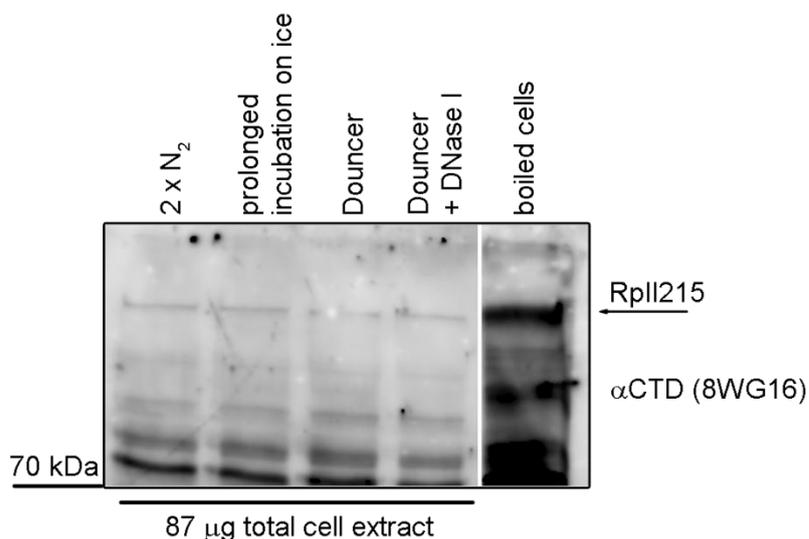
### III.1.1.2 Optimization of RNA polymerase II extraction from *Drosophila* S2 cells

It should be noted that tagging of RNAPII subunits (see III.1.1.1.1 and III.1.1.1.2) and optimization of RNAPII extraction from *Drosophila* S2 cells were performed in parallel. Therefore, we used wild-type S2 cells for initial lysis optimization.

For the optimization of cell lysis, we started with a HEPES based buffer (30 mM HEPES pH7.5, 100 mM KoAc, 2 mM Mg(OAc)<sub>2</sub>, 1 mM DTT, 1x Proteinase Inhibitor without EDTA (Roche)) and Triton-X (1 %) as detergent. To efficiently lyse the resuspended cells, we subjected them to a snap-freezing step. The forming ice crystals lead to disruption of the cell membrane (Steponkus et al., 1983). A centrifugation step separates extracted, soluble proteins from cell debris. The above mentioned lysis buffer and snap-freeze protocol have been

previously established in the laboratory. However, while the described protocol retrieves cytoplasmic proteins very efficiently, we could not detect bands corresponding to RplI215 by using  $\alpha$ Rpb1 antibodies (data not shown). This indicates that the nuclear membrane was not disrupted.

We thus tested next whether we can extract RNAPII by adding or increasing physical disruption steps within our protocol. To this end, we either repeated the freeze-thaw cycle or increased the incubation time on ice. Additionally, we investigated whether mechanical disruption with a Dounce homogenizer would facilitate the release of RplI215 from the nucleus. In order to establish a robust NET-Seq protocol, it is essential to purify actively engaged (i.e. chromatin bound) RNAPII. This is why we additionally tested mechanical disruption using a Dounce homogenizer in the presence of 10 Units/ml of DNase I to determine whether enzymatic disruption would increase extraction of RNAPII from the chromatin (see Fig. 9).



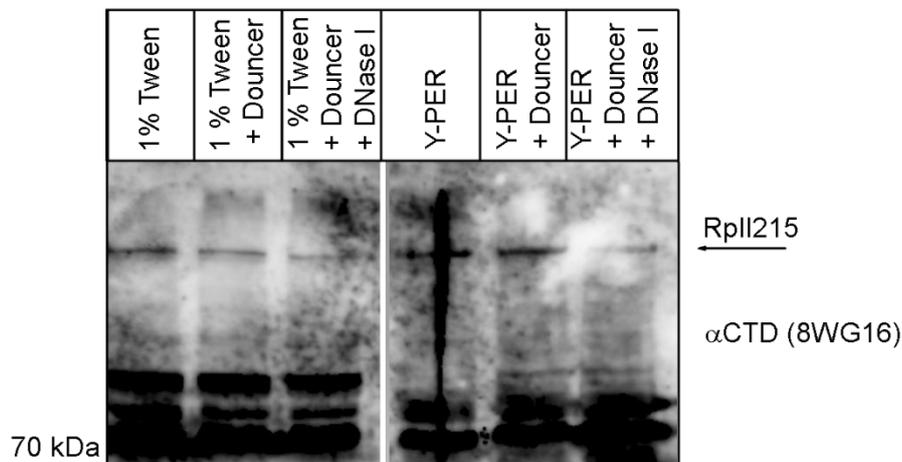
**Fig. 9: The ‘standard’ laboratory protocol was modified to test the influence of different disruption strategies on RNAPII release from the nucleus. (2x N<sub>2</sub>)** the snap-freezing step (snap freezing of resuspended cells in liquid nitrogen and subsequent thawing at room temperature was repeated two times. Following this step, cells were incubated on ice for 10 min and handled as described in V.2.2.2.1). **(Prolonged incubation on ice)** instead of an incubation of 10 min on ice subsequent to the snap-freeze steps, the cell extract was incubated a total of 40 min. **(Douncer)** Following a single snap-freeze step, the cell extract was homogenized with 15 strokes in a Dounce homogenizer. **(Douncer + DNase I)** lysis was performed as described for (Douncer), with the addition of DNase I to a final concentration of 10 U/ml.

Since neither physical disruption nor additional DNase I treatment resulted in increased RNAPII isolation efficiency, we next tested the influence of detergents on extraction of RNAPII.

Since the NET-Seq protocol relies on structural integrity of the RNAPII complex to protect nascent transcripts during immunoprecipitation, we focused on mild, non-ionic detergents. Non-

ionic detergents do not denature proteins, but influence lipid-lipid and lipid-protein interactions, while leaving protein-protein interactions intact (Seddon et al., 2004).

In addition to testing Triton-X-100 as detergent (see Fig. 9 and Fig. 11), we supplemented the ‘standard’ lysis buffer with 1 % Tween-20 and compared it to the commercially available Y-PER reagent, which is a Tris-based buffer with low ionic strength and a proprietary detergent. But even in combination with a Dounce homogenizer  $\pm$  DNaseI, the extraction efficiency was rather poor (compare Fig. 10 to Fig. 9).



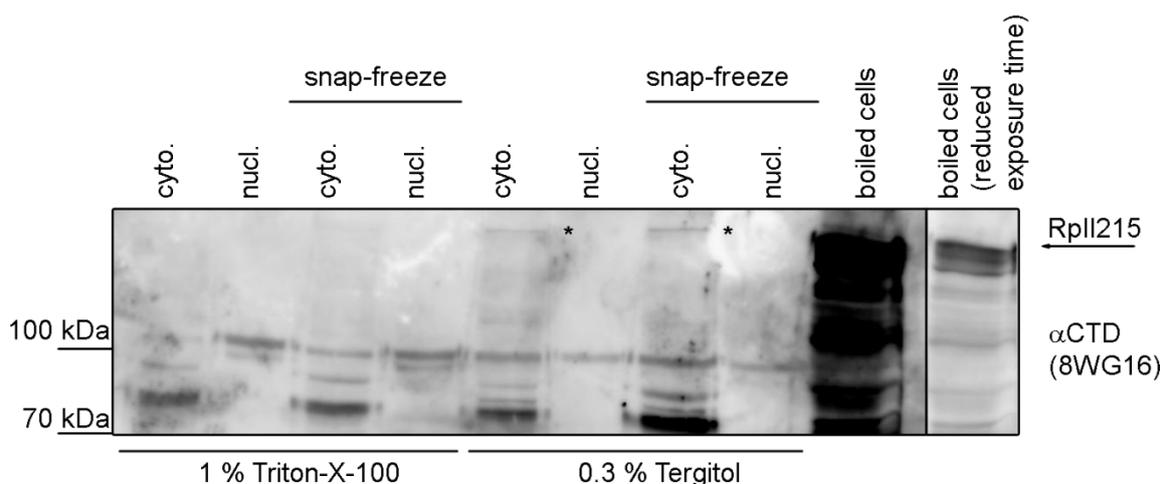
**Fig. 10: Investigating the influence of different detergents on RNAPII extraction from the nucleus.** The lysis buffer was composed as described above, but supplemented with 1 % Tween instead of 1 % Triton-X-100. Extraction steps were performed as described above and we additionally assessed whether mechanical disruption with 15 strokes with a Dounce homogenizer and the presence/absence of DNase I would facilitate RNAPII recovery from the nucleus. We tested the same conditions in parallel with the commercially available lysis reagent Y-PER.

While the plasma membrane is easily disrupted by physical means, this is not the case for the nuclear membrane. The inner nuclear membrane is also not solubilized by most non-ionic detergents, such as Triton-X-100 (Brawerman, 1974). Tight association to the chromatin, as it is the case for actively engaged RNAPII complexes, further impairs extraction, as untreated chromatin is insoluble in most nuclear extraction protocols.

We therefore turned to cellular fractionation protocols, which make use of the mentioned properties. A hypotonic buffer is used (often in combination with additional physical disruption) to permeabilize the plasma membrane. Centrifugation yields the cytoplasmic extract (supernatant (SN)) and intact nuclei (pellet). The isolated nuclei are resuspended in a hypertonic buffer (often supplemented with a strong detergent to permabilize the nuclear membrane, e.g.

Tergitol-type NP-40). High salt facilitates disruption of the nuclear membrane and forces DNA into solution.

To test this lysis approach, we adapted the Abcam Nuclear Fractionation protocol and compared the influence of different detergents on RNAPII extraction efficiency (1 % Triton-X-100 was compared to 0.3 % Tergitol-type NP-40). Cell pellets were first resuspended in a hypotonic buffer (10 mM HEPES pH 7.5, 1.5 mM MgCl<sub>2</sub>, 10 mM KCl, 1 mM DTT, 1 % Triton-X-100 or 0.3 % Tergitol-type NP-40, 1x Proteinase Inhibitor without EDTA (Roche)). Resuspended cells were then subjected to a freeze-thaw cycle (as described above) or directly incubated on ice for 20 min (according to Abcam's protocol). Subsequently, the lysate was cleared by centrifugation and the supernatant taken as the cytosolic fraction. In a next step, the pellet was resuspended in a high salt buffer (5 mM HEPES, pH 7.5, 1.5 mM MgCl<sub>2</sub>, 1 mM DTT, 10 % Glycerol, 1x Proteinase Inhibitor without EDTA (Roche)). NaCl was added to a final concentration of 300 mM. The extract was further homogenized by 20 full strokes in a Dounce homogenizer on ice. Samples were incubated for an additional 20 min on ice, and then centrifuged at top speed, 4 °C for 20 min.

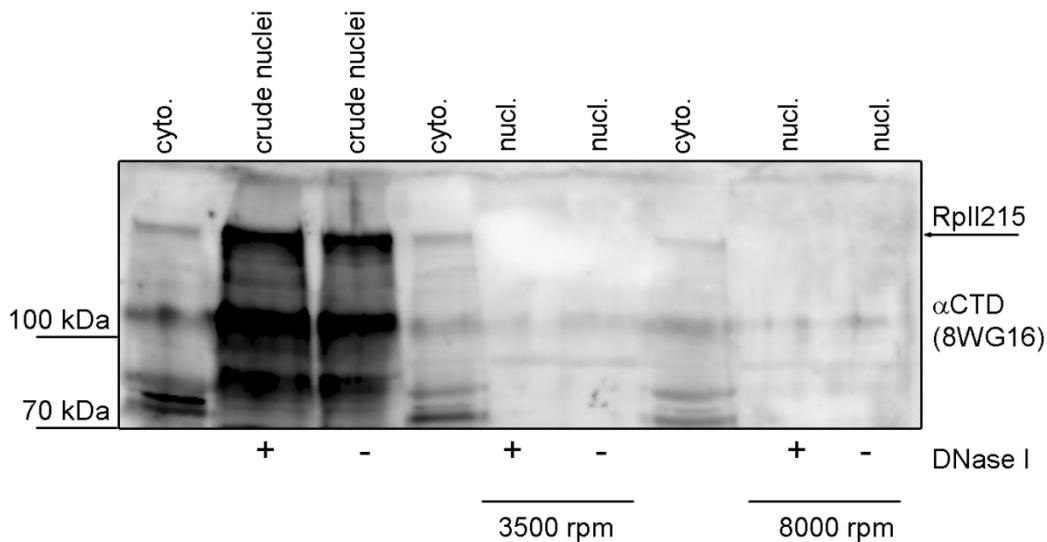


**Fig. 11: Cellular fractionation and comparison of the influence of different detergents on RNAPII extraction from the nucleus.** A cellular fractionation protocol from Abcam was either performed according to the Abcam protocol or with addition of a freeze-thaw cycle. The cytosolic (**cyto.**) and nuclear fractions (**nucl.**) were loaded on an 8 % SDS-PAGE and subjected to Western blotting. In addition, we tested both fractionation protocols using either Triton-X-100 (1 % final concentration) or Tergitol-type NP-40 (0.3 % final concentration) as detergent. Comparison to boiled cells allows to assess RNAPII extraction efficiency qualitatively. Cytosolic fractions showed a weak band for RplI215 when Tergitol-type NP-40 was used as detergent. However, no RplI215 could be retrieved in the nuclear fractions.

Addition of Tergitol-type NP-40 disrupted the nuclear membrane and thereby led to leakage of RplI215 into the cytosolic fraction. In contrast, we could not detect bands corresponding to

RpII215 when we used Triton-X-100 as a detergent. This is in accordance with the observation that Triton-X-100 treatment does not perturb the nuclear membrane (Brawerman, 1974). We thus decided to use Tergitol-type NP-40 as detergent from this point on.

While a weak RpII215 signal was visible in the cytosolic fraction, we could not retrieve RpII215 in the nuclear fraction. RNAPII appears to be tightly associated with the chromatin. Therefore, we reasoned that the final centrifugation step of the nuclear fraction might pellet chromatin as well as chromatin-associated proteins.

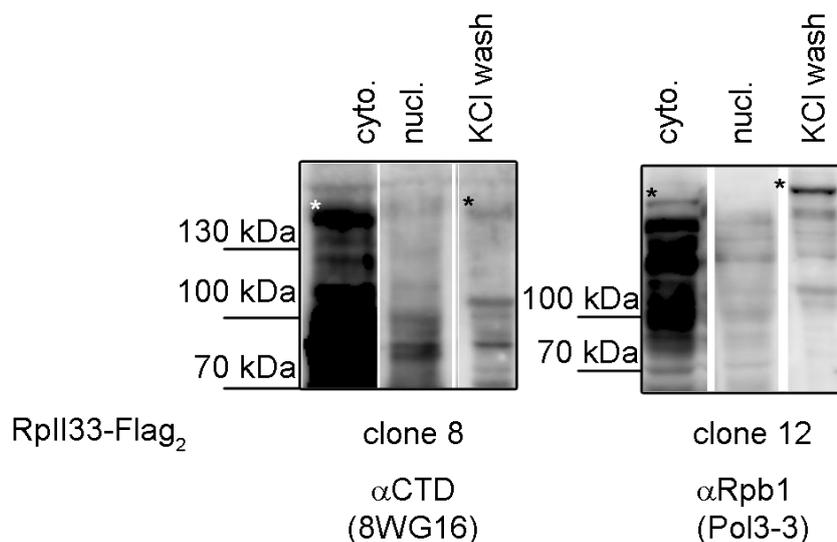


**Fig. 12: RNAPII is tightly associated to chromatin and pelleted by centrifugation of nuclear fractions.** RpII215 bands were readily visible in crude nuclei (nuclear extracts prior to centrifugation). A reduction of centrifugation speed did not circumvent pelleting RNAPII and additional DNase I treatment did not facilitate RNAPII release from chromatin.

This could be confirmed when crude (i.e. not centrifuged) nuclear extract was included on the Western blot. A band corresponding to RpII215 was readily detectable in the crude extracts, however centrifugation results in pelleting RNAPII. Additional DNase I treatment (3 U Dnase I) of the nuclear extract did not digest chromatin sufficiently enough to prevent pelleting of RNAPII. Reducing duration (10 instead of 20 min) and speed of the final centrifugation step did also not alleviate the problem (see Fig. 12).

KCl has been reported to facilitate extraction of chromatin associated proteins, and as suggested for isolation of chromatin components of mammalian cells (Schnitzler, 2001), we resuspended nuclear fraction pellets in an additional chromatin-extraction buffer (20 mM HEPES pH 7.5, 3 mM MgCl<sub>2</sub>, 1x Proteinase Inhibitor without EDTA (Roche)). We then added an equal volume of high KCl chromatin-extraction buffer (0.6 M KCl, 20 mM HEPES pH 7.5,

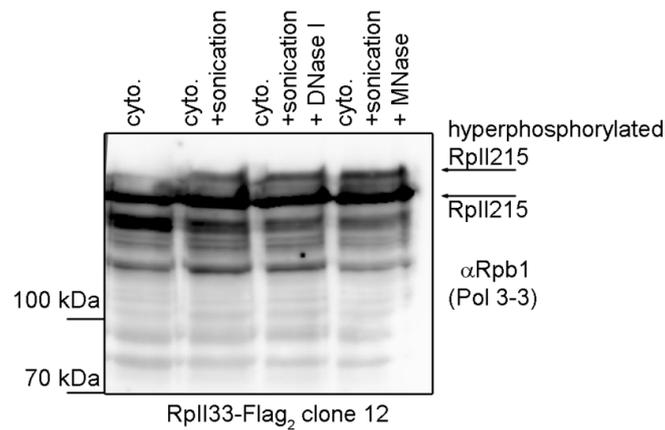
3 mM MgCl<sub>2</sub>, 1x Proteinase Inhibitor without EDTA (Roche), 10 % glycerol) in a dropwise manner, which is thought to release soluble proteins (Abmayr et al., 2001).



**Fig. 13: Chromatin-wash of the nuclear fraction using a high KCl buffer facilitates RNAPII release from chromatin.** Cytosolic and nuclear fractions were prepared as described above. The chromatin pellet was subsequently resuspended in buffer C (with a final composition of 20 mM HEPES pH 7.5, 3 mM MgCl<sub>2</sub>, 1x Proteinase Inhibitor without EDTA (Roche), 0.3 M KCl, 5 % glycerol.) The KCl wash was performed with extracts from RplI33-Flag<sub>2</sub> clone 8 and clone 12. The comparison shows that the efficiency of the KCl wash varies. Bands corresponding to RplI215 are marked with an asterisk (\*).

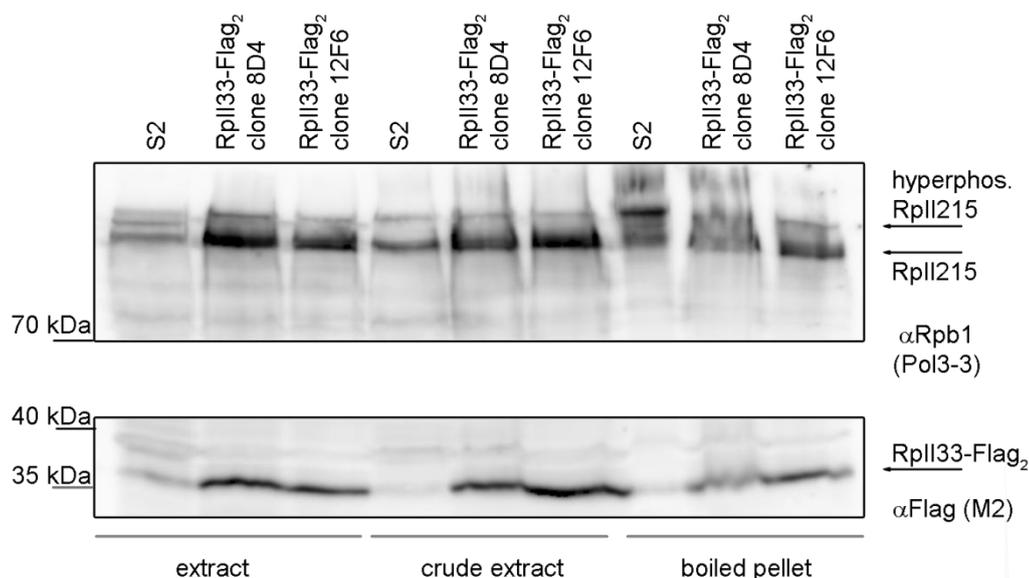
This additional step helped to retrieve RNAPII from the chromatin fraction, however with sometimes low efficiency (see Fig. 13).

Another step that significantly improved extraction of RNAPII was sonication of the samples. Even when using solely the hypotonic extraction buffer, sonication facilitated extraction of RNAPII, especially in its hyperphosphorylated form, which is an indicator of engaged, active RNAPII (see also Fig. 14).



**Fig. 14: Sonication facilitates release of the hyperphosphorylated RplI215 from the nucleus.** Extracts from the cell line RplI33-Flag<sub>2</sub> clone 12 were prepared using the hypotonic extraction buffer (cyto) and additionally subjected to sonication  $\pm$  enzymatic digestion by DNaseI or MNase. Sonication was sufficient to recover the hyperphosphorylated form of RplI215. Since hyperphosphorylation of the CTD is a hallmark of engaged RNAPII, we decided to include a sonication step in our lysis protocol. Additional treatment with either DNase I or MNase did not result in a further increase in the signal for RplI215 or its hyperphosphorylated form.

To combine the strength of the KCl chromatin wash with chromatin shearing, we adapted the composition of the previously used buffers. The composition of the hypotonic buffer was unchanged, but after the snap-freezing step we added an equal volume of hypertonic buffer (10 mM HEPES pH 7.5, 1.5 mM MgCl<sub>2</sub>, 0.6 M KCl, 0.3 % Tergitol-type NP-40, 1 mM DTT, 10 % Glycerol, 1x Proteinase Inhibitor without EDTA (Roche), 1x PhosStop (Roche)) in dropwise manner to the hypotonic extract. To ensure efficient shearing of the chromatin, the pellet was resuspended in a fixed volume to reach a cell density of  $1 \times 10^8$  cells/ml. After 20 min incubation on ice (and vortexing every 5 minutes), samples were sonicated using Diagenode's Bioruptor with the following instrument settings: 12 times 20 sec on, 60 sec off, high. Samples were then centrifuged at top speed, 4 °C for 15 min. Using this protocol, extraction of RNAPII was efficient and the signal from centrifuged extracts was comparable to crude extracts. However, it should be noted that some protein was left in the chromatin pellet (see Fig. 15).

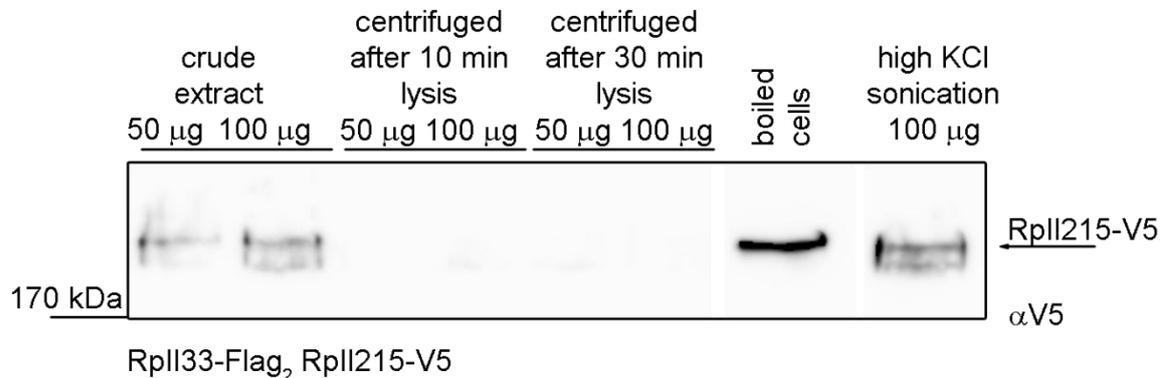


**Fig. 15: RplI215 release from chromatin is reproducibly facilitated by hypertonic lysis with a KCl containing buffer and sonication.** Wild-type S2 cells and the clonal cell lines RplI33-Flag<sub>2</sub> clone 8D4 and clone 12F6 were subjected to lysis as described above. The direct comparison of the centrifuged extracts to crude extracts shows that RNAPII is retrieved in a reproducible and efficient manner. However, some RNAPII is still left in the chromatin pellet (compare boiled pellet fractions). Arrows indicate bands corresponding to un- and hyperphosphorylated RplI215, and RplI33-Flag<sub>2</sub>.

Final adaptations to the lysis protocol were made by Romy Böttcher, who demonstrated that the dropwise addition of a hypertonic buffer is not necessary. Instead, cells can be resuspended directly in a single lysis buffer (10 mM HEPES pH 7.5, 0.4 M KCl, 1.5 mM MgCl<sub>2</sub>, 1 % Tergitol-type NP-40, 1 mM DTT, 10 % glycerol, 1x Proteinase Inhibitor without EDTA (Roche), 1x PhosStop (Roche)). Also, it is not necessary to adjust the lysis volume to a defined cell density for efficient lysis. Finally, the concentration of Tergitol-type NP-40 was increased to 1 %. After resuspension of the pellet in lysis buffer, cells are snap-frozen, incubated on ice for 20 min (occasional shaking), sonicated as described above and centrifuged at 4 °C, top speed, 15 min.

Since mechanical disruption by grinding was described as lysis method in the first report of the NET-Seq technique (Churchman and Weissman, 2011; Churchman and Weissman, 2012), we made a final attempt for an alternative lysis method using a coffee grinder (Gastroback) as described in Materials and Methods (see V.2.2.2.5). We used the optimized buffer composition (as described above), but increased the molarity of the buffer component (HEPES) from 10 mM to 100 mM to ensure that the pH will stay stable at 7.5 even in the presence of dry ice. Briefly, cells were resuspended in lysis buffer and powdered dry ice was added until the sample was completely frozen. The frozen cells were then ground in a pre-cooled coffee grinder and

subsequently incubated at room temperature until thawing was complete. Subsequently, the samples were incubated on ice for the times indicated in Fig. 16.



**Fig. 16: Comparison of lysis by the optimized protocol (high KCl, sonication) and protein extraction by grinding.** We cryogenically ground cells in a coffee grinder (detailed in Materials and Methods (V.2.2.2.5)) and incubated extracts for additional 10 or 30 min on ice prior to centrifugation. While RplI215 bands are readily visible in the crude extract, no bands are detectable in the centrifuged extracts.

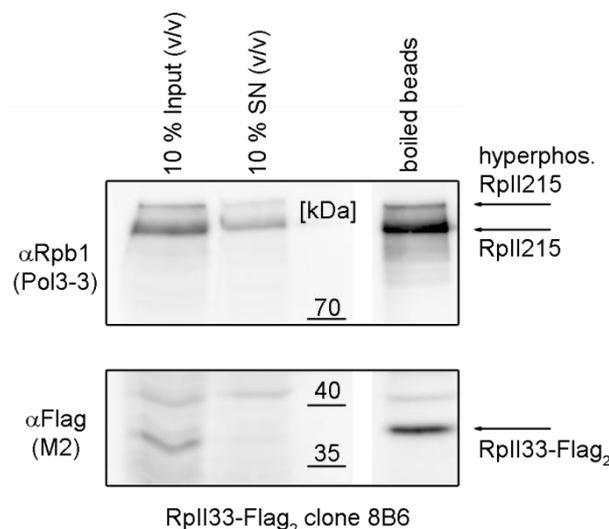
As observed for various approaches of cell lysis (see above), RplI215 protein was detectable only in the crude extract. Centrifugation appears to pellet the RNA polymerase II with the insoluble chromatin fraction. This observation underlines the importance of the sonication step included in the optimized high KCl lysis protocol.

### III.1.1.3 Optimization of RNA polymerase II immunoprecipitation and specific elution using RplI33-Flag<sub>2</sub> tagged cell lines

Once we established a robust lysis protocol to retrieve RNAPII by combining the strengths of Tergitol-type NP-40 as detergent for disruption of the nuclear membrane, the high ionic strength of KCl and sonication as a way to shear the chromatin, we investigated whether we can enrich for RNAPII complexes by immunoprecipitation. Given the high sensitivity of the downstream analysis by deep sequencing, we additionally aimed to perform specific elution of the immunoprecipitated complexes in order to improve the signal-to-noise ratio.

When working with RplI33-Flag<sub>2</sub> clones, we pre-incubated magnetic Protein G beads (Dynabeads) with  $\alpha$ Flag antibody (M2, Sigma) for 1 h to overnight. Pre-incubated beads were then washed and resuspended in the lysis extracts. After 1-2 hours of rolling at 4 °C the supernatant was removed and beads were washed with IP buffer (25 mM HEPES pH 7.5, 150 mM NaCl, 12.5 mM MgCl<sub>2</sub>, 10 % glycerol, 0.3 mM Tergitol-type NP-40, 1 mM DTT, 1x

Proteinase Inhibitor without EDTA (Roche), 1x PhosStop (Roche)). Immunoprecipitation of RpII33-Flag<sub>2</sub> was successful (depletion of the corresponding band) in the supernatant (SN) fraction in Fig. 17. Also, RpII215 and its hyperphosphorylated form was co-purified, thus indicating that immunoprecipitation of RpII33-Flag<sub>2</sub> enables the enrichment of functional RNAPII complexes (Fig. 17).



**Fig. 17: Immunoprecipitation of RpII33-Flag<sub>2</sub> using Protein G beads that were pre-incubated with  $\alpha$ Flag (M2).** 10 % input as well as 10 % of the IP supernatant (SN) was saved and prepared for Western blotting. Bands corresponding to RpII33-Flag<sub>2</sub> and RpII215 are indicated by an arrow. We were able to enrich RpII33-Flag<sub>2</sub> (compare boiled beads lane to input and SN lane). In addition, there is a clear co-enrichment of RpII215 and its hyperphosphorylated form. This indicates that we successfully purified the functional RNA polymerase II complex.

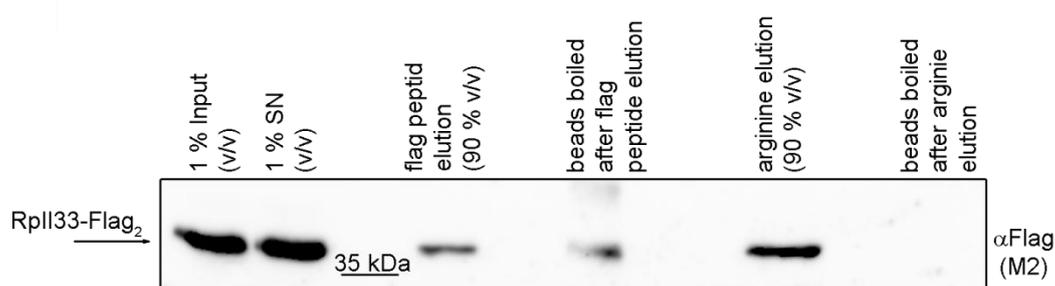
Theoretically, the use of purified Flag-peptide allows for specific elution of the immunoprecipitated protein (complexes). We tested Flag elution under several conditions (indicated in Table 1), but could not find a condition to allow for efficient elution. One possible explanation is that we have introduced a tandem array of Flag-peptides at the C-terminus of RpII33. This tag might show higher affinity for the M2 antibody compared to a 3X Flag peptide-tag, which can be eluted according to manufacturer's instructions using the 3X Flag peptide (Sigma).

**Table 1: Conditions tested for specific elution of RpII33-Flag<sub>2</sub> using the 3X Flag-peptide.** All experiments were performed using RpII33-Flag<sub>2</sub> clone 8B6 cells. The standard elution buffer is comprised of 110 mM KCl, 10 mM HEPES pH 7.5, 1.5 mM MgCl<sub>2</sub>, 0.3 % Tergitol-type NP-40, 1 mM DTT, 1x PhosStop (Roche), 1x Proteinase Inhibitor without EDTA (Roche). Recovery of RpII33-Flag<sub>2</sub> was assessed by Western blotting (comparison of eluate fraction to boiled beads). n.d. = not detectable on a Western blot.

tested conditions		c(3X Flag peptide) in $\mu\text{g/ml}$		
temperature	elution buffer	100	250	500
ice (30 min)	standard	n.d.	very inefficient	n.d.
	+ 10 mM EDTA	very inefficient	n.d./very inefficient (two independent replicates)	n.d.
	+ 10 mM EDTA + 100 mM Glycin	inefficient	inefficient	very inefficient
room temperature (30 min)	+ 10 mM EDTA	n.d.	very inefficient	n.d.
	+ 10 mM EDTA + 100 mM Glycin	very inefficient	inefficient	inefficient

In a final attempt to elute RpII33-Flag<sub>2</sub>, we tested the flag elution protocol after instructions from a Cold Spring Harbor recipe<sup>1</sup>. Beads were resuspended in elution buffer (10 mM HEPES pH 7.5, 0.3 M KCl, 1.5 mM MgCl<sub>2</sub>, 0.05 % Tergitol-type NP-40, 300  $\mu\text{g/ml}$  3X Flag peptide) and incubated for 30 min at room temperature.

We compared this to an elution using a 0.75 M arginine solution (pH 3.5), where beads were incubated for 5 min at 37 °C and then neutralized with 0.14 x Volume of Tris-HCl (pH 8,0) buffer (Futatsumori-Sugai et al., 2009). The latter elution method was very efficient, since no Western blot signal for the boiled beads was detectable. This is in stark contrast to elution using the 3X flag peptide, where the Western blot signal of the elution fraction is comparable to the signal intensity for boiled beads.



**Fig. 18: Comparison of elution methods to elute immunopurified RpII33-Flag<sub>2</sub>.** RpII33-Flag<sub>2</sub> cells were lysed as described above and the lysate added to Dynabeads Protein G preincubated with  $\alpha\text{Flag}$  (M2) for immunoprecipitation. 1 % of input and 1 % supernatant (SN) was saved for Western blotting. Subsequent to IP, RpII33-Flag<sub>2</sub> was eluted from the beads using either 3X Flag peptide or arginine. Beads were boiled to assess elution efficiency.

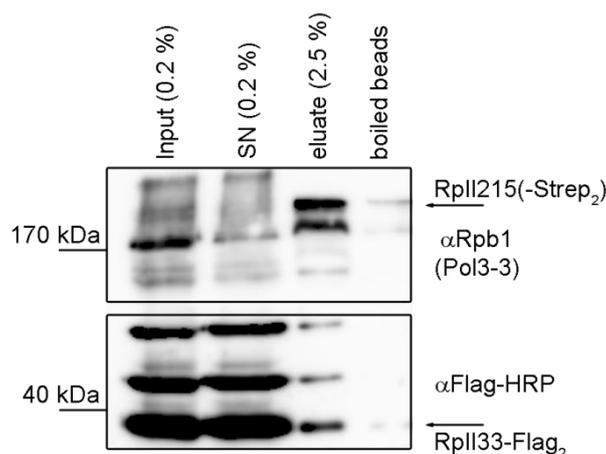
<sup>1</sup> <http://cshprotocols.cshlp.org/content/2010/7/pdb.rec12246.full>

Although the arginine elution was very efficient in eluting RpII33-Flag<sub>2</sub>, it is unclear how the co-purifying RNA might be affected by the decrease in pH. In addition, the elution step might result in release of nucleic acids that unspecifically bind to the beads.

Indeed, NET-Seq libraries prepared by IP against RpII33-Flag<sub>2</sub> and eluted by arginine showed a substantial degree of reads that neither mapped to *E. coli*, nor to *D. melanogaster*. Further investigation revealed that those reads originate from bacterial contamination (data not shown). This contamination problem was significantly reduced when we directly treated beads with Proteinase K to release the co-immunoprecipitated RNA (data not shown).

#### III.1.1.4 Optimization of RNA polymerase II immunoprecipitation and specific elution using RpII33-Flag<sub>2</sub> RpII215-Strep<sub>2</sub> tagged cell lines

As described in III.1.1.2 we decided to tag also the biggest subunit of the RNAPII complex, RpII215, at its C-terminus. We hypothesized that pulling directly on RpII215 might improve recovery of the associated transcripts. To allow for gentle and specific elution, we chose to modify RpII215 with a Strep-tag. This peptide sequence shows high affinity to streptavidin and can be eluted by competition with biotin. We used magnetic Strep-Tactin (an engineered derivative of streptavidin) beads (IBA) with the previously optimized lysis and immunoprecipitation protocol. As Fig. 19 shows, this approach was successful to enrich for RpII215-Strep<sub>2</sub> and the smaller subunit RpII33-Flag<sub>2</sub>. In addition, the elution by biotin was highly efficient.

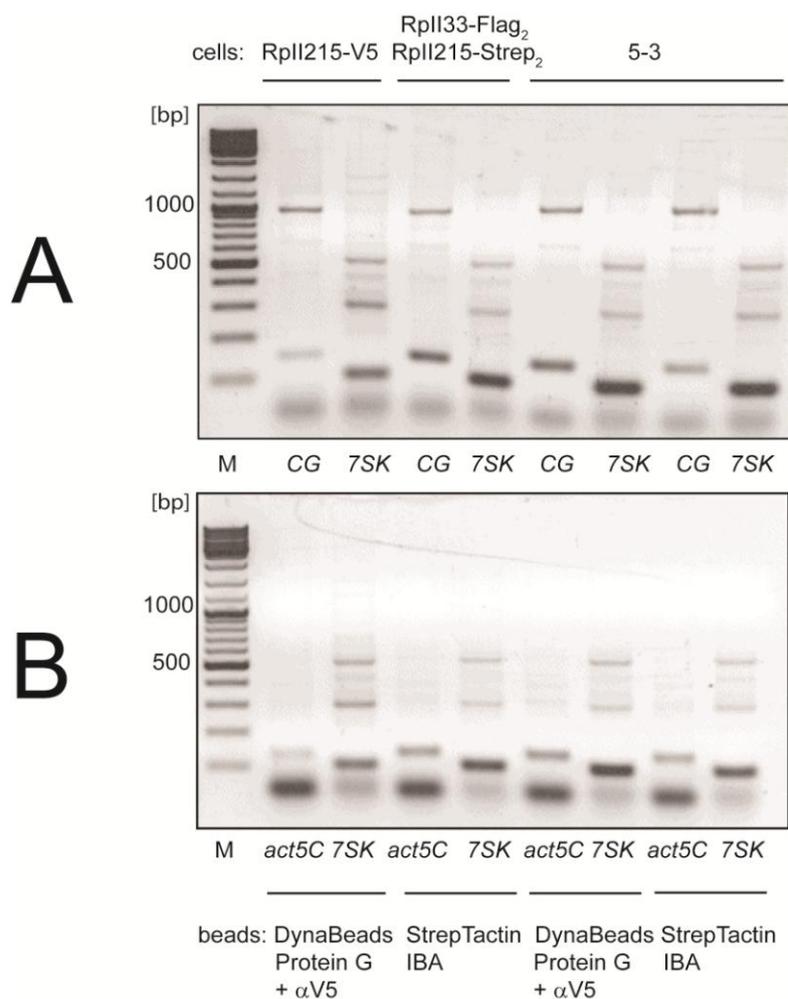


**Fig. 19: Immunoprecipitation and specific elution of RpII215-Strep<sub>2</sub> using Strep-Tactin Type2HC beads and 1x Biotin-elution buffer.** RpII33-Flag<sub>2</sub> RpII215-Strep<sub>2</sub> cells were lysed under optimized conditions (see III.1.1.2) and the lysate was used as input for the IP. ~ 0.2 % input as well as 0.2 % supernatant (SN) were saved and prepared for Western blotting. Elution with 1x Biotin buffer retrieved RpII215-Strep<sub>2</sub> (indicated by an arrow). There was clearly a co-purification of RpII33-Flag<sub>2</sub> when we probed the membrane with  $\alpha$ Flag-HRP (indicated by an arrow). This validates the approach as method to specifically elute the functional RNA polymerase II complex

### III.1.1.5 Evaluation of co-purified RNA from RpII215-Strep<sub>2</sub> IP by deep sequencing and NET-RT-PCR

Based on the so far optimized lysis, immunoprecipitation and elution conditions, we next wanted to evaluate the signal-to-noise-ratio of the co-purified RNA from the NET-Seq experiment. We performed lysis, IP and biotin elution for RpII33-Flag<sub>2</sub> RpII215-Strep<sub>2</sub> and for the untagged 5-3 cell line (control). Co-purified RNA was incorporated into a library and the two libraries were pooled (IS02). However, we detected a high degree of background in the deep sequencing data, especially for the 30 nt long 2S rRNA and RNAPIII transcribed loci (e.g. tRNAs), which results in a poor signal-to-noise ratio.

We thus investigated whether unspecific binding of RNAs to the beads might be the problem. To this end, we compared immunoprecipitation using cell extracts from RpII33-Flag<sub>2</sub> RpII215-Strep<sub>2</sub> cells and magnetic strep-tactin beads (IBA) with immunoprecipitation of using cell extract from RpII215-V5 clone 16 (which was provided by Volker Nitschko) and Protein G (Dynabeads) that were preincubated with  $\alpha$ V5. To evaluate the signal-to-noise ratio, we isolated co-immunoprecipitated RNA, reverse transcribed it and performed a PCR for RNAPII associated loci (*act5C*, *CG15098*) as well as for a RNAPIII transcribed locus (*7SK*). Although RT-PCR is no quantitative method, there is a correlation between template input and amplicon abundance. For the Strep-Tactin Beads we know from the deep sequencing data (IS02) that background is rather high (see above). This is also reflected in the RT-PCR gel, where the amplicon signal for *7SK* is even stronger than for *act5C* or for *CG15098*. A reduced background would be indicated by a stronger amplicon signal for *act5C* or *CG15098* compared to *7SK*. However, the NET-RT-PCR using the Dynabeads shows the same poor signal-to-noise ratio (see Fig. 20). An alternative explanation for the strong co-enrichment of *7SK* is given in reports that identify the *7SK* snRNA as inhibitor of the positive transcription elongation factor p-TEFb. p-TEFb is a cyclin-dependent kinase that is required for progression of promoter-proximal paused RNAPII into elongation (Nguyen et al., 2001; Peterlin et al., 2012; Quaresma et al., 2016; Yang et al., 2001). Furthermore, there is a study that indicates association of human RNAPII to RNAPIII transcribed loci, including the *7SK* locus (Raha et al., 2010).



**Fig. 20: Specificity of immunoprecipitation tested by NET-RT-PCR.** Lysis and immunoprecipitation were performed as described above using the noted cell lines and the indicated beads.

The eluate (for Strep-Tactin beads) or the washed beads were digested with Proteinase K to release co-immunoprecipitated nucleic acids. RNA was further purified by TRIZOL treatment and DirectZol purification columns, then subjected to reverse transcription using random hexamer primers. Specific PCRs were performed using either primers for *CG15098*, *actin5C* or *7SK*. **(A)** The signal-to-noise ratio of co-purified RNA from the IP against tagged RpII215 can be assessed by comparison of signal intensities from RT-PCR products for *CG15098* (RNAPII associated) and *7SK* (RNAPIII associated). **(B)** Same as in **(A)**, but only for *act5C* (RNAPII associated) in comparison to *7SK*.

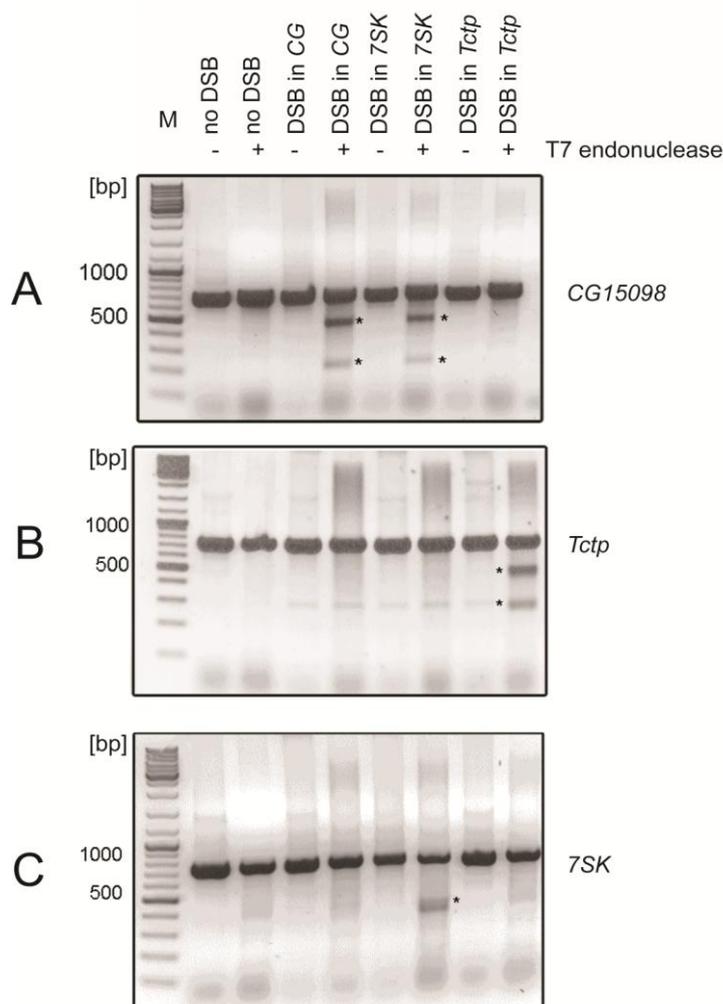
### III.1.2 Preliminary results on transcription at induced DNA double-strand breaks

Using the RpII33-Flag<sub>2</sub> RpII215-Strep<sub>2</sub> cell line and the optimized lysis protocol, we wanted to gain insight into the RNAPII recruitment to DSB sites. To this end, we transfected RpII33-Flag<sub>2</sub> RpII215-Strep<sub>2</sub> cells that also stably express Cas9 with sgRNA expression templates targeting different genomic loci and compared the results to an untreated control. *CG15098* is strongly expressed in *Drosophila* S2 cells and contains three introns. Data from Karin Merk (Merk et al.,

2017) showed that introduction of a DSB by CRISPR/*cas9* leads to a robust siRNA response upstream of the break. In contrast to this, a DSB in the intronless and also highly expressed *Tctp* gene does not trigger an efficient siRNA response (Merk et al., 2017). To investigate whether we can detect antisense transcription at this locus, we also transfected an sgRNA expression template that targets *Tctp*.

Another very intriguing question is whether antisense transcription at DSB is performed by the respective RNA polymerase (e.g. RNAPII for mRNA genes). To this end, we induced a break in the *7SK* gene that is transcribed by RNA polymerase III.

48 h after transfection, cells were harvested and cell lysis and IP were performed using the optimized protocol (see III.1.1.2 and III.1.1.4). To investigate whether 48 h were sufficient for introduction of DSB and DSB repair to occur, we isolated gDNA from transfected cells and performed a T7 endonuclease assay. T7 endonuclease recognizes and cleaves non-perfectly matched DNA and can be used to detect mutagenic repair. As shown in Fig. 21 error-prone repair had occurred 48 h after transfection of the different sgRNAs. Surprisingly, introduction of a DSB in the *7SK* gene by sgRNA transfection also resulted in a T7 endonuclease assay signal when we performed a PCR and T7 endonuclease digestion using *CG15098* specific primers (Fig. 21). This is also reflected in the deep sequencing data obtained for the DSB in *7SK* sample (compare Fig. 23 and Table 5). Although greatest care was taken to avoid cross-contamination with different sgRNA expression templates, this might be the reason for the unexpected observation.



**Fig. 21: T7 endonuclease assay to detect error-prone DSB repair for the respective loci.** DSBs were introduced into the Cas9 expressing S2 cell line 5-3 by sgRNA expression template (indicated above) transfection, untransfected cells serve as control (no DSB). 48 h post transfection, cells were harvested and gDNA was extracted. A specific PCR for the three different loci was conducted and subjected to T7 endonuclease digestion. Fragments of the expected size are indicated by an asterisk (\*). **(A)** A specific PCR for *CG15098*. PCR amplicons were digested using the T7 endonuclease. The expected fragment lengths are 420 nt and 207 nt. **(B)** PCR amplicons for the *Tctp* gene were generated and digested to yield fragments of 504 nt and 273 nt. **(C)** Specific PCR products for the *7SK* locus were digested and resulted in fragments of 354 nt and 421 nt.

We performed cell lysis and IP under the optimized conditions and eluted with biotin as described above (see III.1.1.2 and III.1.1.4). The co-immunoprecipitated RNA was retrieved, purified and size-selected on a 20 % urea-PAGE, where we excised bands in the range ~22–50 nt. Purified RNA was subjected to 3' linker ligation. Since a free 3' OH is required for 3' ligation, information of the read orientation is conserved during library preparation. 3' ligated RNA was again subjected to gel purification and size-selection (~30-100 nt) on a 15 % urea-PAGE. Since RNA might have been subject to partial degradation at the 5'-end during the immunoprecipitation step, we treated RNA with T4 Polynucleotidkinase. This step ensures that

co-purified RNA species carry a 5' phosphate that is essential for 5' ligation. After 5' ligation, RNA is reverse transcribed and purified. Co-purified RNA from all four samples (no DSB, DSB in *CG15098*, DSB in *Tctp*, DSB in *7SK*) was incorporated, as described above, into a library. All four libraries were pooled (to yield IS03) to allow for direct comparison.

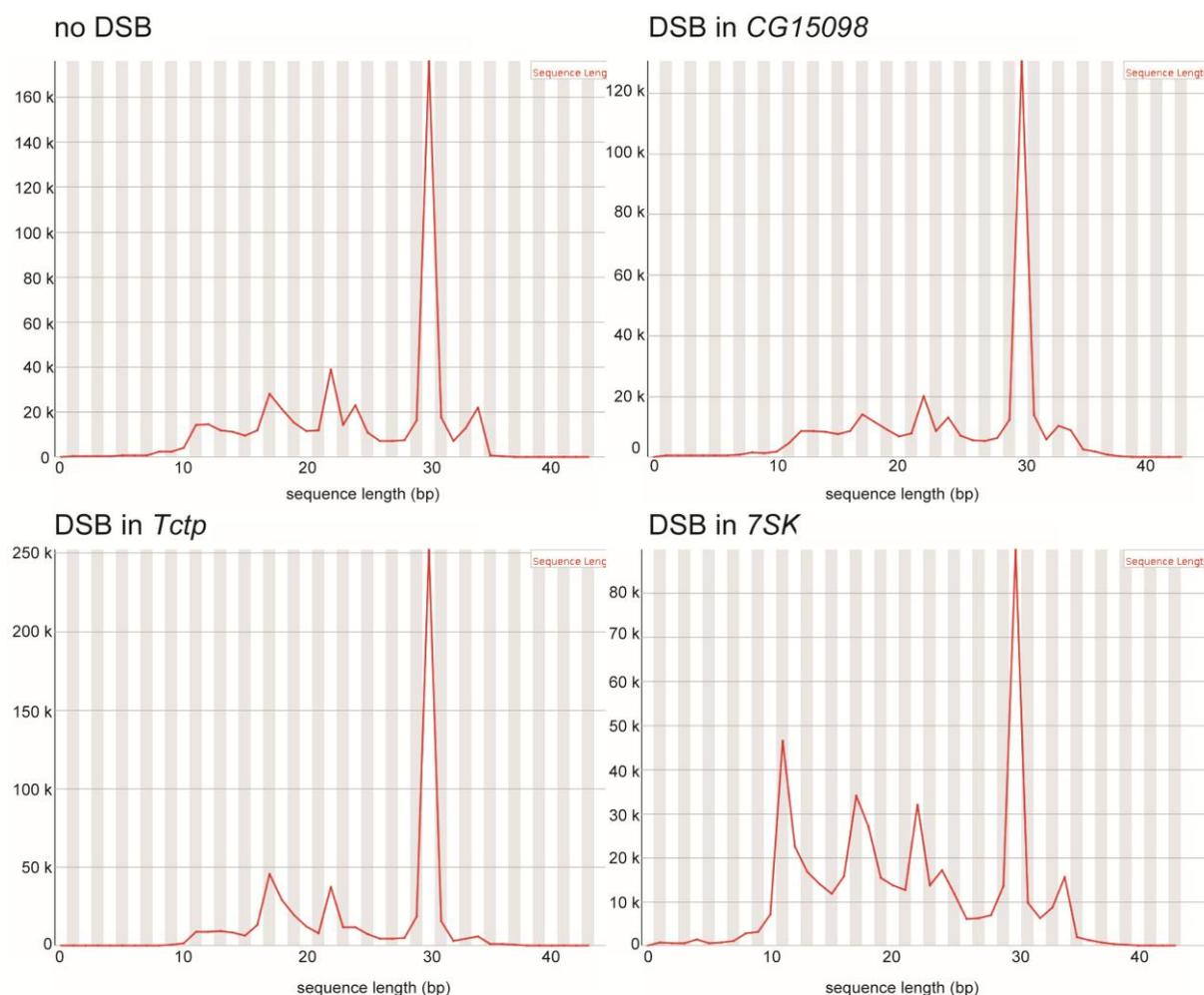
### III.1.2.1 Quality of NET-Seq libraries

3' linker is used in excess in the 3' linker reaction, in order to ensure quantitative ligation of small RNAs to the linker. Free 3' linker is visible as a distinct 17 nt band on the urea-PAGE after 3' ligation and can thus be removed from the sample by size-selection. Free 3' linker can directly react with the 5' linker without incorporation of insert RNA, giving rise to “linker-to-linker products”. A high percentage of “linker-to-linker” can thus result in reduced sequencing depth, as these reads contain none of the intended sequences. IS03 showed a low percentage of “linker-to-linker” (Table 2). The bioinformatic Galaxy Tool ‘Clip Adaptor Sequence’ removes the 3' adaptor and the sequence downstream of the 3' adaptor. “Clipped reads” therefore represent reads for which incorporation of RNA inserts was successful. IS03 shows a high percentage of clipped reads (Table 2). We thus conclude that the technical quality of IS03 was rather high.

**Table 2: Assessment of the technical quality of IS03.** Reads that neither belong to the “linker-to-linker” nor the “clipped reads” category, are categorized as “others”. The low percentage of linker-to-linker and others indicates high technical quality.

all lengths	no DSB	DSB in <i>CG15098</i>	DSB in <i>7SK</i>	DSB in <i>Tctp</i>
<b>total reads</b>	53,192,020	37,122,068	65,267,134	51,172,140
<b>% of summarized total reads (all four libraries)</b>	26 %	18 %	32 %	25 %
<b>linker-to-linker</b>	1,713,849	247,121,800	1,703,839	160,273
<b>%</b>	3.2%	6.7%	2.6%	0.3%
<b>clipped reads</b>	49,236,062	32,756,830	48,197,250	48,761,455
<b>%</b>	92.6%	88.2%	73.8%	95.3%
<b>others</b>	2,242,109	1,894,020	15,366,045	2,250,412
<b>%</b>	4.2%	5.1%	23.5%	4.4%

The length distribution profile of clipped reads showed a similar overall pattern (with exception of *7SK*) with a more or less pronounced shoulder between 10-16 nt, small peaks at 17 nt and 22 nt and a very prominent peak at 30 nt. This corresponds to the 2S rRNA, which is a major contaminant in all samples (compare Fig. 22 and Table 3).



**Fig. 22: Distribution of sequence lengths over all sequences.** Length distribution of clipped reads was determined using the Galaxy Tool FastQC on the Lafuga Galaxy Server.

**Table 3: Contribution of 2S rRNA reads to all clipped reads.**

	sample			
Size selection (in nt)	No DSB	DSB in <i>CG15098</i>	DSB in <i>Tctp</i>	DSB in <i>7SK</i>
$\geq 17$	53.47 %	47.40 %	74.27 %	50.04 %
$\geq 20$	51.73 %	47.71 %	45.07 %	71.98 %
$\geq 22$	52.63 %	49.36 %	72.88 %	45.36 %
$= 30$	90.61 %	89.93 %	94.17 %	86.89 %

Further analysis was performed for reads with lengths  $\geq 17$  nt, because there are several reports that the length of the nascent transcript that is protected by the RNA polymerase complex is on average 20 nt long (17-22 nt) (Choder and Aloni, 1988) and can extend up to  $\sim 30$  nt for arrested RNAPII complexes (Brabant and Acheson, 1995; Gu et al., 1996). Since our protocol did not include a nuclease step (to degrade RNA that is not protected by the active center of RNAPII), we did not exclude reads longer than 30 nt from analysis. 2S rRNA sequences were

bioinformatically removed prior to further analysis. Table 4 shows the percentage of clipped reads mapped to various precomputed *D. melanogaster* sequence collections. It should be noted that introns and exons also include rDNA and tDNA sequences and are thus not the best indicator whether enrichment of RNAPII nascent transcripts was successful. On the other hand, mRNA transcripts represent RNAPII transcribed genes. Also “miscRNAs” comprise RNAPII associated loci, since besides ribosomal RNA (of which the 2S rRNA sequence was filtered out prior to mapping) it also includes small nucleolar RNA (snoRNAs) and small nuclear RNA (snRNAs). Both, snoRNAs and snRNAs, are either transcribed by RNAPII or RNAPIII (Henry et al., 1998).

However, the high content of tRNA mapping reads in all libraries (16.24 % (for DSB in *Tctp*) – 36.77 % (no DSB)) reveals another substantial contaminant, especially since there are reads mapping to each of the 313 tRNAs in the *Drosophila* genome. On the other hand, RNAPII occupancy at RNAPIII transcribed genes was reported for mammalian cells (Raha et al., 2010) and this phenomenon might contribute to the high abundance of tRNA mapping reads for RNAPII NET-Seq from *Drosophila* S2 cells.

**Table 4: Percentage of clipped reads of a read length  $\geq 17$  nt after *in silico* removal of 2S rRNA mapping to the indicated sequence collections from Flybase.** Clipped reads were size selected *in silico* using the Galaxy tool Filter-FastQ on the Lafuga Server. Reads were then mapped to various sequence collections using Bowtie with no mismatch allowed.

Size $\geq 17$ nt (after <i>in silico</i> removal of 2S rRNA)	sample			
	No DSB	DSB in <i>CG15098</i>	DSB in <i>Tctp</i>	DSB in <i>7SK</i>
Mapping to				
<i>E. coli</i>	6.27 %	6.65 %	4.25 %	8.00 %
<i>D. mel.</i> genome	60.66 %	56.05 %	44.87 %	51.38 %
<i>D. mel.</i> introns	38.44 %	23.16 %	17.54 %	29.10 %
<i>D. mel.</i> exons	59.22 %	54.69 %	44.17 %	50.40 %
<i>D. mel.</i> mRNA transcripts	6.78 %	6.46 %	3.98 %	6.51 %
<i>D. mel.</i> miscRNAs	16.74 %	27.60 %	24.33 %	16.85 %
<i>D. mel.</i> tRNAs	36.77 %	21.24 %	16.24 %	27.52 %

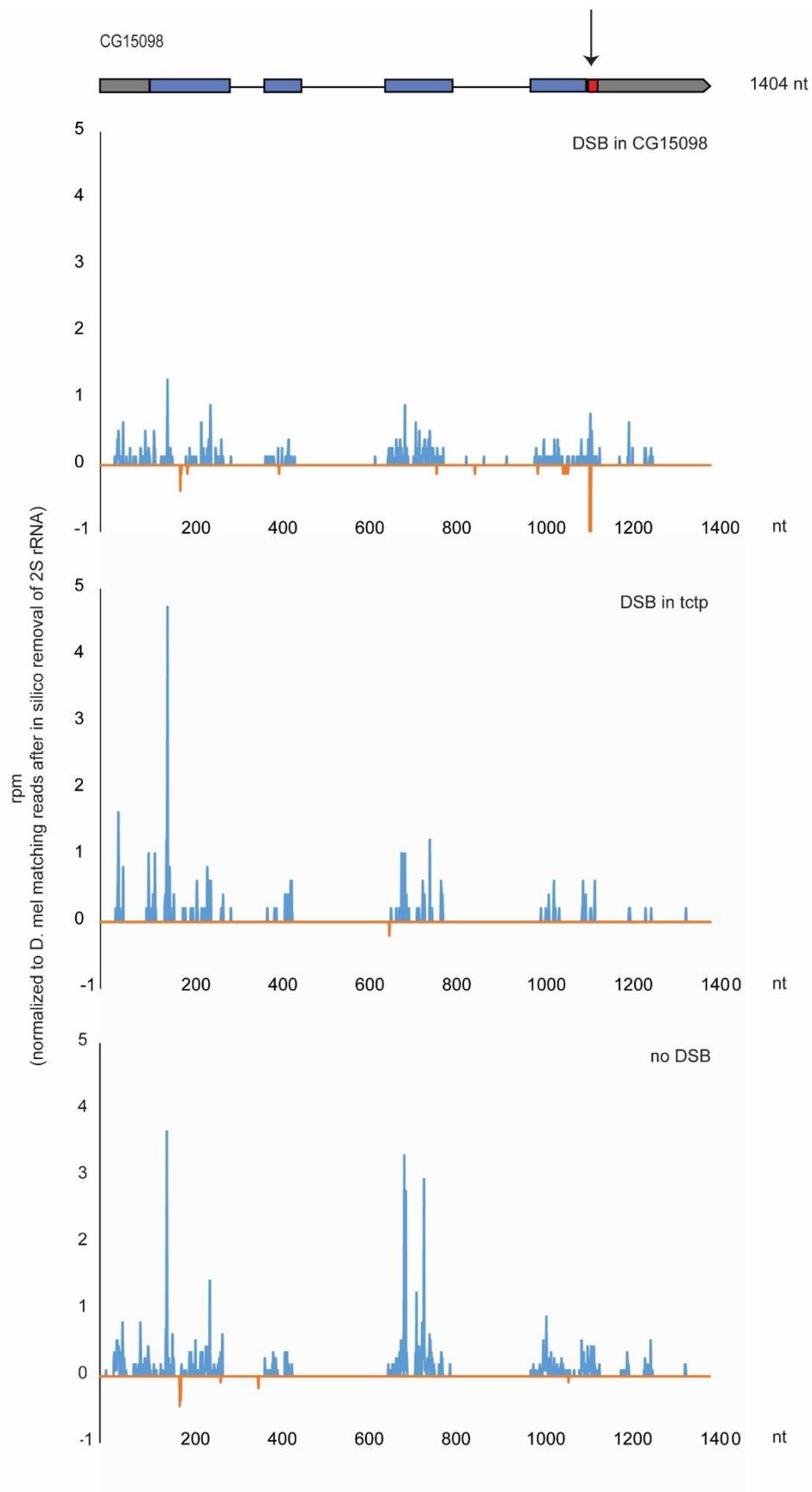
The mapping data reveals two major contaminants, namely the 2S rRNA and tRNAs. Especially in comparison to a publication of NET-Seq in yeast where the authors reported that 55 % of the reads align to rRNA and 2 % to tRNA (Churchman and Weissman, 2012), it becomes evident that further optimization is needed to specifically enrich for RNAPII transcripts and improve the signal-to-noise ratio.

---

### III.1.2.2 Transcriptional response to a DSB

Data has shown that a DSB at an endogenous locus leads to an siRNA response upstream of the break (Merk et al., 2017). Depending on the locus and the targeted position within the gene, the strength of the siRNA response varied. For introduction of a break in *CG15098*, we thus selected the sgRNA expression template for transfection (CRISPR-Oligo #750) that had elicited the strongest response (Merk et al., 2017) and targets the 3' UTR of *CG15098*.

Compared to the previously investigated siRNA response, the number of reads mapping to *CG15098* is very little. However, one has to keep in mind the low signal-to-noise ratio (due to 2S rRNA and tRNA contamination) and the fact that NET-Seq reveals RNAPII occupancy on a genome-wide scale, while small RNA library preparation allows for specific enrichment of siRNAs species. Compared to the sense reads, which originate mainly from UTRs and exons, antisense reads are a very rare event, but are detectable even in the absence of a DSB (albeit to a lower extent) (see Table 5).



---

**Fig. 23: Sense and antisense reads mapping to *CG15098* for three different conditions.** Clipped reads were bioinformatically size selected to have a length  $\geq 17$  nt. Bowtie was used for mapping of the reads to *CG15098* (release r6.14). The customized perl script 'map\_in\_interval' was used to generate the read distribution over the locus with a bin size of 1. Reads were normalized to *D. mel.* matching reads after bioinformatic removal of the 2S rRNA to obtain reads per million (rpm). Sense reads are depicted in blue, antisense reads in orange. Please note that the scale of the minus axis was manually set to  $-1$  rpm in the upper panel (DSB in *CG15098*) to allow for direct comparison of the different conditions. We observe a high number of reads that map in antisense to the CRISPR oligo (23.22 rpm which amounts to 93.78 % of all antisense matching reads for *CG15098*).

**Table 5: Number of clipped reads of the indicated read length mapping to *CG15098*.** Size selection was performed *in silico* using the Galaxy tool Filter-FastQ. Subsequent to mapping clipped reads to *CG15098* (release r6.14) using Bowtie (no mismatch allowed), we ran the customized perl script 'count\_hits\_to\_fasta' to retrieve the number of reads mapping to the sense and antisense strand of *CG15098*. We performed mapping after size-selection for different sizes of co-purified RNA. Indicated are the numbers of sense, antisense, and all gene matching reads for both absolute and normalized values (reads per million (rpm)). To obtain reads per million, we performed normalization to *D. mel.* matching reads and removal of 2S rRNA reads. The fraction % antisense/gene matching is also indicated and elevated when we induced a DSB in *CG15098*. 17 nt is the minimal length that is protected by the active center of the RNAPII complex. Note that we observed a strong signal from CRISPR matching reads for DSB in *CG15098* in the size range  $\geq 17$  nt. This is in accordance with the CRISPR oligo targeting length of 19 nt. 181 antisense reads (of a total 193) map to the CRISPR target site, which amounts to 33.52 % of gene matching reads. We therefore removed antisense reads mapping to the region covered by the sgRNA from the analysis (marked with an asterix (\*)). To exclude possible degradation products from analysis as well as technical artifacts from CRISPR matching reads, we also size-selected *in silico* for RNAs  $\geq 20$  nt. To exclude contaminating siRNAs we also size-selected for RNAs  $\geq 22$  nt.

Reads mapping to <i>CG15098</i>	Absolute number			Normalized (rpm)			antisense (% of sense+antisense)
	sense	antisense	all gene matching	sense	antisense	all gene matching	
$\geq 17$ nt							
No DSB	868	14	882	77.76	1.25	79.02	1.6
DSB in <i>CG15098</i>	347	12*	359*	44.52	1.54*	46.06*	2.2*
DSB in <i>Tcp</i>	220	1	221	45.52	0.21	45.43	0.5
DSB in <i>7SK</i>	475	13	488	57.53	1.57	59.11	2.7
$\geq 20$ nt							
No DSB	592	4	596	59.46	0.40	59.87	0.7
DSB in <i>CG15098</i>	236	6	242	33.93	0.86	34.80	2.5
DSB in <i>Tcp</i>	163	1	164	37.36	0.23	37.59	0.6
DSB in <i>7SK</i>	314	6	320	44.03	0.84	44.87	1.9
$\geq 22$ nt							
No DSB	452	4	456	48.87	0.43	49.31	0.9
DSB in <i>CG15098</i>	180	3	183	28.10	0.47	28.57	1.6
DSB in <i>Tcp</i>	125	1	126	30.44	0.24	30.69	0.8
DSB in <i>7SK</i>	22	0	22	42.85	0	42.85	0

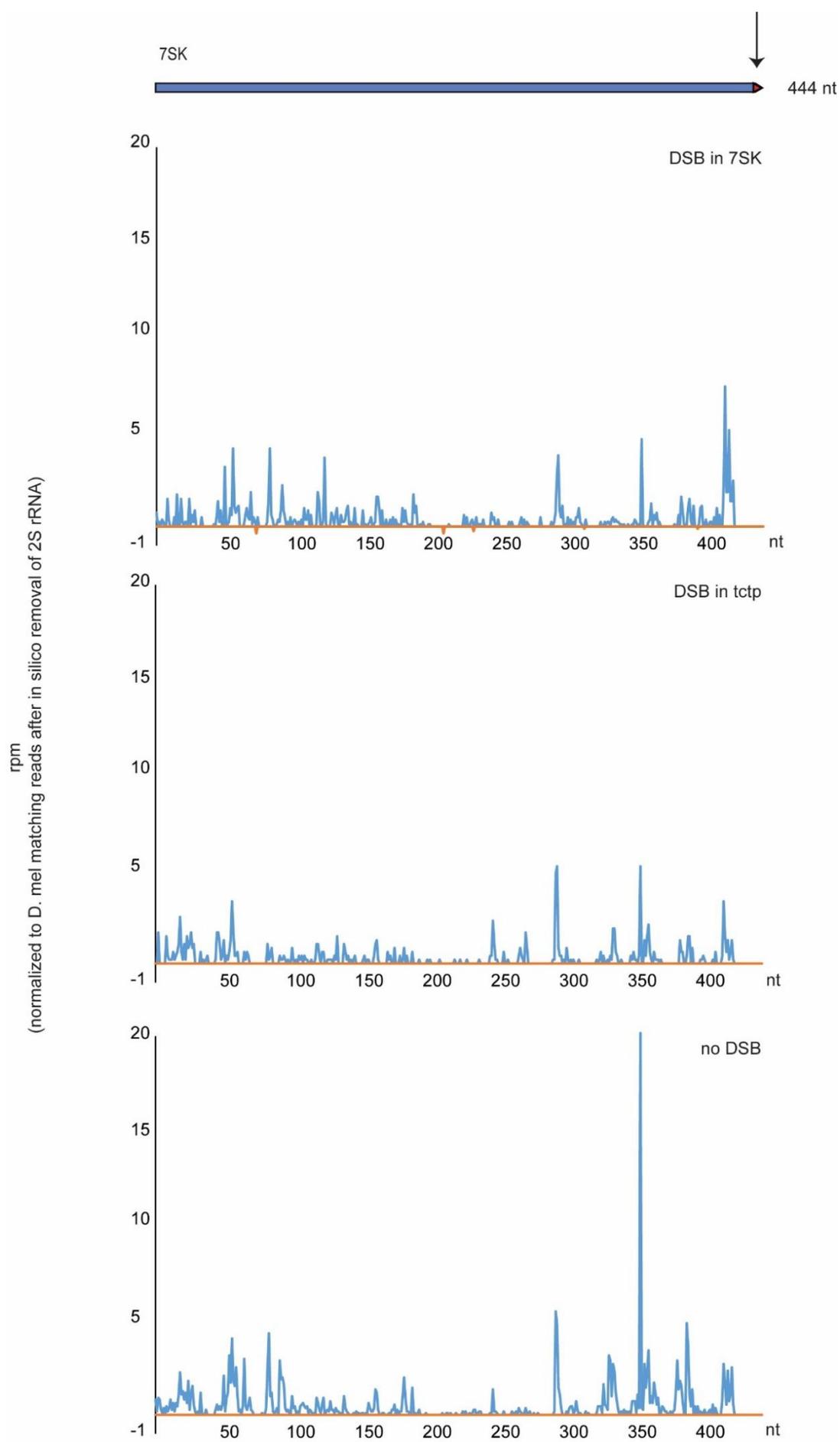
For the intronless and highly expressed *Tctp* gene no antisense transcripts could be detected. This was the case for all four libraries. Even if we introduced a DSB in the 3' UTR (CRISPR-Oligo #1173) this did not trigger antisense transcription. This is in accordance with findings by Karin Merk (Merk et al., 2017), who did not observe an efficient siRNA response to a break in *Tctp*.

**Table 6: Number of clipped reads of the indicated read length that map to *Tctp*.** Size selection was performed *in silico* using the Galaxy tool Filter-FastQ. After mapping of clipped reads to *Tctp* (release r6.15) using Bowtie (no mismatch allowed), we run the customized perl script 'count\_hits\_to\_fasta' to retrieve the number of reads mapping to the sense and antisense strand of *Tctp*. As before, we size-selected for different RNA lengths,  $\geq 17$  nt taken into consideration the minimal length of nascent transcript protected by RNAPII,  $\geq 20$  nt to enrich for *bona fide* transcripts (and exclude degradation products),  $\geq 22$  nt to exclude siRNAs. We present the absolute number of sense, antisense and gene matching reads that map to *Tctp*, as well as the normalized reads per million (rpm) to normalize for all *D. mel.* matching reads after *in silico* removal of 2S rRNA. Finally, the percentage of antisense reads to all gene matching is given.

Reads mapping to <i>Tctp</i>	Absolute number			Normalized (rpm)			antisense (% of sense+antisense)
	sense	antisense	all gene matching	sense	antisense	all gene matching	
$\geq 17$ nt							
No DSB	868	0	868	77.76	0	77.76	0
DSB in <i>CG15098</i>	668	1	669	85.70	0.13	85.83	0.2
DSB in <i>Tctp</i>	268	0	268	55.09	0	55.09	0
DSB in <i>7SK</i>	672	0	672	81.40	0	81.40	0
$\geq 20$ nt							
No DSB	641	0	641	64.39	0	64.39	0
DSB in <i>CG15098</i>	517	1	518	74.34	0.14	74.48	0.2
DSB in <i>Tctp</i>	210	0	210	48.13	0	48.13	0
DSB in <i>7SK</i>	518	0	518	72.64	0	72.64	0
$\geq 22$ nt							
No DSB	529	0	529	57.20	0	57.20	0
DSB in <i>CG15098</i>	446	1	447	69.64	0.16	69.79	0.2
DSB in <i>Tctp</i>	175	0	175	42.62	0	42.62	0
DSB in <i>7SK</i>	38	0	38	74.02	0	74.02	0

The small non-coding RNA *7SK* is transcribed by RNAPIII. We introduced a DSB at this locus to investigate whether RNAPII is recruited to DSBs in general or whether the respective, transcribing RNA polymerase performs the antisense transcription.

For the *7SK* locus, we observed specific enrichment of antisense reads in response to the introduced break. No antisense reads were obtained in the other libraries. However, we did not retrieve antisense reads that were longer than 21 nt, while we could detect one 30 nt long antisense transcript after introducing a DSB in *CG15098* mapping to the *CG15098* gene. This argues against RNAPII recruitment to the site of the break but rather for a contamination with siRNAs that are generated due to the double-strand break. In addition, we found that reads mapping to the *7SK* locus were highly abundant in all libraries. This might be an indication of the above mentioned poor signal-to-noise ratio. Alternatively, co-enrichment of the *7SK* snRNA might be indicative of its function as transcriptional regulator due to its interaction with p-TEFb (Nguyen et al., 2001; Peterlin et al., 2012; Quaresma et al., 2016; Yang et al., 2001).



**Fig. 24: Sense and antisense reads mapping to *7SK* for three different conditions (cut in *7SK*, cut in *Tctp* and uncut control).** Clipped reads were size-selected for lengths  $\geq 17$  nt using the tool Filter-FastQ on the Lafuga Galaxy Server. Bowtie was used for mapping the reads to *7SK* (release r6.15). To obtain the depicted read distribution over the *7SK* locus the customized perl script ‘map\_in\_interval’ was used with a bin size of 1. Reads were then normalized to *D. mel.* matching reads (after *in silico* removal of 2S rRNA) to obtain reads per million (rpm). Sense reads are shown in blue, antisense reads in orange.

**Table 7: Number of clipped reads of the indicated read length (size selection was performed *in silico* using the Galaxy tool Filter-FastQ) that map to *7SK*.** Subsequent to mapping clipped reads to *7SK* (release r6.15) using Bowtie (no mismatch allowed), we run the customized perl script ‘count\_hits\_to\_fasta’ to retrieve the number of reads mapping to the sense and antisense strand of *7SK*. Again, we size-selected for different RNA lengths,  $\geq 17$  nt to take into consideration the minimal length of nascent transcript protected by RNAPII,  $\geq 20$  nt to enrich for *bona fide* transcripts (and exclude degradation products),  $\geq 22$  nt to exclude siRNAs. Shown are the absolute number of sense, antisense, and gene matching reads that map to *Tctp*, as well as the normalized reads per million (rpm) to normalize for all *D. mel.* matching reads after *in silico* removal of 2S rRNA. Finally, the percentage of antisense reads to all gene matching is given.

Reads mapping to <i>7SK</i>	Absolute number			Normalized (rpm)			antisense (% of sense+antisense)
	sense	antisense	all gene matching	sense	antisense	all gene matching	
$\geq 17$ nt							
No DSB	2450	0	2450	219.50	0	219.50	0
DSB in <i>CG15098</i>	2021	0	2021	259.30	0	259.30	0
DSB in <i>Tctp</i>	618	0	618	127.04	0	127.04	0
DSB in <i>7SK</i>	1255	10	1265	152.01	1.21	153.22	0.8
$\geq 20$ nt							
No DSB	1807	0	1807	181.51	0	181.51	0
DSB in <i>CG15098</i>	1647	0	1647	236.82	0	236.82	0
DSB in <i>Tctp</i>	470	0	470	107.72	0	107.72	0
DSB in <i>7SK</i>	859	9	868	120.45	1.26	121.71	1
$\geq 22$ nt							
No DSB	1407	0	1407	152.14	0	152.14	0
DSB in <i>CG15098</i>	1360	0	0	212.35	0	212.35	0
DSB in <i>Tctp</i>	385	0	0	93.76	0	93.76	0
DSB in <i>7SK</i>	41	0	41	79.86	0	79.86	0

## III.2 Biological role of DSB-triggered small RNAs

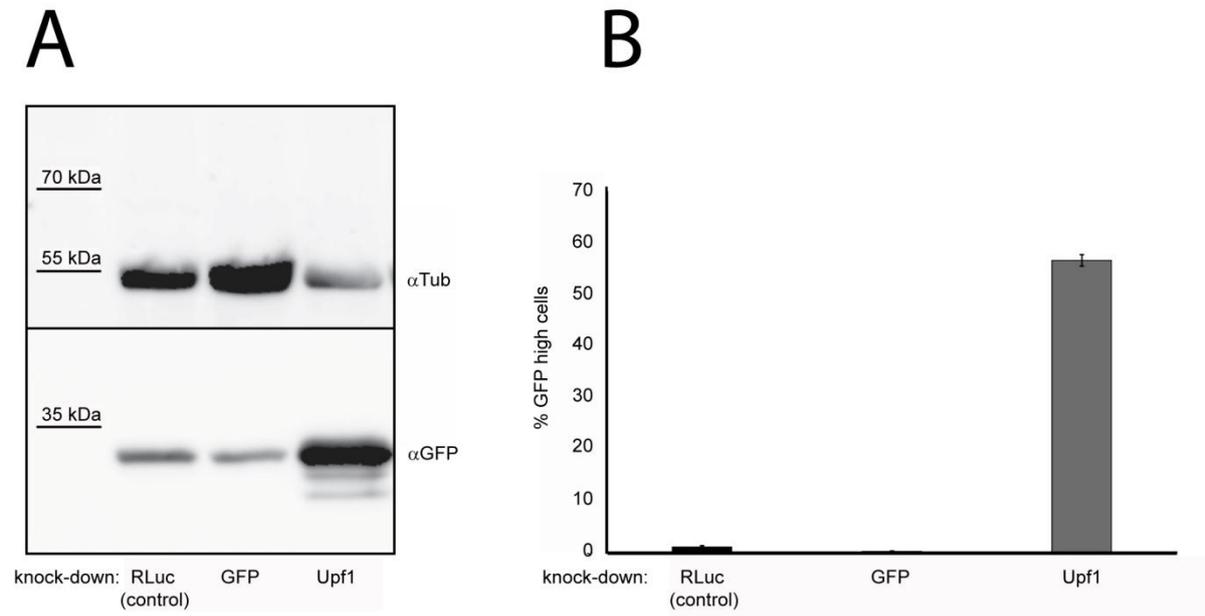
Parts of this chapter have been previously published in Schmidts *et al.* “Homology directed repair is unaffected by the absence of siRNAs in *Drosophila melanogaster*”, *Nucleic Acids Research*, 2016, Vol. 44, No. 17: 8261-8271.

### III.2.1 Single-strand annealing repair is unaffected by the absence of DSB-induced siRNAs

To investigate the biological function of DSB-triggered small RNAs in *Arabidopsis thaliana*, a recent publication by Wei *et al.* used a  $\beta$ -glucuronidase SSA reporter (see also II.4.1). The reporter assay indicated severely impaired SSA upon mutation of the DNA damage signaling kinase ATR, thus validating the reporter. SSA efficiency was also markedly reduced if plants were deficient for the small RNA biogenesis factors Dicer-like protein 2, 3 and 4 or AGO2. This suggests a link between small RNA biogenesis factors and DSB repair. In addition, SSA was impaired when plants lacked RNA polymerases IV and V. The latter two proteins are components of the RNA-dependent DNA methylation (RdDM) pathway. RNA polymerase IV provides a single-stranded RNA for the generation of a double-stranded RNA by the RNA-dependent RNA polymerase 2 (RDR2). siRNAs generated by this pathway are generally associated with Ago4 and mediate DNA methylation through RdDM (Matzke and Mosher, 2014). In contrast to this, RNA polymerase V transcripts provide a scaffold for interaction with siRNAs that are generated by the concerted action of RNA polymerase IV, RDR2, DCL3 and Ago4 and various silencing factors (Matzke and Mosher, 2014).

To test the putative link between SSA and DSB-induced siRNAs in *Drosophila*, an SSA reporter was established *in vitro* and *in vivo*. Similar to the *A. thaliana* DGUS.US-1 reporter (detailed in II.4.1), the SSA reporter used throughout this study is comprised of two copies of full length GFP which are separated by a restriction site for the highly specific homing endonuclease I-SceI (Pierce et al., 1999). For a schematic overview of the SSA reporter used in this study see II.4.1 and Fig. 3. In the uncut state only the first GFP open-reading frame is translated, whereas the second copy of GFP is part of the 3' UTR due to a stop codon upstream of the I-SceI restriction site. Since the poly(A) signal is provided downstream of the second GFP, expression of the construct triggers nonsense-mediated decay. This is in accordance with literature where long 3' UTRs have been described to induce NMD in *Drosophila* (Behm-Ansmant et al., 2007; Giuliani et al., 2014). In line with this hypothesis, we observed a substantial increase in GFP

levels upon knock-down of Upf1, which we assessed by Western blotting and flow cytometry (see Fig. 25).



**Fig. 25: Knock-down of the NMD factor Upf1 leads to de-repression of GFP expression from the long reporter construct.** GFP expression levels in SSA reporter were assessed by Western blotting **(A)**, or flow cytometry **(B)**. For both assays a knock-down was performed for RLuc (knock-down control), GFP or Upf1. **(A)** The Western blot membrane was probed with  $\alpha$ Tubulin (for loading control) and  $\alpha$ GFP. **(B)** The percentage of GFP high cells was measured after knock-down of the different factors and transfection of pRB2, an expression vector for the firefly luciferase that does not influence the reporter constructs and is used as a transfection control. Figure adapted by permission from Oxford University Press (Schmidts et al., 2016), copyright (2016).

### III.2.1.1 Assessing the putative link between DSB-triggered siRNAs and SSA in *Drosophila* S2 cells

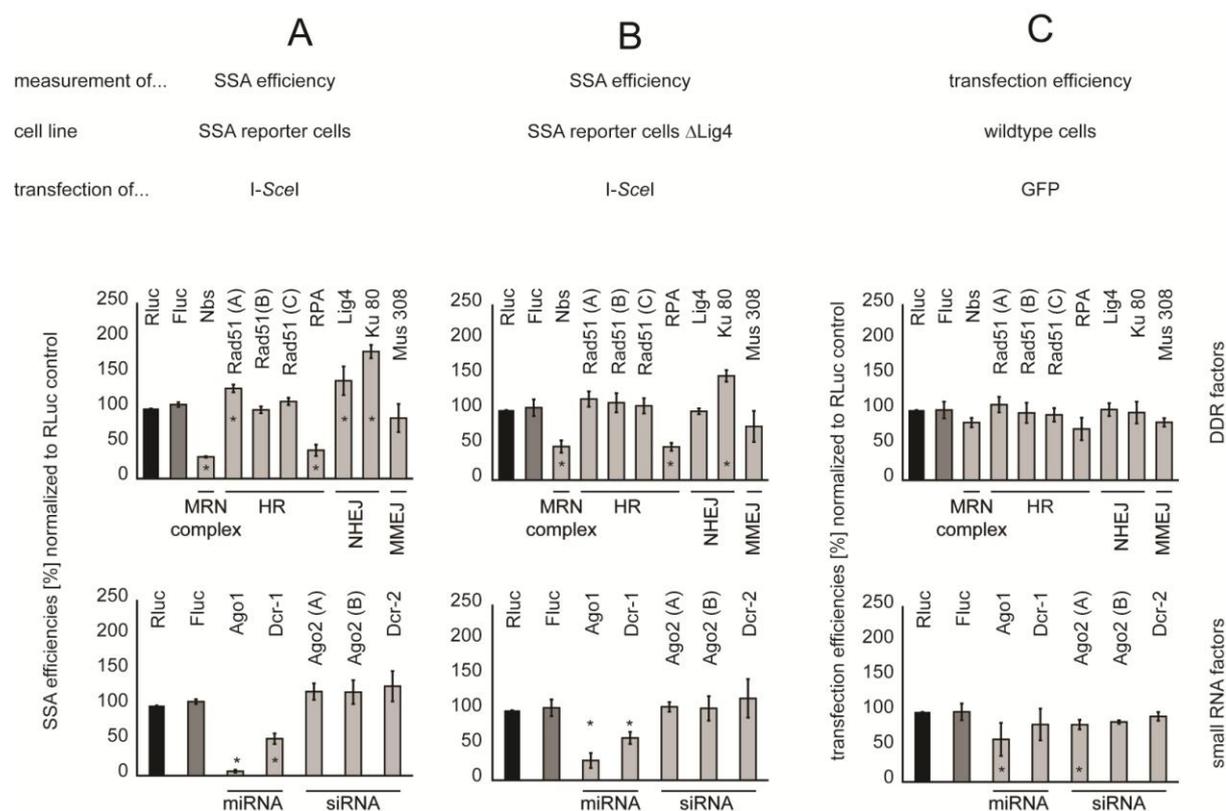
Upon introduction of the I-SceI nuclease, the SSA reporter construct is cleaved between the two GFP open reading frames (ORFs). Repair by SSA is accompanied by a substantial increase in GFP fluorescence. In contrast to this, repair via NHEJ does not result in a change in GFP fluorescence, since the 3' UTR of the construct is not considerably shortened. For all *in vitro* experiments we assessed SSA efficiencies via flow cytometry.

To ensure that the SSA reporter is able to detect changes in SSA repair rates, we first subjected reporter cells to knock-down of known DSB repair factors. Knock-down of Nbs, a component of the MRN complex, resulted in drastically reduced SSA efficiencies. We observed a similar phenotype when we depleted RPA, a well established HR factor. The fact that Rad51 is not needed in SSA for homology search (Ivanov EL, 1996) is reflected in the data. We tested

three different Rad51 dsRNA constructs, however SSA repair rates were not decreased by either one. There is even a measurable trend towards higher SSA rates upon depletion of Rad51. A similar observation was made in mammalian cells where DSB repair pathway choice was shifted towards SSA in the absence of functional Rad51 (Stark et al., 2004).

Knock-down of the NHEJ factors Ku80 and Lig4 resulted in elevated SSA repair rates, since depletion of NHEJ factors presumably shifts the repair pathway choice towards HR (Fig. 26). On the other hand, depletion of the MMEJ factor Mus308 did not result in significant changes in SSA repair rates, thus Mus308 does not seem to affect repair pathway choice in this experimental setup.

Having verified that the SSA reporter could detect changes towards both higher and lower SSA efficiencies, we probed whether depletion of miRNA pathway factors (Ago1, Dcr-1) and siRNA pathway components (Ago2, Dcr-2) (Förstemann et al., 2007; Hartig et al., 2009) would result in measurable changes in SSA repair rates. Inactivation of the miRNA biogenesis factors, but not of siRNA pathway factors resulted in an apparent reduction of SSA efficiency (see Fig. 26).



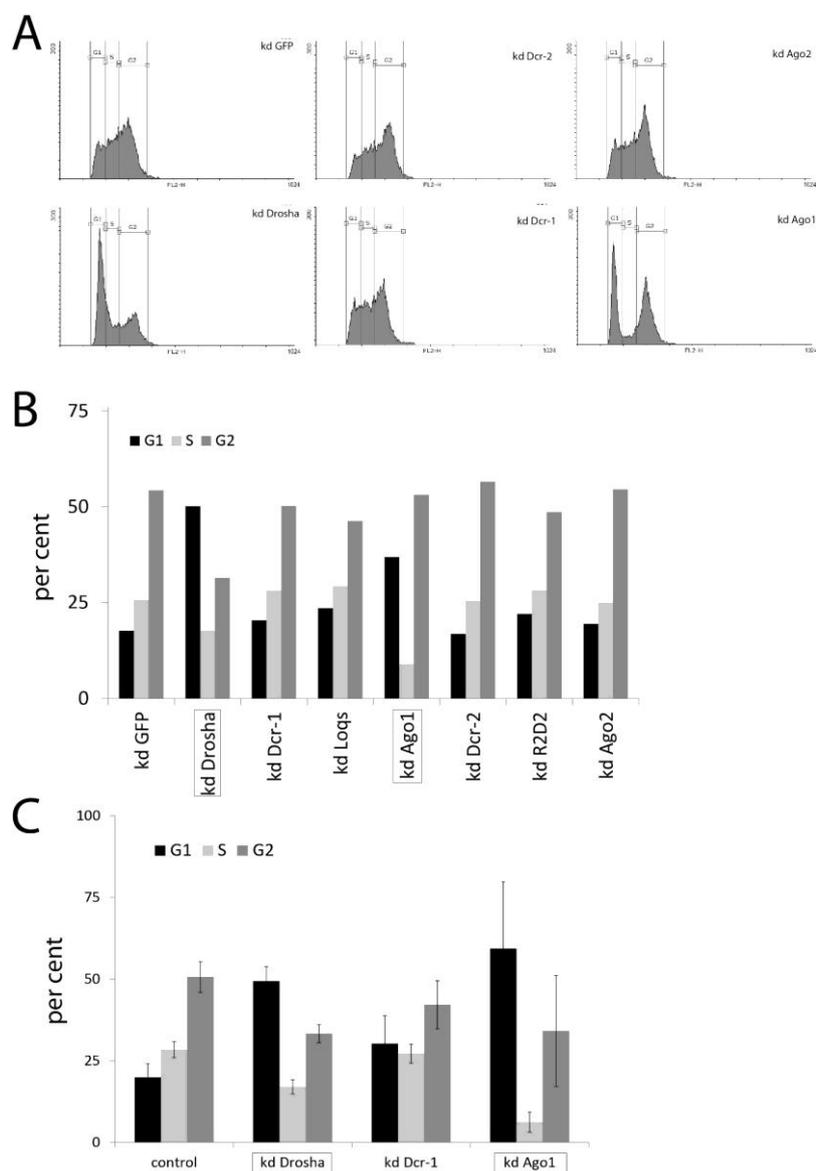
**Fig. 26: HDR efficiencies of the SSA reporter are unaffected by knock-down of siRNA factors.**

(A and B) HDR efficiencies were determined by flow cytometry after knock-down of individual factors and subsequent transfection with a plasmid encoding the *I-SceI* nuclease. Knock-down of *Renilla* luciferase (RLuc) served as knock-down control and was used for normalization. Depicted in (A) are HDR efficiencies for the parental SSA

reporter cell line, while **(B)** shows HDR efficiencies for the  $\Delta lig4$  SSA reporter cell line. **(C)** To assess the influence of knock-down of individual factors on the transfection efficiency, we transfected wild-type (S2) cells with a plasmid encoding GFP following the respective knock-down. Five days later, GFP-expressing cells were quantified via flow cytometry. Again, knock-down of RLuc was used as control for normalization. Error bars represent the standard deviation (SD); for Rad51, three independent dsRNA designs were tested, two designs for Ago2. Significant changes compared to the *Firefly* Luciferase (FLuc) are indicated by an asterisk (\*) ( $P < 0.05$ ; two-sided t-Test equal variance,  $n \geq 3$  biological replicates). Figure adapted by permission from Oxford University Press (Schmidts et al., 2016), copyright (2016).

Since the induction of a DSB depends on transient transfection of an I-SceI expression vector, we also performed a transfection control, to measure the influence of the various knock-downs on transfection efficiencies. While the observed reduction in SSA efficiency for knock-down of the MRN complex factor Nbs and the HR protein RPA, was not due to reduced transfection efficiencies, we found that the apparent reduction of SSA rates after Ago1 depletion can be partly explained by a reduced transfection efficiency.

It has been reported that miRNAs directly regulate DSB repair factor expression (Di Francesco et al., 2013; Martin et al., 2013; Yan et al., 2010). In addition to that, miRNA factor depletion can result in cell cycle abnormalities (as reviewed in (Yang and Qi, 2015)), which in turn might affect HR repair. Consistent with this hypothesis, we found that knock-down of Drosha and Ago1, but not Ago2 or Dcr-2 led to cell cycle arrest of cultured *Drosophila* cells in G1 (see Fig. 27). A similar observation was made in human cells, where the G1- to S-phase transition was inhibited by depletion of DICER (Liu et al., 2015).



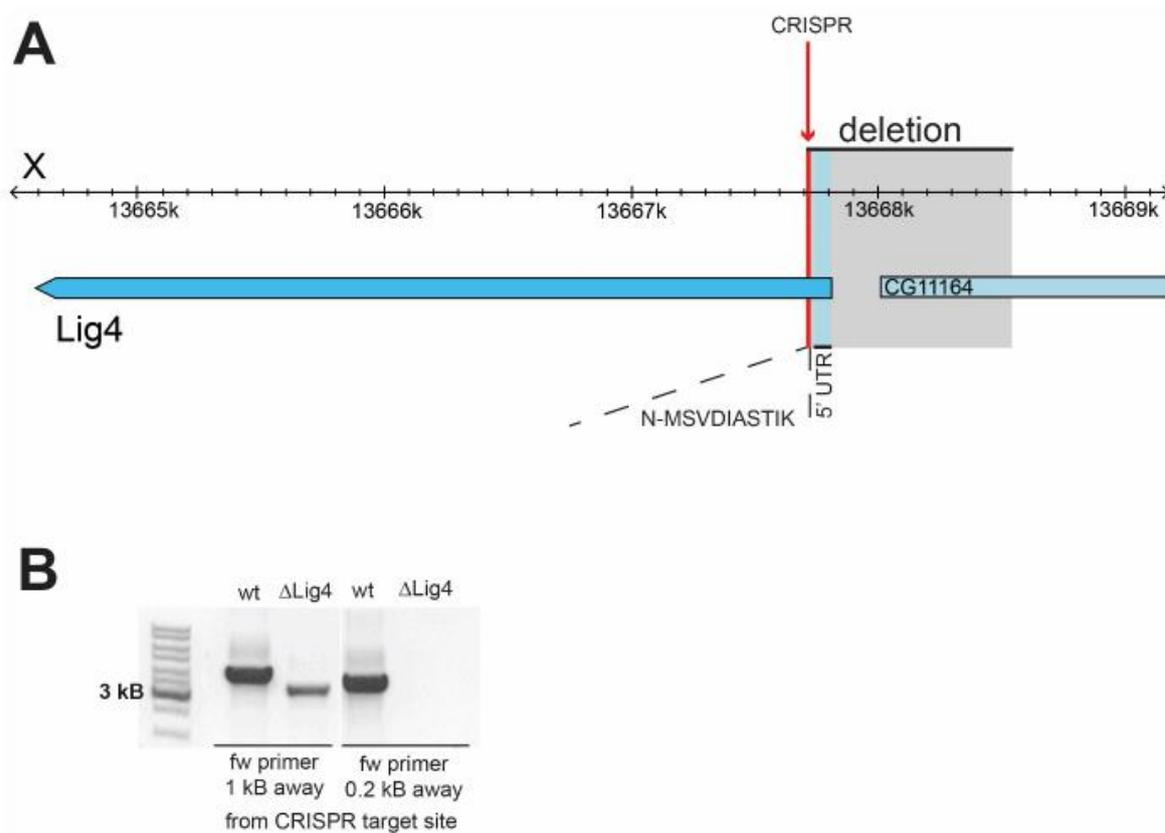
**Fig. 27: Depletion of certain small RNA biogenesis factors leads to cell cycle perturbations. (A)** Representative propidium iodide (PI) FACS profiles after control (dsGFP) or small RNA biogenesis factor knock-down; depletion of Drosha and Ago1 leads to clear accumulation in the G1-phase. **(B)** Comparative bar graph for a larger panel of cell cycle profiles following knock-down ( $n=1$ ). **(C)** The G1 accumulation after depletion of Drosha and Ago1, but not Dcr-1, is reproducible ( $n=3$ ). Experiments were conducted by Dr. Mirkovic-Hösle. Figure adapted by permission from Oxford University Press (Schmidts et al., 2016), copyright (2016).

While the results from the SSA annealing reporter clearly argue against an essential contribution of the siRNA biogenesis factors Ago2 and Dcr-2, it is not trivial to dissect the apparent decrease in SSA repair rates after knock-down of miRNA factors. Possible effects on viability, changes in the cell cycle distribution and lower transfection rates in the case of Ago1 are likely to result in the observed decrease in SSA efficiency. In addition, we cannot exclude further contribution of other miRNA-related regulatory events.

To our current knowledge, three major DSB repair pathways exist in *Drosophila*, HR (which we assess via the SSA reporter), non-homologous end-joining (which is dependent on Ku proteins and DNA ligase 4 / *Lig4*) and an alternative end-joining pathways that relies on microhomologies of 5-25 nt (McVey and Lee, 2008) and shares the initial resection steps with homology directed repair.

Most interestingly, the expression of the *Drosophila* homolog of polymerase theta (encoded by the *mus308* gene in *Drosophila*) (Chan et al., 2010; Czech et al., 2008; Okamura et al., 2008), is regulated by siRNA factors. *CG4068* is one of the long RNA hairpin loci within the *Drosophila* genome which is processed via the siRNA pathway and represses *mus308* transcripts. Thus, it is likely that knock-down of Dcr-2 might lead to up-regulation of polymerase theta. Okamura *et al.* were able to show that *mus308* transcript levels are up-regulated upon knock-down of Dcr-2 (Okamura et al., 2008). So far, there is no study on polymerase theta protein levels after Dcr-2 knock-down. Several attempts to tag *mus308* in order to investigate polymerase theta protein levels were not successful (data not shown). It is thus a formal possibility that depletion of siRNA factors leads to up-regulation of *mus308*. Although SSA is thought to be favored for homology stretches that are longer than 20 nt (Chang et al., 2017), we cannot rule out that the SSA reporter is analogously re-arranged due to a hyper-activated MMEJ system upon depletion of siRNA factors and hereby compensates putative effects of siRNA depletion on SSA.

In order to shift the balance between HR and end-joining activities towards HR to detect decreases in SSA efficiencies more easily, we used the CRISPR/*cas9* system to generate a *lig4* deletion in the reporter background. After introduction of the sgRNA expression template, cells were subjected to single cell dilutions. One of the retrieved clones showed a ~ 0.8 kB deletion, which includes the *ligase4* transcription start site and its first ten amino acids (see Fig. 28).



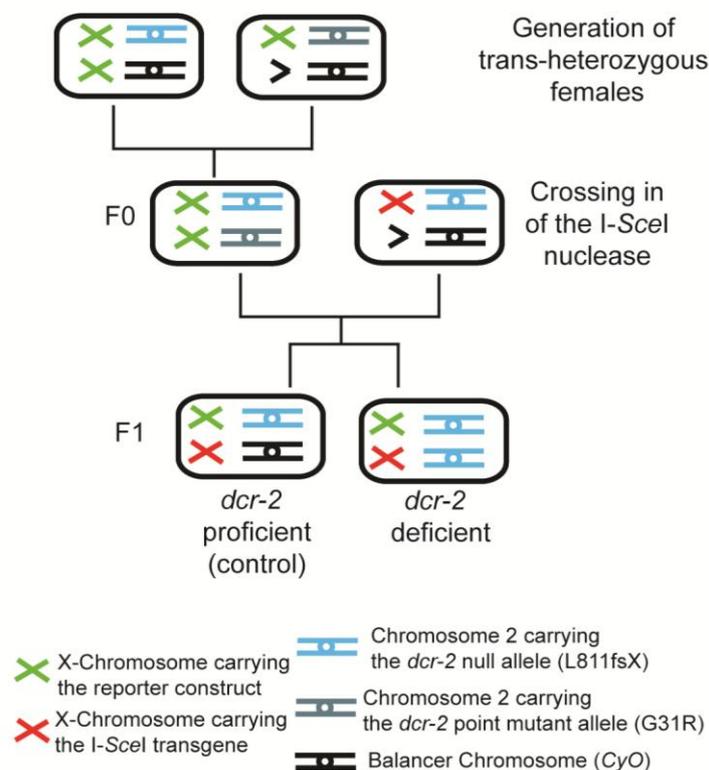
**Fig. 28: Introducing a *lig4* deletion in the SSA reporter cell line.** In order to introduce a *lig4* deletion, we designed a CRISPR sgRNA that targets the beginning of the *lig4* ORF (its position is indicated by the red line and the red arrow). **(A)** Following single cell dilution, we recovered a clone with an 827 bp deletion (grey area). The deletion also affects the 5' UTR and the sequence encoding for the first ten amino acids (lightblue) (verified by sequencing). **(B)** The genomic deletion prevents amplification of the *lig4* gene if the forward primer is chosen within the deleted region. Figure adapted by permission from Oxford University Press (Schmidts et al., 2016), copyright (2016).

In theory, the overall balance between HR and end-joining should be shifted towards HR in a *lig4* knock-out. As a result, we would expect that the potentially compensating effect of MMEJ after RNAi factor knock-down is abolished. The observation that knock-down of Lig4 in the  $\Delta lig4$  SSA reporter background did not result in changes on SSA repair rates validates our  $\Delta lig4$  cell line. However, when measuring SSA repair rates subsequent to knock-down and transfection of the I-*SceI* expression vector, we made the same observations as for the parental SSA reporter cell line (compare Fig. 26 A to Fig. 26 B). Knock-down of the essential DSB repair factors Nbs and RPA resulted in decreased SSA efficiencies. Similarly, depletion of miRNA factors led to an apparent reduction in repair rates, while SSA was unaffected by the absence of RNAi factors. Thus, we conclude that in cell culture DSB-triggered siRNAs do not contribute to repair via SSA.

### III.2.1.2 Assessing the putative link between DSB-triggered siRNAs and SSA *in vivo*

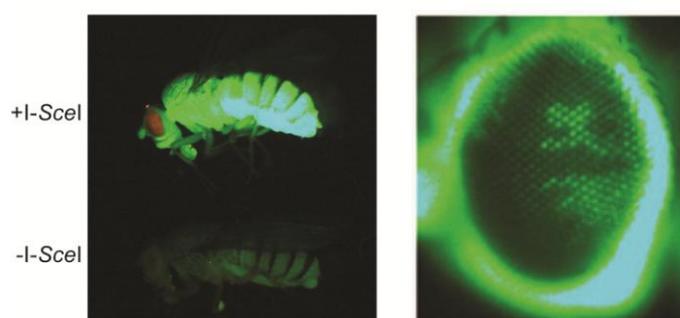
Having the SSA reporter integrated into cultured *Drosophila* cells provided us with the possibility to validate our reporter and see how depletion of various factors affects SSA efficiency. For example, we could detect an apparent decrease in SSA repair rates upon depletion of microRNA factors, which was most likely due to indirect effects on transfection efficiencies, cell cycle and viability. *In vivo*, the corresponding experiment would have been almost impossible to perform, since mutants for miRNA factors are not viable. However, when performing the knock-down of a factor, it is always a concern whether the depletion of the respective factor is sufficient to completely abrogate its function. As mentioned above (see II.4.1), we had also integrated our SSA reporter *in vivo*. We tested our hypothesis that SSA repair is unaffected by break-induced siRNAs in *dcr-2* deficient flies.

To ensure that *dcr-2* is not loaded as maternal factor into the egg (Laver et al., 2015), we set up crosses in a manner that ensured that the progeny was both, maternally and zygotically, *dcr-2* deficient (see Fig. 29 for a representative cross). Using *dcr-2* heterozygous fly strains with a SSA reporter background, we first performed a cross to obtain trans-heterozygous animals, i.e. carrying two different *dcr-2* mutant alleles. This step was necessary, since fly strains often harbor background mutations, which lead e.g. to sterility in homozygous mutants. *dcr-2* trans-heterozygous animals were then crossed with flies that stably and ubiquitously express I-*SceI* and are heterozygous for a *dcr-2* mutant allele.



**Fig. 29: Schematic overview of a representative cross for the *in vivo* SSA assay.** To ensure that *dcr-2* deficient animals were maternal and zygotic null mutants, we first set up a cross to retrieve *dcr-2* trans-heterozygote females (F0) that are also homozygous for the SSA reporter constructs. We crossed *dcr-2* trans-heterozygotes to *dcr-2* heterozygote males that carry the *I-SceI* transgene. Female F1 progeny carries the *I-SceI* transgene as well as the SSA reporter. We investigated the role of *dcr-2* in fully deficient *dcr-2* F1 females and used the *dcr-2* proficient (*dcr-2/CyO*) females for control. Figure adapted by permission from Oxford University Press (Schmidts et al., 2016), copyright (2016).

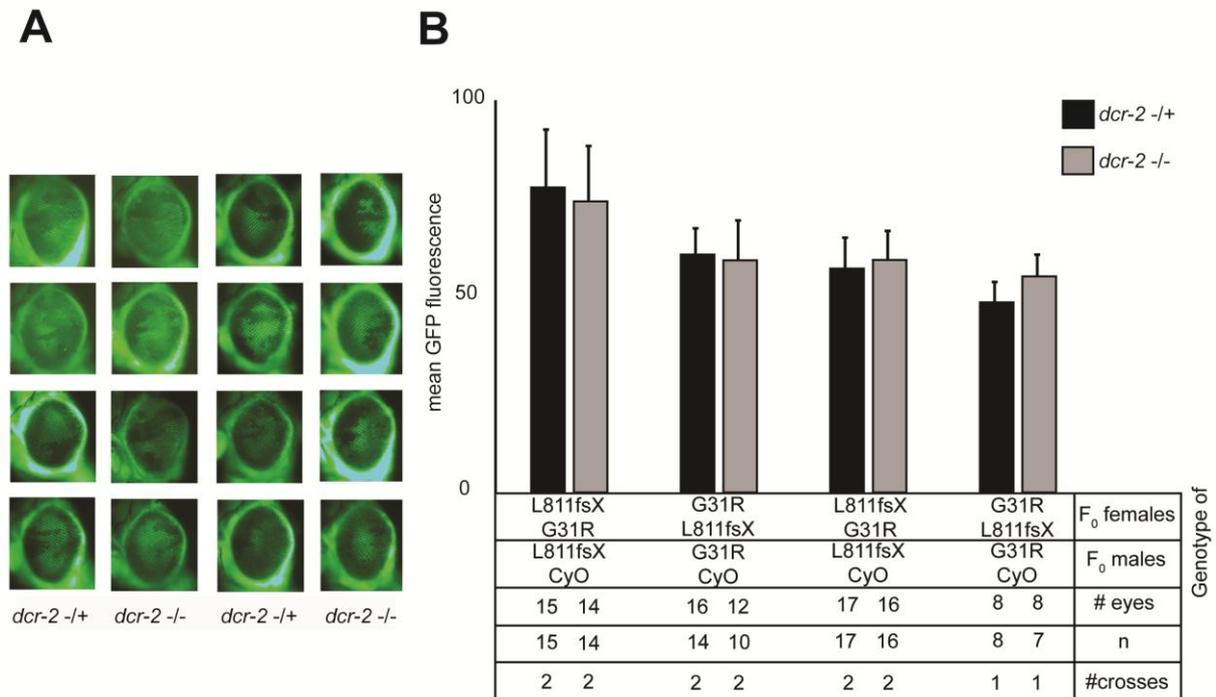
Upon activation of the zygotic genome, *I-SceI* will introduce a DSB between the two copies of GFP. Only repair by SSA, but not by NHEJ, will result in patches with high GFP fluorescence in the F1 adults. While individual patches are not discernible in the fly body, the fly eye shows a mosaic pattern of highly and lowly fluorescent areas. It is thus a good proxy to assess the balance between repair by SSA and end-joining (Fig. 30).



**Fig. 30: *I-SceI* expressing animals are mosaic and allow assessing the balance between SSA and end-joining repair pathway choice.** If the DSB introduced by *I-SceI* is repaired by SSA, this results in a substantial increase in GFP fluorescence. However, animals are mosaic. Individual patches have undergone SSA, while other areas underwent repair by end-joining. While individual patches are not discernible in the fly body, the fly eye allows investigating the balance between SSA and NHEJ. Figure adapted by permission from Oxford University Press (Schmidts et al., 2016), copyright (2016).

We first tested SSA repair efficiencies with two *dcr-2* mutant alleles that were retrieved in a mutagenesis screen for RNAi mutants (Lee et al., 2004). The *dcr-2*<sup>L811fsX</sup> allele carries a frameshift mutation that results in a premature stop codon and no Dcr-2 protein is expressed from this allele (Lee et al., 2004). On the other hand, Dcr-2 protein can be expressed from the *dcr-2*<sup>G31R</sup> allele, however, the point mutation lies within the ATP binding site of the helicase domain and renders *dcr-2* catalytically inactive (Lee et al., 2004). We set the cross up as depicted in Fig. 29 and as described above, and tested all possible orientations (see Fig. 31). We subjected *dcr-2* deficient animals and the heterozygous controls to fluorescence microscopy and quantified the mean GFP fluorescence within the fly eye. Comparison of *dcr-2* deficient (*dcr-2* *-/-*) to *dcr-2* proficient (*dcr-2* *+/-*) animals did not reveal significant differences (see Fig. 31). This is consistent with the cell

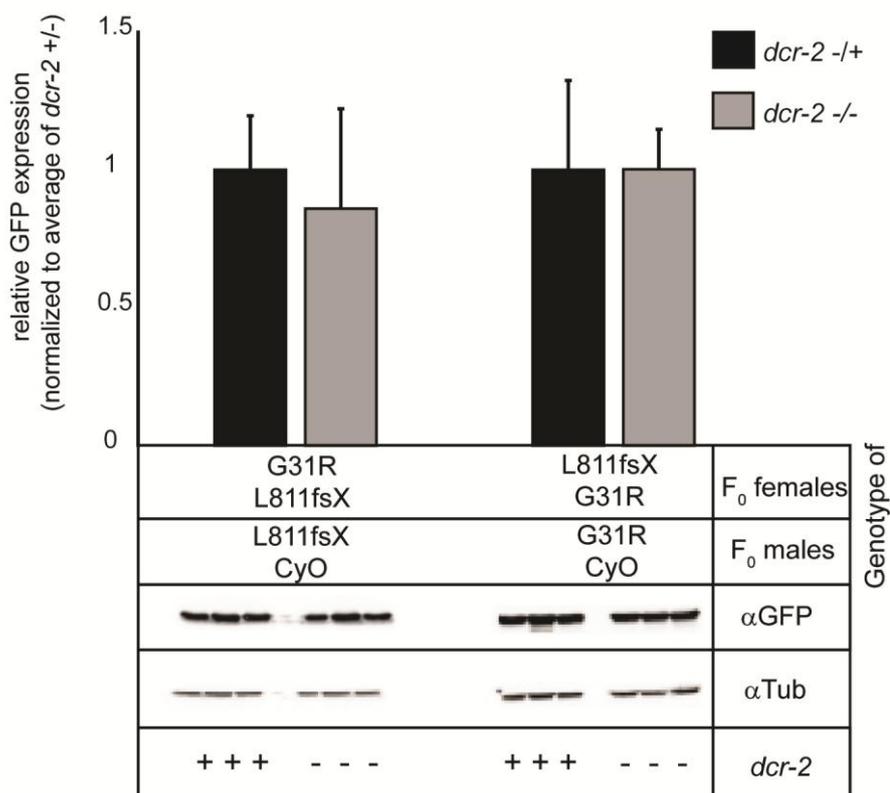
culture assay (see also III.2.2.1), where we also observed that SSA repair is unaffected by the absence of RNAi factors.



**Fig. 31: Measurement of mean GFP fluorescence within the fly eyes of *dcr-2* deficient animals and heterozygous control. (A)** Representative images of fly eyes of *dcr-2* proficient (*Dcr-2*  $-/+$ ) or *dcr-2* deficient (*dcr-2*  $-/-$ ) animals. **(B)** Quantification of the mean GFP fluorescence within the fly eye. We used ImageJ to measure the mean GFP fluorescence. The orientation of the crosses, the number of eyes and animals is indicated below the graphs, as is the number of independent crosses. Black bars represent results from heterozygous controls, depicted in grey are results for *dcr-2* deficient animals. Figure adapted by permission from Oxford University Press (Schmidts et al., 2016), copyright (2016).

We could validate this finding using an additional *dcr-2* mutant allele C473Y (Lee et al., 2004) (data not shown).

In order to obtain an independent assessment of GFP expression, we measured GFP levels of whole fly extract from individual flies. However, the differences in GFP expression between individual animals were more pronounced than between *dcr-2* mutants and control animals (Fig. 32).



**Fig. 32: Analysis of mean GFP protein expression of *dcr-2* proficient and deficient animals assessed by Western blot.** Protein extracts from individual flies were subjected to Western blotting. The blot was probed with antibodies against GFP and Tubulin (loading control to allow for quantification). Bars represent quantification of the Western blot band signals. Depicted in black is the mean GFP protein expression of heterozygous control animals, shown in grey the mean GFP expression of *dcr-2* deficient animals. Figure adapted by permission from Oxford University Press (Schmidts et al., 2016), copyright (2016).

In summary, the results from the SSA reporter argue against an involvement of break-derived siRNAs, both in cell culture and *in vivo*.

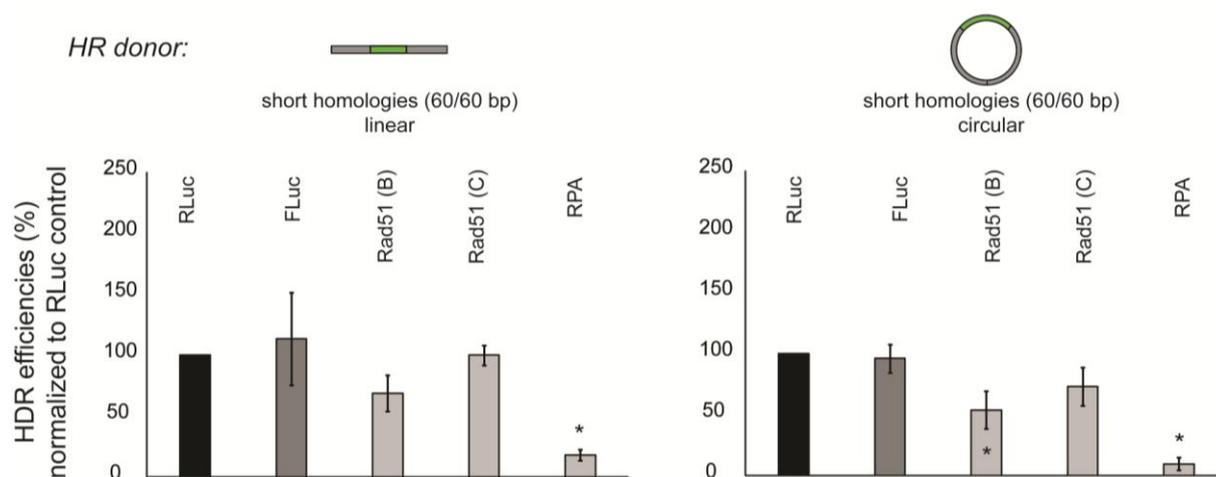
### III.2.2 Rad51-dependent HDR repair is unaffected by the absence of DSB-induced siRNAs

Publications by Wei *et al*, Gao *et al* and Francia and colleagues suggest a function of DSB-triggered siRNAs in Rad51-dependent repair and DNA damage signaling in cultured mammalian cells (Francia, 2012; Gao et al., 2014; Wei et al., 2012). To probe for the possible influence of DSB-induced small RNAs, all three reports made use of the DR-GFP reporter, which was introduced 1999 by the Jasin laboratory (see II.4.2 for details on the reporter). The uncut reporter construct yields no GFP fluorescence. However, if repair occurs via intra-chromatid

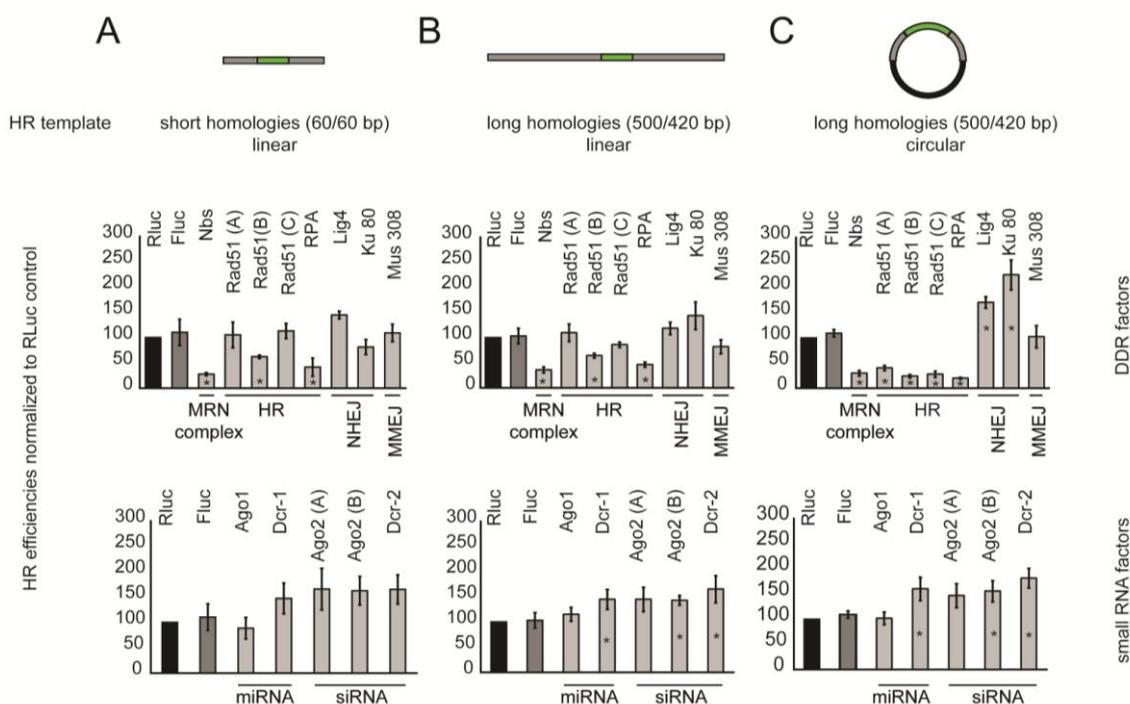
recombination or unequal sister-chromatid recombination, GFP is expressed, which allows for readout using flow cytometry. In the above mentioned studies, depletion of DICER, DROSHA and AGO2 led to a reduction of GFP expression compared to a knock-down control.

In order to test whether this interplay between DSB-triggered small RNAs and Rad51-dependent HDR is conserved in *Drosophila*, we considered our approach for C-terminal tagging by use of the CRISPR/*cas9* system as a reporter. As described previously (Böttcher et al., 2014; Kunzelmann et al., 2016), we programmed Cas9 to introduce a break adjacent to the stop codon at a gene of interest. For this study we chose the *act5C* gene, since it is highly expressed and was previously used as model locus to establish the C-terminal tagging protocol (Böttcher et al., 2014; Kunzelmann et al., 2016). We co-transfected different HR donors together with the sgRNA expression template. Every tested donor was comprised of homology arms upstream and downstream of the break which flank a GFP coding sequence. Integration of the HR donor at the *act5C* gene results in a fluorescence signal from the Act5C-GFP fusion protein and allows for detection of integration events by flow cytometry (Böttcher et al., 2014; Kunzelmann et al., 2016). There are reports in literature that the dependence of HR on Rad51 correlates with homology length. For example in yeast, Rad51-independent HR repair is favored for homologies that are shorter than 100 nt (Ira and Haber, 2002). Therefore we varied the length of the homology stretches upstream and downstream of the break (60/60 nt and 500/420 nt). In addition to that, we tested the influence of HR donor topology (linear and circular).

We first probed whether we could detect changes in HDR repair rates via our GFP knock-in approach. Knock-down of the MRN complex member Nbs resulted in a substantial decrease in HDR efficiencies, as did depletion of RPA. The results from the different HR donor lengths and topologies give insight into the HR donor requirements of Rad51. Upon knock-down of Rad51 with three different independent Rad51 dsRNA constructs, we found no strong effect with linear donor molecules, independent from the homology arm length. The relative HR efficiencies did not fall below  $62.2 \pm 3.4 \%$  (short homologies, Figure 34 A) and  $64.01 \pm 4.5 \%$  (long homologies, Figure 34 B) of the control treatment. However, we observed a substantial drop in HDR efficiencies if we provided a circular HR donor with long homology arms. In this experimental setup HR efficiencies dropped to  $23.8 \pm 3.1 \%$  of the control treatment (Figure 34 C). In a separate experiment we also investigated whether the dependence on Rad51 correlates solely with the topology of the HR donor. We circularized the HR donor with the short (60/60 nt) homology arms (see also Fig. 33) and compared linear vs. circular HR donors with short homology stretches. As described above, there was no strong dependence on Rad51 (efficiencies did not drop beyond  $68.74 \%$  (linear) and  $54.03 \%$  (circ) compared to the knock-down control).



**Fig. 33: Assessing the influence of HR donor topology on the dependence of HDR repair on Rad51.** Top: schematic representations of the provided HR donors. We transfected an sgRNA expression construct targeting *act5C* into S2 cells that stably express Cas9 protein. The HR donors depicted above were co-transfected, HDR events are indicated by the expression of Act5C-GFP fusion protein. Readout was performed by flow cytometry. The percentage of Act5C-GFP positive cells to all viable cells was quantified. Knock-down of Renilla luciferase (RLuc) served as control for normalization. For the experiment we tested two different Rad51 dsRNA constructs. We performed a t-Test to detect significant changes compared to the *Firefly* luciferase (FLuc) knock-down control. Significant changes are indicated by an asterisk (\*) ( $P < 0.05$ ; two-sided t-Test, equal variance;  $n = 3$  biological replicates). Left: a linear donor with short homologies (60/60 nt) was provided for repair. We could only detect significant changes for depletion of RPA. Right side: The short homology donor was circularized, however the results mirror the results we obtained for transfection of the linear donor.



**Fig. 34: HDR efficiencies of genome editing are unaffected by knock-down of siRNA factors.** Top: schematic representation of the HR donors employed; homology arms are not drawn to scale. S2-cells stably

expressing the Cas9 protein were transfected with an sgRNA expression construct and the indicated homology donor to append a GFP moiety to the *act5C* gene at its chromosomal locus. Seven days later, cells were analyzed by flow cytometry and the Act5C-GFP positive population was quantified (absolute values were 3–8 % for the RLuc control). Knock-down of *Renilla* luciferase (RLuc) served as control for normalization. We tested three independent knock-down constructs for Rad51, two for Ago2. The asterisk (\*) indicates significant changes ( $P < 0.05$ ; two-sided t-Test equal variance,  $n \geq 3$  biological replicates). **(A)** PCR-based HR donor with 60 nt of homologous sequence at either end; although the MRN complex and RPA appear to be required for homology-directed repair with this donor, the role of Rad51 is limited. Only one knock-down construct resulted in a significant reduction; for the small RNA biogenesis factors no significant change was observed. **(B)** Extension of the homology arms in the linear configuration (targeting vector pIS1 linearized) results in values largely comparable with the ones observed in (A). For the small RNA factors, a slight apparent increase of HR efficiency is detected; we assume this to be a technical artefact because detection of Act5C-GFP is facilitated in the absence of repressive siRNAs generated at the break. **(C)** If the HR donor is supplied as a circular targeting vector, the dependence on Rad51 is far more pronounced. In this case, all three knock-down constructs resulted in a highly significant reduction of HR efficiency. For the small RNA biogenesis factors, however, the results are essentially unchanged. Figure adapted by permission from Oxford University Press (Schmidts et al., 2016), copyright (2016).

As observed for our SSA reporter (compare Fig. 26), knock-down of NHEJ factors significantly increased HDR efficiencies when the HR donor was circular and provided long homology stretches. The ability to detect changes in HDR efficiency and the dependence on Rad51, when a circular HR donor with long homologies is provided, validates our tagging approach as reporter for HDR and allows comparison of our findings to the studies conducted in the mammalian cells. In contrast to previous reports, we did not observe any reduction in HR rates upon knock-down of RNAi factors. If anything, HR efficiencies appeared to be elevated upon depletion of the siRNA factors Dcr-2 and Ago2 (Fig. 34). This observation is likely due to a technical reason. As reported in (Kunzelmann and Förstemann, 2017), siRNAs are generated upon introduction of a tag and lead to repression of the GFP fusion protein. Detection of Act5C-GFP fusion protein is thus facilitated in the absence of the RNAi factors Dcr-2 or Ago2 (Kunzelmann and Förstemann, 2017). It also should be noted that in this experimental setup we assessed initial recombination frequencies without previous marker selection. Taken together, we did not find a link between DSB repair and DSB-triggered siRNAs in *Drosophila*, as was the case for the SSA reporter assay.

### **III.2.3 The DSB stress response is unaffected by the absence of DSB-triggered siRNAs**

In the study so far, we introduced only single DSBs at defined loci by either the endonuclease I-*SceI* or by programming Cas9 to cut at the *act5C* gene. In contrast to previous reports, we did not

---

observe an impact on repair efficiencies in the absence of RNAi factors neither if repair occurred via a donor *in cis* (SSA) nor *in trans* (knock-in of GFP). We next investigated whether Dcr-2 might contribute to the general DSB stress response. Treatment with the topoisomerase inhibitor camptothecin (CPT) leads to the formation of a covalent link between the nicked DNA and topoisomerase I. Nicks can thus not be re-ligated, and the complex poses a steric hindrance to the replication machinery. Ultimately, numerous DSBs are formed throughout the genome due to collapsed replication forks. *brca2* encodes an important factor in Rad51-dependent HR repair and accordingly, *brca2* mutant flies show hypersensitivity to camptothecin (Thomas et al., 2013). We hypothesized that we would observe a similar phenotype if siRNA factors (e.g. Dcr-2 and Ago2) indeed played an essential role in HR repair.

To detect potential hypersensitivity, indicated by an impaired development into adult flies, we supplemented fly food with CPT at varying concentrations or the solvent control (DMSO). We noted the number of RNAi proficient and deficient progeny retrieved from a cross of heterozygous parents. We then calculated the proportion of RNAi defective mutant flies to investigate a potential involvement of siRNA factors in the general DSB stress response. The results from this CPT screen are represented in Table 8.

We first set up a cross to obtain flies that carry a wild-type chromosome 2 over the balancer *CyO*. From this cross we retrieved the expected ratio of straight-winged progeny. We can thus exclude that the balancer chromosome *CyO* confers resistance to CPT treatment (Table 8, cross 1). When we crossed flies that were heterozygous for the catalytically inactive *ago2*<sup>V966M</sup> allele (Kim et al., 2007) with flies harboring the P-element insertion allele *ago2*<sup>321</sup> (Hain et al., 2010)(Table 8, cross 9) we also did not observe a phenotype under CPT treatment. In contrast to this, we observed severely impaired hatching of *dcr-2* deficient animals when we crossed heterozygous *dcr-2* mutant flies (Table 8, cross 2 and cross 3). This was the case for crosses that carry the frameshift to stop codon mutation *dcr-2*<sup>L811fsX</sup> or the point mutation within the helicase domain *dcr-2*<sup>G31R</sup>, both of which we also used to investigate the role of Dcr-2 in SSA (compare Fig. 29, 31 and 32 as well as III.2.1.1).

**Table 8: Assessing the general DSB stress response of *dcr-2* and *ago2* mutant flies upon treatment with camptothecin.**

(*CyO* and *TM3* are homozygous lethal balancer chromosomes with functional *Dcr-2* or *Ago2* alleles and a visible marker). Table adapted by permission from Oxford University Press (Schmidts et al., 2016), copyright (2016).

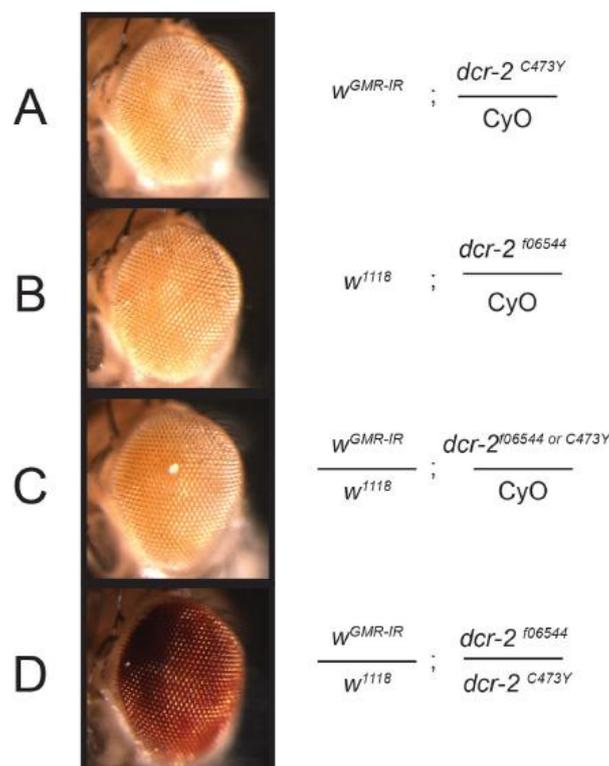
Cross	genotype	phenotype	expect. ratio	untr. control	c (CPT) in mM				
					0 (DMSO)	0.01	0.03	0.1	0.3
1	+/ <i>CyO</i>	<i>Cy</i>	2	75	43	112	142	69	84
	x	<i>non-Cy</i>	1	54	38	94	70	38	52
	+/ <i>CyO</i>	# ind. crosses		(1)	(2)	(3)	(3)	(2)	(3)
2	<i>dcr-2<sup>G21R</sup></i> / <i>CyO</i>	<i>Cy</i>	2	234	236	105	74	220	131
	x	<i>non-Cy</i>	1	100	141	62	41	21	0
	<i>dcr-2<sup>L811βX</sup></i> / <i>CyO</i>	# ind. crosses		(3)	(2)	(2)	(2)	(3)*	(3)(*)
3	<i>dcr-2<sup>L811βX</sup></i> / <i>CyO</i>	<i>Cy</i>	2	155	157	8	100	219	135
	x	<i>non-Cy</i>	1	61	82	2	5	2	0
	<i>dcr-2<sup>L811βX</sup></i> / <i>CyO</i>	# ind. crosses		(3)	(2)	(1)	(2)*	(3)*	(3)(*)
4	<i>dcr-2<sup>L811βX</sup></i> / <i>CyO</i>	<i>Cy</i>	1	219	186	183	117	31	20
	x	<i>non-Cy</i>	1	191	163	196	68	1	0
	<i>dcr-2<sup>L811βX</sup></i> ;P{ <i>Dcr-2</i> }	# ind. crosses		(4)	(3)	(5)	(4)	(1)*	(1)(*)
5	<i>dcr-2<sup>L811βX</sup></i> / <i>CyO</i>	<i>Cy</i>	2	62	55	45	108	126	23
	x	<i>non-Cy</i>	1	41	36	19	37	4	0
	<i>dcr-2<sup>Dj(2R)ED3385</sup></i> / <i>CyO</i>	# ind. crosses		(3)	(3)	(2)	(3)	(3)*	(2)(*)
6	<i>dcr-2<sup>Dj(2R)ED3385</sup></i> / <i>CyO</i>	<i>Cy</i>	1	68	49	77	112	95	117
	x	<i>non-Cy</i>	1	81	87	112	114	8	0
	<i>dcr-2<sup>L811βX</sup></i> ;P{ <i>Dcr-2</i> }	# ind. crosses		(2)	(3)	(3)	(3)	(3)*	3(*)
7	<i>dcr-2<sup>06544</sup></i> / <i>CyO</i>	<i>Cy</i>	2	93	81	66	18	12	62
	x	<i>non-Cy</i>	1	51	43	37	23	9	25
	<i>dcr-2<sup>Dj(2R)ED3385</sup></i> / <i>CyO</i>	# ind. crosses		(2)	(2)	(2)	(2)	(2)	(2)
8	<i>dcr-2<sup>06544</sup></i> / <i>CyO</i>	<i>Cy</i>	2	175	92	100	230	146	137
	x	<i>non-Cy</i>	1	83	52	69	135	89	66
	<i>dcr-2<sup>L881βX</sup></i> / <i>CyO</i>	# ind. crosses		(4)	(3)	(3)	(5)	(3)	(4)
9	<i>ago2<sup>321</sup></i> / <i>TM3,Sb</i>	<i>Sb</i>	1	51	85	49	72	60	58
	x	<i>non-Sb</i>	1	64	89	56	111	91	52
	<i>ago2<sup>V966M</sup></i>	# ind. crosses		(2)	(2)	(2)	(3)	(3)	(3)

\*distribution different from expected,  $P < 10^{-5}$ ,  $\chi^2$  test.

(\*)no exact  $\chi^2$  test calculation possible because of 0 observed events in one category.

Attempts to rescue this phenotype using an ectopically placed and functionally validated *dcr-2* rescue transgene were not successful. This and the fact that both of the alleles tested (individually and in the trans-heterozygous combination) were derived from the same genetic screen (Lee et al., 2004) argues against a direct involvement of Dcr-2 in the DSB stress response and rather for a background mutation present in the strain that was mutagenized. Therefore, we combined the alleles with independent *dcr-2* mutant chromosomes. The deficiency strain *Df(2R)ED3385* harbors a deletion of ~118 kb that also includes the *dcr-2* locus. Again, we observed CPT hypersensitivity when we performed the cross with flies carrying the *dcr-2*<sup>811fsX</sup> allele (Table 8, cross 5) and again, we did not succeed in rescuing the phenotype with the wild-type *dcr-2* transgene (Table 8, cross 6).

The second independent *dcr-2* mutant chromosome carries a piggyBAC transposon insertion 29 nt downstream of the annotated transcription start site. We validated that the *dcr-2*<sup>06544</sup> allele is defective in transgenic RNAi interference (Fig. 35).

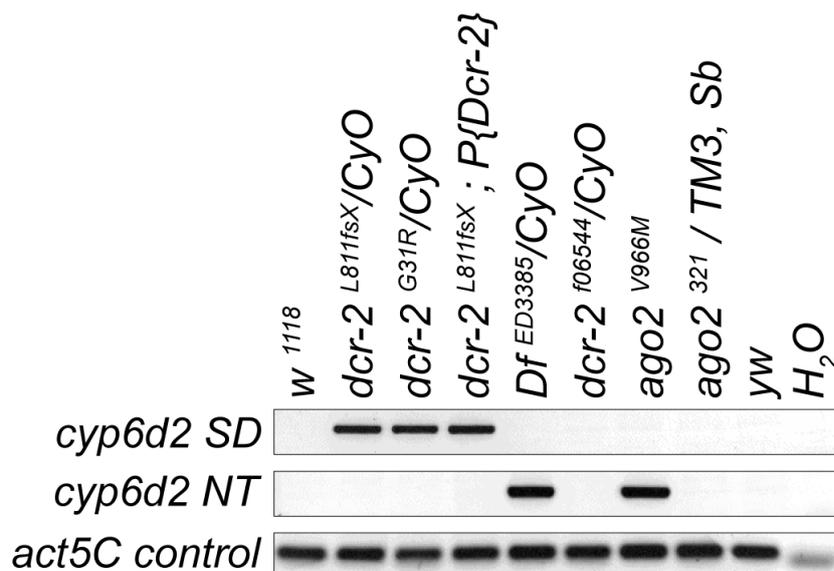


**Fig. 35: The *dcr-2*<sup>06544</sup> allele has impaired function in transgenic RNA interference.** (A) The tester strain contains an inverted repeat transgene on the X-chromosome together with a wild-type *w* gene. This triggers RNA interference and results in light orange eye-color as long as the RNAi pathway is functional. To test for Dcr-2 function, the strain carries a non-functional point mutation (C473Y) together with a *CyO* balancer chromosome. (B) The *dcr-2*<sup>06544</sup> allele was crossed to the tester strain described in (A); this line has a mutant *w*<sup>1118</sup> allele on the X-chromosome but light orange eye color because of the *mini-w* marker gene embedded in the piggyBAC mutagen that causes the *dcr-2*<sup>06544</sup> allele. (C) Heterozygous offspring from the cross can be identified because of the *CyO* balancer

chromosome; we could not detect any difference between heterozygous female offspring. The *dcr-2* allele can therefore not be determined phenotypically. **(D)** The dark red eye color clearly indicates that trans-heterozygous *dcr-2<sup>C473Y</sup>/dcr-2<sup>06544</sup>* female offspring from the cross is deficient in transgenic RNAi directed against the *w* gene. Figure adapted by permission from Oxford University Press (Schmidts et al., 2016), copyright (2016).

Trans-heterozygous combination of *dcr-2<sup>06544</sup>* with either the deficiency *Df(2R)ED3385* or *dcr-2<sup>L811fsX</sup>* did not show increased CPT sensitivity (Table 8, cross 7 and 8).

We next tested our hypothesis that some of our *dcr-2* mutant fly strains carry a background mutation. We reasoned, based on the results we obtained from experiments using the deficiency strain *Df(2R)ED3385*, that such a background mutation should reside within the deleted region. Alternatively, the *Df(2R)ED3385* might carry the same background mutation elsewhere on the chromosome as well. For example, McVey and colleagues recently reported the isolation of commonly found *cyp6d2* mutant alleles which led to a CPT hypersensitivity phenotype in various fly stocks (Thomas et al., 2013). Since this might provide an explanation for our observations, we tested the fly strains for the two known mutant *cyp6d2* alleles that show either impaired splicing (SD – splicing defective) or inhibited transcript accumulation (NT – no transcript) (Thomas et al., 2013). We repeated the published PCR assay to screen the stocks used in our CPT experiment for *cyp6d2* deficiency. Indeed, we could detect the SD allele in the *dcr-2<sup>L811fsX</sup>* strain, the strain carrying the *dcr-2* rescue transgene as well as the *dcr-2<sup>G31R</sup>* strain. In addition to this, we found the NT allele to be present in the *Df(2R)ED3385* strain and the *ago2<sup>V966M</sup>* strain. In accordance with the results from the CPT hypersensitivity assay, the *dcr-2<sup>06544</sup>* piggyBAC allele and the *ago2<sup>321</sup>* allele did not carry either *cyp6d2* variant.



**Fig. 36:** PCR-based assay to detect *cyp6d2* splicing defective (SD) or no transcript (NT) variants. We performed the published PCR assay to screen the fly strains used throughout this study for common variants of the

*cyp6d2* gene using the previously described PCR primers and assay conditions (Thomas et al., 2013). The genotype of the respective fly strain is indicated above. We performed an *act5C* control to ensure the integrity of gDNA retrieved from flies. Figure adapted by permission from Oxford University Press (Schmidts et al., 2016), copyright (2016).

Taken together, we conclude that RNAi factors do not confer resistance to DSB stress. All genetic combinations that displayed a hypersensitivity to CPT were deficient for *cyp6d2*. This underlines the results we obtained from the SSA assay and the GFP-knock-in approach to assess HDR rates: faithful DSB stress response and break repair occurs in the absence of break-derived siRNAs.

---

## IV Discussion

### IV.1 Further optimization is needed to improve specific enrichment of reads mapping to RNAPII transcribed loci in NET-Seq libraries

Ribosomal RNA contamination is a very common obstacle in establishing small RNA library protocols (Matts et al., 2014). Theoretically, the RNAPII immunoprecipitation should be sufficient to deplete ribosomal RNAs and to specifically enrich for nascent transcripts, since it includes several washing steps. However, it is difficult to assess the affinity of the magnetic purification beads for contaminating RNA species. We performed a direct comparison between Protein G beads (DynaBeads) and Strep-Tactin Type2HC beads (IBA), but did not find major differences in the obtained signal-to-noise ratio. There is a variety of experimental approaches that might allow for depletion of background. For example addition of nucleases not only during lysis steps but also during the immunoprecipitation could improve the signal-to-noise ratio. Nascent transcripts would be protected by the active center of RNA polymerase II. However this approach has some caveat, only 20-30 nt (Brabant and Acheson, 1995; Choder and Aloni, 1988; Gu et al., 1996) of the nascent transcript are protected by the polymerase complex. Nuclease addition will degrade protruding parts of nascent transcripts and make it more difficult to discern between *bona fide* nascent RNA and degradation products or e.g. mature siRNAs. A different approach might be to include a fractionation step in the lysis protocol and deplete cytosolic mature rRNAs prior to nuclear extraction. In addition to this, there are several protocols to deplete specific rRNA species by base complementarity. Addition of DNA oligos complementary to 2S rRNA and subsequent treatment with RNase H to specifically degrade RNA-DNA 2S rRNA hybrids represents a possible approach (Li et al., 2009).

It might be feasible to even include such a 2S rRNA depletion step in the context of the immunoprecipitation or the immunoprecipitation washing steps. Another common approach is using biotinylated DNA oligos that are immobilized to magnetic beads. The supernatant is thus depleted for 2S rRNA (Seitz et al., 2008). Furthermore, there is a recent publication that suggests

depletion of 2S rRNA in the context of 5' ligation using terminator oligos that are complementary to 2S rRNA (Wickersheim and Blumenstiel, 2013) if there are concerns about additional enzymatic steps (possible source of contamination) or the use of magnetic beads (potential loss of material).

Mature tRNAs have a length between ~ 76-90 nt. Although we selected for ~22-50 nt long RNA species for unligated RNA and for ~ 30-100 nt long RNA species after 3' ligation on an urea-PAGE, this might not be selective enough and might explain tRNA contamination.

#### **IV.2 Further optimization of the NET-Seq protocol is needed to obtain a pattern of DSB-triggered antisense transcription**

We introduced DNA double-strand breaks at three different loci to gain insight in the early biogenesis steps of the long dsRNA precursor molecule, which is a prerequisite to siRNA production. Since we know from a genome-wide screen that there is a genetic requirement for various factors of the splicing machinery in DSB-induced siRNA production (Merk et al., 2017), we compared two RNAPII transcribed loci: *CG15098* and *Tetp*. The *CG15098* gene contains three introns and showed a robust siRNA response to a DSB (Merk et al., 2017). Using our NET-Seq approach, we could detect antisense transcripts for the locus. Their number is about 10 fold lower than for the sense transcripts mapping to *CG15098*. However, we detected antisense transcripts even in the absence of a break. In response to a DSB the fraction of antisense to all gene matching reads increased. This might be an indication that there is indeed recruitment of RNA polymerase II to the break. Yet, we were not able to obtain an antisense transcription profile at the break site. Aside to the drawback of the poor signal-to-noise ratio, antisense transcription might also be a rare event.

Using a different approach, namely strand specific RT-qPCR, to detect sense and antisense transcripts in the presence/absence of a break in *CG15098*, Anna Gatz could detect antisense transcription in two out of five replicates (Gatz, 2017). This is a further indication that DSB-triggered antisense transcription is rare or the half-life of antisense transcripts is very limited.

It should also be noted that the NET-Seq experiment allows a 'snap'-shot of ongoing transcription at a given time point. In contrast to this, the siRNA response to a break was measured several days after introduction of a DSB. siRNAs are stable, accumulate over time and can therefore be readily measured by small RNA sequencing (Merk et al., 2017).

We harvested cells 48 h after transfection with the sgRNA expression template, and while the T7 endonuclease assay showed that breaks were introduced and repaired by error-prone repair pathways, it remains elusive whether the chosen time frame was optimal in terms of RNAPII recruitment to the break site. A time-resolved RNAPII ChIP experiment might be conducted to probe RNAPII recruitment to the break. In addition to that, it is also possible that antisense transcription might be cell-cycle dependent. When antisense transcripts hybridize with complementary sense transcripts, they become a substrate for Dcr-2. There have been reports that RNAPII and Dicer interact in human cells (White et al., 2014), which might also be the case for *Drosophila* cells. Mature antisense transcripts might thus be a very unstable species. Repetition of the NET-Seq experiment under a Dcr-2 knock-down condition should increase the half-life of antisense transcripts. In theory, the immunoprecipitation step in the NET-Seq protocol allows for purification of protected, nascent antisense transcripts. However, in this study, the recovery of nascent antisense transcripts was not sufficient to obtain an antisense transcription profile.

For *Tctp*, an intronless gene, no antisense transcripts were detected, even though we obtained an overall higher number of reads mapping to this locus compared to *CG15098*. This is consistent with the previous finding that a DSB in *Tctp* does not elicit an efficient siRNA response (Merk et al., 2017). Also, no antisense transcripts could be detected by strand-specific RT-qPCR in response to a break in *Tctp* (five replicates) (Gatz, 2017). Taken together, the independent approaches indicate that RNAPII is not recruited to the break if the break occurs at an intronless gene. The missing recruitment finally results in the absence of an siRNA response to a DSB for this intronless locus.

To dissect whether there is a RNA polymerase that is dedicated to DSB-triggered antisense transcription or whether the respective locus-associated polymerase fulfills this function, we decided to introduce a DSB into the *7SK* locus. The *7SK* small non coding RNA is transcribed by RNA polymerase III and we know from a previous study that a DSB in the *7SK* gene triggers an siRNA response (unpublished data). Performing the NET-Seq experiment, we could detect RNAs mapping to the antisense strand, specifically due to the DSB. However, none of them were longer than 21 nt. This hints at a contamination with siRNAs rather than at specific recruitment of RNAPII to this RNAPIII transcribed locus.

Further optimization of the NET-Seq protocol should allow obtaining a comprehensive transcription pattern at break sites. Adapting the NET-Seq protocol for RNAPII and RNAPIII might give insight whether antisense transcription at a DSB recruits the polymerase that also mediates sense transcription at a given locus or whether there is one polymerase dedicated to DSB-induced antisense transcription.

Recent reports also show that there is a general drop in transcription at break sites and this transient transcriptional silencing depends on DNA damage signaling by ATR (Iannelli et al., 2017; Kruhlak et al., 2007; Pankotai et al., 2012; Shanbhag et al., 2010). Using NET-Seq after performing knock-down of the signaling kinase orthologues in *Drosophila*, Tefu and Mei-41, might give insight whether DSB-induced antisense transcription also shows a dependence on the DNA damage response. Similarly, once a robust NET-Seq protocol is established, it would be possible to dissect which factors are essential for early biogenesis steps. A link between siRNAs and splicing was described (Merk et al., 2017). Knock-down of splicing factors would allow gaining mechanistic insight into the interplay between splicing and antisense transcription. Finally, it would be feasible to harvest cells at different time points, to identify times of maximal recruitment of RNAPII to the break.

With the established protocol of strand-specific RT-qPCR there is now also an independent approach to verify results obtained from NET-Seq experiments (Gatz, 2017).

### **IV.3 Homology directed repair and the general DSB stress response are unaffected by the absence of DSB-triggered siRNAs**

In the present study, we also investigated the biological role of break-induced siRNAs. The occurrence of DSB-triggered small RNAs is conserved and has been observed for the fungus *Neurospora crassa*, the plant *Arabidopsis thaliana*, the fly *Drosophila melanogaster*, and mammalian cultured cells.

A current hypothesis in the field is that break-derived small RNAs mediate faithful break repair via different pathways of homologous recombination. For *N. crassa* there is an evident hypersensitivity for DNA damaging conditions for deletion mutants of RNAi factors (Lee et al., 2009). In addition to this, there is a clear genetic inter-dependence between HR and siRNAs that are generated from repetitive loci in *N. crassa* (Yang et al., 2015). In apparent accordance with these findings, studies in *A. thaliana* and mammalian cultured cells also suggested an involvement of break-derived siRNAs in HR repair (Gao et al., 2014; Wei et al., 2012). However, for the latter two model organisms there are potential caveats. In *A. thaliana* the multitude of small RNA species and small RNA biogenesis factors, 4 Dicer-like proteins and 10 Argonaute proteins, considerably complicates the dissection of the influence of DSB-triggered siRNAs on DSB repair. The study by Wei *et al.* indicates that biogenesis of break-derived siRNAs might be preferentially generated by the action of DCL-3, however a reduced SSA efficiency was also reported for DCL-2 and DCL-4 (Wei et al., 2012). This suggests functional redundancy. Indeed, none of the Dicer-like proteins exclusively generate only a single class of small RNAs

---

(Gascioli et al., 2005). It is therefore very likely that the biogenesis of other small RNA species was affected in the DCL deletion mutants, and this in turn might result in a myriad of possible indirect effects.

In the present study, we made use of a GFP-based SSA reporter in order to relate to the study conducted in *A. thaliana* (Gao et al., 2014; Wei et al., 2012). Similar to the GUS reporter (see II.4.1), the break is induced by the highly specific endonuclease I-SceI. As for the GUS reporter, the uncut construct does not generate a signal. Only upon successful SSA repair an increase in GFP fluorescence is measurable (see also II.4.1). The SSA reporter was incorporated both in *Drosophila* cultured cells and in transgenic flies. Introduction of a single DSB by I-SceI led to an siRNA response against the reporter, which is very similar to previous observations using a linearized plasmid (Michalik et al., 2012).

In contrast to *A. thaliana* and the mammalian system, *Drosophila* cells have designated small RNA biogenesis factors. Dcr-1 and Ago1 are dedicated to the biogenesis of microRNAs (Förstemann et al., 2007; Lee et al., 2004; Tomari et al., 2007) and mutants of either factor are not viable. Dcr-2 and Ago2 on the other hand mediate siRNA biogenesis (Förstemann et al., 2007; Lee et al., 2004; Tomari et al., 2007). While flies with mutations in *dcr-2* and *ago2* are viable (see Fig. 31 and Table 8), there is a phenotype. Males with *dcr-2* and *ago2* mutant alleles show impaired fertility, which is caused by the reduced expression of hairpin RNAs. Furthermore, there are reports on zygotic defects in chromatin structure and mitosis for *ago2* mutant embryos (Deshpande et al., 2005). Despite these minor caveats, the model organism *Drosophila melanogaster* provides the unique opportunity to dissect the biological role of DSB-induced siRNAs and DSB repair via homology directed repair with minimal side-effects on the miRNA pathway.

When we used our SSA reporter to investigate the putative interplay between RNAi and HR-repair, we found that SSA was unaffected by the absence of siRNA factors in cultured cells. Accordingly, *dcr-2* mutant flies displayed functional and faithful SSA. In contrast to this, knock-down of miRNA factors in the SSA cell culture assay led to an apparent impairment of SSA repair. Likely explanations for this observation are indirect effects due to shifts in cell cycle (Fig. 27), effects on viability and reduction of transfection efficiencies (see Fig. 26 C).

Given that SSA is a special subtype of homology directed repair that uses a homology donor *in cis* and is independent from Rad51, we also probed whether we could detect a link between DSB-triggered siRNAs and HR with a donor *in trans*. Such a link was suggested for mammalian cultured cells (Wei et al., 2012). In the mammalian studies a GFP based reporter for short-tract gene conversion was used. Knock-down of DICER and AGO2 was found to reduce HR efficiencies. However, it should be noted that small RNA biogenesis pathways in mammalian cells are intertwined. The only Dicer protein processes both, miRNA and siRNA. Accordingly,

*Dicer* mutant mice are not viable. Although there four Argonaute proteins in the mammalian system, only AGO2 shows catalytic activity and is the preferential acceptor for both small RNA species. It is thus very likely that the effects on HR repair observed by Wei *et al.* can be at least partly explained by indirect effects. It is well established that microRNAs regulate the expression of DNA repair factors (Di Francesco *et al.*, 2013; Martin *et al.*, 2013; Yan *et al.*, 2010) and influence cell cycle distribution (Fig. 27 and (Yang and Qi, 2015)), in turn DSB repair pathway decision is dependent on the respective cell cycle stage. Indeed, a recent publication showed that overexpression of miRNAs of the let-7 family led to rescue of HR deficiency that was observed upon a DICER knock-down (Liu *et al.*, 2015).

To test whether Rad51-dependent HDR would be affected by the absence of siRNAs, we exploited our approach to tag proteins at their endogenous locus (Böttcher *et al.*, 2014; Kunzelmann *et al.*, 2016). To enable readout via flow cytometry, we knocked-in a GFP moiety at the *act5C* locus. However, we observed that the dependence on Rad51 itself strongly relies on the topology of the HR donor and the length of the homology stretches provided by the HR donor (see Fig. 34). The dependence on Rad51 was particularly strong when the HR donor was circular and had long stretches of homology. However, depletion of siRNA factors did not lead to a decrease in Rad51-dependent homology directed repair. This underlines the results obtained from our *in vitro* and *in vivo* SSA reporter assay.

To investigate the biological role of break-derived siRNAs we first set up GFP based reporters to re-evaluate experiments that were performed in *A. thaliana* and mammalian cells. However, in contrast to the mentioned studies, we were not able to find any link between faithful HR repair and DSB-triggered siRNAs in *Drosophila*. For both, the SSA reporter and the GFP knock-in approach, we introduced a single DSB at a defined position. We finally wanted to probe whether RNAi factors play a role in a general DSB stress response. We chose to assess this question by treatment with camptothecin, an inhibitor of topoisomerase I. This chemical agent leads to the generation of multiple, undirected DSBs throughout the genome.

We tested the response to the DSB stress *in vivo* in the presence or absence of the RNAi factors Dcr-2 and Ago2. Experiment readout was complicated by a background mutation in the *cyp6d2* gene that was present in several of the strains used in this study and which we could detect using a previously published PCR based assay (Thomas *et al.*, 2013). The mutated splicing defective or no transcript allele caused hypersensitivity to CPT (see Table 8 and Fig. 36) in the homozygous or trans-heterozygous combination. We could show that said hypersensitivity to CPT was not caused by mutations in the *dcr-2* gene: First, we could not rescue hypersensitivity to CPT by a wild-type, ectopically expressed *dcr-2* transgene (Table 8). Second, crosses with fly strains that carry wild-type *cyp6d2* did not show hypersensitivity. This was the case for crosses of

---

the RNAi defective *dcr-2*<sup>f06544</sup> allele. In addition, trans-heterozygous *ago2* mutants did not show impaired development into adult flies under CPT treatment.

In summary, our investigation of the putative interplay between break-derived siRNAs and HR repair showed that various subtypes of homology directed repair as well as the general DSB stress response are unaffected by the absence of break-induced siRNAs. Our findings are further substantiated by the observation that transcription is a prerequisite to an siRNA response to a break (Michalik et al., 2012). If DSB-triggered siRNAs played an essential role in homologous recombination repair, this would implicate that untranscribed loci or intergenic loci are not repaired via HR. This would be detrimental to the cells and a major cause for genomic instability. Furthermore, the phenotypes of *dcr-2* and *ago2* mutant flies are rather mild. Essential factors in HR repair such as *spnA* (encoding the Rad51 protein) or *okra* (encoding the Rad54 protein) display severe phenotypes when mutated, such as lethality or sterility due to inability to resolve meiotic crossovers. If the postulated crucial role in HR repair was true, then mutants of *dcr-2* or *ago2* should phenocopy DSB repair mutants.

Since readout of all the conducted assays was performed days after induction of the DSB, we cannot exclude that there is a redundant backup mechanism in the absence of siRNA factors and DSB-induced siRNAs to facilitate or mediate faithful HDR. Also it should be noted that although depletion of siRNA factors causes most likely no perturbation in microRNA biogenesis, there is the class of hairpin-derived endogenous siRNAs that are processed in a Dcr-2 dependent manner. *mus308*, encoding the MMEJ factor DNA polymerase theta may be a target for repression by the hairpin CG4068 (Babiarz et al., 2008; Okamura et al., 2008). Attempts to tag *mus308* at its endogenous locus to measure protein abundance in presence or absence of *dcr-2* were not successful. However, when we tested knock-down of various repair factors we observed for both, the SSA assay and the GFP-knock in assay, that depletion of Lig4 had a stronger influence on repair than depletion of Mus308 (Fig. 26 and Fig. 34).

Also, we did not investigate whether there is a change in end-joining repair pathway choice. The clear focus of the present study lies in reviewing whether the putative link between break-derived siRNAs and HR repair is conserved.

The results from this thesis, together with the previous observation that an siRNA response only occurs within transcribed units, led us to speculate that break-derived siRNAs rather have a function in RNA surveillance. Aberrant transcripts from the damaged gene might pose a threat to genome integrity or cellular physiology. siRNAs directed against such transcripts would thus act to render those transcripts harmless. Another possible explanation is that break-derived siRNAs are part of a self vs. non-self recognition mechanism to protect cells against transposition from transposons, during which DSB are introduced. This hypothesis is

substantiated by the observation that break-induced qiRNAs in *N. crassa* are directed against repetitive DNA (Yang et al., 2015). A break is even a prerequisite to the biogenesis of qiRNAs that target transposons (Zhang et al., 2013).

Once a robust, optimized NET-Seq protocol is established, it should be feasible to also investigate whether a break at a transcribed locus will result in run-off transcription of RNAPII or rather to its stalling. The latter scenario would display similarity to the identification of transposons in *Cryptococcus neoformans* (Dumesic et al., 2013), where stalled spliceosomes trigger transposon defense.

## V Materials and Methods

### V.1 Materials

#### V.1.1 Generation of sgRNA and sgRNA expression templates

For C-terminal tagging, cells that stably express Cas9 (cell line 5-3) were transfected with an *in vitro* transcribed sgRNA (for tagging of *RpII33*) or a sgRNA expression template (for all other loci modified in this study) to introduce a DSB adjacent to the stop codon. sgRNA expression templates for *in vitro* (IVT) or *in vivo* transcription were generated by an overlap PCR using the oligos indicated in Table 9 (for *RpII33*) or Table 10 (for all other loci).

**Table 9: Oligos used for generation of the sgRNA directed against the C-terminus of RpII33.** IVT was performed as described in (Michalik et al., 2012) and the *in vitro* transcribed sgRNAs were purified using a PCR purification kit (Qiagen).

Oligo #	Name	Sequence (5' -> 3')	notes
266	CRISPR_RPII33	taatagcactactattACTTGT TATATGTAAATCAAAG tttagagctag	Oligo contains the T7 promoter sequence to allow for IVT followed by a 19 nt long region complementary to the RpII33 (upper case letters) and an overhang that shows complementarity to oligo # 12
12	cas9_gRNA_scaffold	GTTTTAGAGCTAGAA ATAGCAAGTTAAAAT AAGGCTAGTCCGTTA TCAACTTGAAAAAGT GGCACCGAGTCGGTGC C	Serves as template for sense primer # 226 and the targeting sequence and the antisense primer # 11 only used as gRNA scaffold for tagging of <i>RpII33</i>
11	cas9_gRNA_ivt_as	GCACCGACTCGGTGC CACT	binds antisense to the gRNA scaffold # 12

**Table 10: Oligos used for overlap-extension PCR to generate sgRNA templates that are transcribed *in vivo* by RNAPIII.** For all PCRs we used the oligo # 252 as sense primer that allows for amplification of the U6Ac promoter from the pRB17 donor plasmid. The target sequence specific CRISPR oligos show an overlap to the pRB17 template, as well as an overlap for the optimized gRNA scaffold # 336. To provide a termination signal for RNAPIII transcription #254 is used as primer antisense to # 336.

Oligo #	Name	Sequence (5' -> 3')	notes
Common reagents for overlap PCR			
252	U6promotor_sense	GCTCACCTGTGATTGCTCCTAC	To amplify the U6Ac promoter from the pRB17 template
336	optimized gRNA scaffold	GTTTtAAGAGCTAtgctgGAAAacagcaTAGCAAGTTtAAATAAAGGCTAGTCCGT TATCAACTTGAAAAAGTGGCACCG AGTCGGTGC	
254	as_scaffold_pol3ter	gcttatttcAAAAAAGCACCGACTCGGT GCCACT	binds antisense to the gRNA scaffold #336
CRISPR Oligos			
1086	CRISPR_Ct_RP2-215	cctattttcaatttaacgtcgTCGAGGAGAGCG AAGACTGgtttaagagctatgctg	
570	Act5C_CRISPR_19nt_opt_scaffold	cctattttcaatttaacgtcgACCGCAAGTGCTT CTAAGAggtttaagagctatgctg	
CRISPR Oligos to generate sgRNA expression templates for introduction of locus specific DSBs			
750	CRISPR_3'UTR	cctattttcaatttaacgtcgTCCAGTGTAGCTT CCCGTtgtttaagagctatgctg	sgRNA directed against the 3' UTR of <i>CG15098</i>
1173	Tctp_CRISPR_830	cctattttcaatttaacgtcgATATCTAATTTCTT TTTACgtttaagagctatgctg	sgRNA directed against the 3' UTR of <i>Tctp</i>
653	7SK_CRISPR_3p	cctattttcaatttaacgtcgGTTTCGCTGCAGCA AAAGAAgtttaagagctatgctg	sgRNA directed against the 7SK locus

### V.1.2 Generation of HR donor templates for C-terminal tagging

For C-terminal tagging we rely on HDR for integration of the tag along with a resistance cassette. This can be achieved by concomitant transfection of an sgRNA (template) and a HR donor for repair. The HR donor was generally provided as linear PCR product using the oligos indicated in Table 11 and the template plasmids listed in Table 12 for PCR.

**Table 11: Oligos used for generation of HR donor templates with short homology arms (60 nt).**

oligo #	Sequence of sense oligo (5' -> 3')	Antisense oligo#	Sequence of antisense oligo (5' -> 3')	Donor plasmid
Tagging of <i>RpII33</i> to generate <i>RpII33-Flag<sub>2</sub></i>				
267	GAGAACAT <sup>T</sup> TGTGGTCATG GGTGTGCAGGTGCTGAAG AACAAGCTGTCAAACCTGC AGACGCAACTCAGCCACGA GTCCCAAACGATGCATTG GCCGTT <sup>g</sup> gatcttccggatggctcgag	268	TGGGCAT <sup>T</sup> TAACATAGACA GCTTGTTC AAGGATTAGG CGGAACATCTTTTATAAAA CTTAAAATAAACTTGTAT ATGTAAAGAAGTTCCTAT TCTCTAGAAAGTATAGGA ACTTCCATATG	pMH4
Tagging of <i>RpII215</i> to generate <i>RpII33-Flag<sub>2</sub></i> , <i>RpII215-Strep<sub>2</sub></i> , Blasticidin <sup>R</sup>				
1087	TACAGTCCCACGAGTCCGG CCTACTCGCCCAGCAGTCC CACG <sup>T</sup> TCGAGGAGAGCGA AGAC <sup>g</sup> gatcttccggatggctcgag	1088	ACGGCTTGACCGCCCCGG AGATCGGGACGTGCTGG GGGAGCTACCCCCGTCCT CCCTTCC <sup>g</sup> aagttcctattctagaaa gtataggaacttccatg	pIW1
Tagging of <i>RpII215</i> to generate <i>RpII33-Flag<sub>2</sub></i> , <i>RpII215-Strep<sub>2</sub></i> , Puromycin <sup>R</sup>				
1087	See above	1088	See above	pSK24
Tagging of <i>act5C</i> with a short HR donor template (60 nt homology) to generate <i>act5C-GFP</i>				
365	TGGATCTCCAAGCAGGAGT ACGACGAGTCCGGCCCCTC CAT <sup>T</sup> TGTGCACCGCAAGTGC TTC <sup>g</sup> gatcttccggatggctcgag	366	CCTCCAGCAGAATCAAGA CCATCCCGATCCTGATCCT CT <sup>T</sup> TGCCAGACAAGCGAT CCTTCGAAGTTCCTATTCT CTAGAAAGTATAGGAACT TCCATATG	pMH3

**Table 12: Overview on plasmids that were used as templates for HR donor template generation.**

Name	Tag	resistance-cassette
pMH4	- Flag <sub>2</sub>	Blasticidin-Resistance
pIW1	- Strep <sub>2</sub>	Blasticidin-Resistance
pSK24	- Strep <sub>2</sub>	Puromycin-Resistance
pMH3	- GFP	Blasticidin-Resistance

All listed template plasmids have been previously published (Böttcher et al., 2014; Kunzelmann et al., 2016)

The *act5C* locus was tagged with a GFP moiety either by a short linear PCR product (see Table 11), harboring 60 nt homology stretches upstream and downstream of the break or with a HR donor that carried longer homology stretches (500 nt/420 nt). Comparison of

different homology lengths and topology (linear and circular) allowed for investigation of requirements for Rad51-dependent HDR repair. We circularized the HR template harboring the short homology (60 nt) regions (Table 11). To this end 1.5 µg HR donor template was treated with 37.5 U T4 polynucleotidekinase (ThermoFischer Scientific) in a reaction mixture supplemented with 1x T4 DNA Ligase buffer (NEB) in a total reaction volume of 75 µl. The sample was incubated for 20 min at 37 °C, and the enzyme inactivated for 30 min at 65 °C. T4 DNA ligase (NEB) was added to a final concentration of 100 U/µl. Ligation was performed overnight at 18 °C. Circularization was confirmed on a 1 % agarose gel.

We also cloned the plasmid pIS1, where the GFP moiety is flanked by longer homology stretches (500 nt/420 nt). For the construction of pIS1, we PCR amplified homology regions of a length of 500 nt upstream of the *act5C* stop codon using primers # 982 and # 983 (see Table 13). We designed primers # 984 and # 985 to amplify 500 nt downstream of the stop codon, however, due to a natural occurring *HindIII* site in the homologous sequence, the downstream homology stretch has a length of 420 nt. Molecular cloning was performed in a sequential manner, we first used the *NoI/XhoI* (upstream) restriction site for integration of the upstream homology arm into pMH3 and confirmed successful integration by sequencing. In the next step we used the *Clal/HindIII* restriction site of the plasmid for introduction of the downstream homology region to yield pIS1.

**Table 13: Overview of oligos designed for construction of pIS1.**

Oligo #	name	Sequence (5'->3')
982	Act5c_upstream_ <i>NoI</i> _sense	AAAgcggcgcTCCGTGACATCAAGGAGAAG
983	Act5c_upstream_ <i>XhoI</i> _antisense	TTTctcgagCCATCCGGAAGATCCGAAGCACTTGCGGT GCACAA
984	Act5c_downstream_ <i>Clal</i> _sense	AAAatcgatGAAGGATCGCTTGTCTGGGC
985	Act5c_downstream_ <i>HindIII</i> _antisense	TTTaaagcttCTTTTTCGCGCTTGGGAAAT

To investigate the role of HR donor topology for Rad51 dependence of HDR repair we also linearized pIS1 using *HindIII* and *NoI*.

### V.1.3 dsRNA constructs

Sequences for generation of dsRNA constructs that are not listed in Table 14 have been previously described in (Hartig et al., 2009; Hartig and Förstemann, 2011; Michalik et al., 2012).

**Table 14: Sequences for generation of dsRNA.**

	Sense primer with T7 extension (sequence 5'->3')	Antisense primer with T7 extension (sequence 5'->3')
dsRNA constructs for C-terminal tagging		
Lig 4	taatacgactcactatagggCCCAATGATCCAAAGTGT TTTG	taatacgactcactatagGGAAGTAGGATGCCTTCGC
Mus308	taatacgactcactatagGCTGGGACTCCACCGAA	taatacgactcactatagggTACCGTCGCCGTCCAGTAA
dsRNA constructs for assessment of HDR		
Fluc	taatacgactcactatagggACAGCTCTGGGTCTACCG G	taatacgactcactatagggAGGGTCTTGCCGGTGTC
Rad51	taatacgactcactatagggGCTCCAAGGAGCTGGATA AA	taatacgactcactatagggGGCTTCTTGGCATCAAACA T
Rad51 Design B	taatacgactcactatagggCATTCCCGGCTTGGGCGG	taatacgactcactatagggTCCAGCACCTCGGATTCAT TC
Rad51 Design C	taatacgactcactatagggCAATGTGGCCTTCACCCGT G	taatacgactcactatagggAGCTCTCCCTGGCGTCTCC
mus308	taatacgactcactatagggTACATATACGGCAACGGG TC	taatacgactcactatagggCTGCAGTTTAACGACTCG GC
Nbs1	taatacgactcactatagggATTGGCATGGCCCTAATAC A	taatacgactcactatagggCCTGTTTGGTCTTCCGTGT T
RPA	taatacgactcactatagggAACGCGTTCTCATCATTTTC C	taatacgactcactatagggCATCTCGTAGGCGTTGTTC A
AGO2 Design B	taatacgactcactatagggGAGGGGCTCTGGAGGAG ATT	taatacgactcactatagggCGTCCACGAGCAGCTACC A

#### V.1.4 Cell lines used in this study

**Table 15: Overview on cell lines that were used in this study.**

Name used in this study	Description	Resistance marker
S2	wild-type <i>Drosophila</i> cells	-
5-3	S2 cells stably expressing Cas9, described in (Böttcher et al., 2014; Kunzelmann et al., 2016)	hygromycin
RpII33-Flag <sub>2</sub> population	5-3 cells were tagged with a tandem Flag tag at the C-terminus of the RNAPII subunit <i>RpII33</i>	hygromycin blastocidin
RpII33-Flag <sub>2</sub> clone 8 and 12	RpII33-Flag <sub>2</sub> cells were subjected to single cell cloning, we retrieved clones 8 and 12	hygromycin blastocidin
RpII33-Flag <sub>2</sub> clone 8B6/clone 12D4	RpII33-Flag <sub>2</sub> clones 8 and 12 were subjected to transient FLP recombinase expression (pMH5) to remove the blastocidin resistance cassette and restore the 3' UTR,	hygromycin

	from subsequent single cell cloning we could retrieve clone 8B6, clone 12D4	
RpII33-Flag <sub>2</sub> RpII215-Strep <sub>2</sub> population	RpII33-Flag <sub>2</sub> clone 8B6 was used as parental cell line for tagging of <i>RpII215</i>	hygromycin blasticidin
RpII33-Flag <sub>2</sub> RpII215-Strep <sub>2</sub> population	RpII33-Flag <sub>2</sub> clone 8B6 was used as parental cell line for tagging of <i>RpII215</i>	hygromycin puromycin
RpII215-V5 clone 16	5-3 cells were tagged with a tandem Flag tag at the C-terminus of the RNAPII subunit RpII215 Clonal cell line was provided by Volker Nitschko	hygromycin blasticidin
2x3	SSA reporter cells, detailed in (Böttcher et al., 2014)	
2x3 $\Delta$ lig4	SSA reporter cells, for which a knockout of ligase 4 was performed by CRISPR/ <i>cas9</i> , detailed in (Schmidts et al., 2016)	

## V.2 Methods

### V.2.1 Cell culture

#### V.2.1.1 General culturing conditions and transfection

Cells were cultured in Schneider's medium (Bio&Sell) that was supplemented with 10 % fetal bovine serum (Bio&Sell) and 1x Pen/Strep.

Selection of successfully tagged cells was performed by addition of either Blasticidin-S (Life Technologies) (final concentration 10-25  $\mu$ g/ml) or Puromycin (Life Technologies) (final concentration 0.5  $\mu$ g/ml) to the medium. For all transfections, Fugene HD (Promega) was used as a transfection reagent according the manufacturer's instructions.

#### V.2.1.2 SSA assay

The SSA reporter cell line 2x3 is described in (Böttcher et al., 2014). It served as parental cell line for generation of the lig4 knockout cell line 2x3  $\Delta$ *lig4*, as described in (Schmidts et al., 2016).

Cells were seeded at a concentration of  $1 \times 10^6$  cells/ml and treated with 0.25 ng/ $\mu$ l dsRNA to perform knock-down of individual factors. Cells were split 1:3 four days after RNAi and transfected with pKF259, the expression vector of the I-*SceI* nuclease, at a final concentration of 0.5 ng/ $\mu$ l. In order to assess the influence of individual dsRNA constructs on transfection efficiencies, we treated wildtype S2 cells as described above and transfected them with pKF63 (Fürstemann et al., 2007) encoding GFP under control of the ubiquitin promoter. 5 days post

transfection flow cytometry was performed (FACSCalibur equipped with a 96 well plate autosampler, Becton Dickinson).

### V.2.1.3 C-terminal tagging

#### V.2.1.3.1 C-terminal tagging of RNAPII subunits

Tagging of the C-terminus of *RpII33* with a Flag<sub>2</sub>-Tag via the CRISPR/*cas9* tagging protocol was performed as previously described (Böttcher et al., 2014). 5-3 cells were seeded with a concentration of  $1 \times 10^6$  cells/ml and depleted for Lig4 by knock-down. After recovery, the cells were seeded again with  $1 \times 10^6$  cells/ml and co-transfected with the *in vitro* transcribed sgRNA (Table 9) and a linear PCR product, which served as HR-donor for repair (Table 10). After several rounds of selection in medium supplemented with blasticidin, a single cell dilution was performed to retrieve clonal RpII33-Flag<sub>2</sub> cell lines. Clonal cell lines RpII33-Flag<sub>2</sub> clone 8 and clone 12 were transiently transfected with pMH5 (Böttcher et al., 2014), an expression vector for the FLP recombinase. Again, we subjected cells to a single cell dilution and retrieved the clonal cell lines RpII33-Flag<sub>2</sub> clone 8B6 and 12D4. Removal of the blasticidin resistance cassette restored the endogenous 3' UTR in those two cell lines.

For tagging of *RpII215* the clonal cell line RpII33-Flag<sub>2</sub> clone 8B6 was used as parental cell line. Tagging of cells with a Strep<sub>2</sub>-tag was performed as described in (Kunzelmann et al., 2016). Cells were seeded at  $1 \times 10^6$  cells/ml in a 24 well plate format (500  $\mu$ l/well) and depleted for Mus308 and Lig4 by knock-down by adding each dsRNA to a final concentration of 1  $\mu$ g/ml. After recovery from the knock-down, we again seeded the cells in a 24 well plate format with a concentration of  $1 \times 10^6$  cells/ml and transfected them with 375 ng sgRNA expression template and 375 ng HR donor template. To enrich for cells that had undergone integration of the Strep<sub>2</sub> tag and the respective selection cassette, we subjected cells to selection using the respective selection medium. We performed single cell dilution to obtain clonal RpII33-Flag<sub>2</sub> RpII215-Strep<sub>2</sub> cell lines.

#### V.2.1.3.2 C-terminal tagging of *act5C* for the HDR assay

5-3 cells were seeded at a concentration of  $0.5-1 \times 10^6$  cells/ml in the wells of a 96 well plate (volume 100  $\mu$ l) and treated with 250 ng/well dsRNA constructs for knock-down of individual factors. After 48 h, cells were co-transfected with the sgRNA- and HR donor template, both added to a final concentration of 0.75 ng/ $\mu$ l. 5 days post transfection the percentage of GFP

positive and negative cells was assessed using flow cytometry (FACSCalibur equipped with a 96 well plate autosampler, Becton Dickinson).

#### **V.2.1.4 Introduction of locus-specific DSBs exploiting the CRISPR/*cas9* system for analysis of the antisense transcription at a break via NET-Seq**

sgRNA expression templates directed against the *7SK* locus or the 3' UTR of *CG15098/Tctp* were generated as described above (V.1.1) using the CRISPR oligos indicated in Table 10.

30 ml RpII33-Flag<sub>2</sub> RpII215-Strep<sub>2</sub> cells were seeded with a cell density of 1.5x10<sup>6</sup> cells/ml into 30 cm dishes (Greiner Bio-One). A total of 3x30 cm dishes was prepared for each condition. Per 30 cm dish 15 µg of the respective sgRNA expression template was transfected in a total volume of 6 ml. The transfection mix for each dish included 180 µl FuGene (Promega). Untransfected cells served as a control. Cells were harvested 48 h post transfection and were either used for the T7-Endonuclease Assay to confirm introduction of DSBs (see V.2.2.1) or for NET-Seq (using the optimized lysis protocol described in V.2.2.2.4, the immunoprecipitation and elution protocol as detailed in V.2.2.3/V.2.2.4 and the small RNA library preparation protocol described in V.2.2.8).

### **V.2.2 *In vitro* experimental procedures**

#### **V.2.2.1 T7-Endonuclease Assay**

48 h after transfection of RpII33-Flag<sub>2</sub> RpII215-Strep<sub>2</sub> cells with sgRNA expression templates, 1.5 ml cells were pelleted and resuspended in 160 µl 1x PBS. Subsequently, gDNA was extracted using the ReliaPrep gDNA extraction kit (Promega) according to manufacturer's instructions. PCR amplification of the specific locus was performed using the following reaction mix with the primers listed in Table 16.

gDNA template	2 µl
10 x PCR buffer (+KCl)	2.5 µl
dNTP-Mix (10 mM Mix)	0.5 µl
MgCl <sub>2</sub> (25 mM)	2.5 µl
Primer sense (10 µM)	0.5 µl
Primer antisense (10 µM)	0.5 µl
Taq (laboratory stock)	0.5 µl

ad 25  $\mu$ l with nuclease free water

**Table 16: Primers used for amplification of the indicated loci for a T7 endonuclease assay.**

Locus	Primer sense	Primer sense sequence	Primer antisense	Primer antisense sequence
<i>CG15098</i>	# 957	GAACGCCATTACTC CTGCT	# 1223	TCTTGTATTGGCCATTTCGT TTG
<i>Tcp</i>	# 752	CTGGCGTGGATGTT GTGCTT	# 1172	CATATGGCGGCGTCTACCTA CG
<i>7SK</i>	# 1288	ACCCTCCGTCACACC TTTG	# 1316	TTGGTGGACCTAATAGCTGA

To generate PCR amplicons the following thermocycling parameters were set on a thermocycler (Sensoquest)

Initialization	95 °C	3 min	
Denaturation	95 °C	30 sec	35 x
Annealing	55 °C	30 sec	
Elongation	72 °C	1 min	
Final Elongation	72 °C	5 min	

Following the PCR, amplicons were denatured at 95 °C for 5 min and annealed by setting a temperature decrease from 95 °C-85 °C with a rate of -2 °C per second and a subsequent temperature drop from 85 °C-25 °C with a rate of - 0.1 °C per second. Amplicons were finally digested by mixing the T7 endonuclease reaction mix (2  $\mu$ l NEB buffer 2, 0.5  $\mu$ l T7-Endonuclease (10 U/ $\mu$ l) (NEB), 7.5  $\mu$ l nuclease free water) and 10  $\mu$ l PCR product. Digestion was performed by incubating samples for 16 min at 37 °C. Digested samples were then directly loaded on a 1 % agarose gel.

## V.2.2.2 Cell lysis

### V.2.2.2.1 Cytoplasmic lysis protocol

The cell pellet was resuspended in the lysis buffer (30 mM HEPES pH7.5, 100 mM KoAc, 2 mM Mg(OAc)<sub>2</sub>, 1 mM DTT, 1x Proteinase Inhibitor without EDTA (Roche)) and subjected to a freeze-thaw cycle (snap-freezing in liquid nitrogen, incubation at room-temperature until thawing is almost complete).

Subsequently, the extract was incubated on ice for 20 min and centrifuged at 4 °C, 13200 rpm for 10 minutes. The supernatant contains soluble cytoplasmic proteins.

#### **V.2.2.2.2 Cellular fractionation protocol (modified from Abcam)**

The cell pellet was resuspended in a hypotonic buffer (10 mM HEPES pH 7.5, 1.5 mM MgCl<sub>2</sub>, 10 mM KCl, 1 mM DTT, 1 % Triton-X-100 or 0.3 % Tergitol-type NP-40, 1x Proteinase Inhibitor without EDTA (Roche)). Resuspended cells were then subjected to a freeze-thaw cycle (as described above) or directly incubated on ice for 20 min (according to Abcam's protocol). After centrifugation (3000 rpm, 4 °C, 20 min) the supernatant, which corresponds to the cytosolic extract was saved for further analysis. The pellet, which is comprised of nuclei, was resuspended in a high salt buffer (5 mM HEPES, pH 7.5, 1.5 mM MgCl<sub>2</sub>, 1 mM DTT, 10 % Glycerol, 1x Proteinase Inhibitor without EDTA (Roche)). NaCl was added to a final concentration of 300 mM. To facilitate disruption of the nuclear membrane, the extract was further homogenized by 20 full strokes in a Dounce homogenizer on ice. Samples were incubated for an additional 20 min on ice and then centrifuged at 13200 rpm, 4 °C for 20 min.

Optionally, a KCl chromatin washing step was performed. To this end, nuclear extraction pellets were resuspended in chromatin-extraction buffer (20 mM HEPES pH 7.5, 3 mM MgCl<sub>2</sub>, 1x Proteinase Inhibitor without EDTA (Roche)). Finally, an equal volume of high KCl chromatin-extraction buffer (0.6 M KCl, 20 mM HEPES pH 7.5, 3 mM MgCl<sub>2</sub>, 1x Proteinase Inhibitor without EDTA (Roche), 10 % glycerol) was added in a dropwise manner. Extracts were then centrifuged (15 min, 4 °C, 13200 rpm).

#### **V.2.2.2.3 Combined fractionation and sonication protocol**

The cell number was determined when harvesting cells. For lysis, cells were resuspended in a fixed volume to reach a cell density of  $1 \times 10^8$  cells/ml = 1 V.

First, the cell pellet was resuspended in a 0.5 V hypotonic buffer (10 mM HEPES pH 7.5, 1.5 mM MgCl<sub>2</sub>, 10 mM KCl, 1 mM DTT, 1 % Triton-X-100 or 0.3 % Tergitol-type NP-40, 1x Proteinase Inhibitor without EDTA (Roche)). Following a freeze-thaw cycle, we added 0.5 V equal volume of hypertonic buffer (10 mM HEPES pH 7.5, 1.5 mM MgCl<sub>2</sub>, 0.6 M KCl, 0.3 % Tergitol-type NP-40, 1 mM DTT, 10 % Glycerol, 1x Proteinase Inhibitor without EDTA (Roche), 1x PhosStop (Roche)) in a dropwise manner to the hypotonic extract. Subsequently, samples were incubated for 20 min on ice (vortexing every 5 minutes). Extracts were sonicated

using Diagenode's Bioruptor with the following instrument settings: 12 times 20 sec on, 60 sec off, high. Adjusting the lysis volume to the number of cells (see above) was thought to ensure efficient shearing of the chromatin. Finally, samples were centrifuged at 13200 rpm, 4 °C for 15 min.

#### **V.2.2.2.4 Optimized lysis protocol**

Note that it is unnecessary to adjust the lysis number to a fixed cell density. Usually, we resuspended a pellet from 10-30 ml cell culture in a lysis volume of 500 µl.

For the optimized lysis protocol, cell pellets were resuspended in the optimized lysis buffer (10 mM HEPES pH 7.5, 0.4 M KCl, 1.5 mM MgCl<sub>2</sub>, 1 % Tergitol-type NP-40, 1 mM DTT, 10 % glycerol, 1x Proteinase Inhibitor without EDTA (Roche), 1x PhosStop (Roche)) and subjected to a single freeze-thaw cycle. After incubation of the samples on ice for 20 min with occasional vortexing (~every 5 minutes), extracts were sonicated using Diagenode's Bioruptor with the following instrument settings: 12 times 20 sec on, 60 sec off, high. A final centrifugation step was performed (13200 rpm, 15 min, 4 °C).

#### **V.2.2.2.5 Cell lysis by cryogenic grinding**

~50 ml dry ice was powdered using a coffee grinder (Gastroback). This ensured that the coffee grinder was pre-cooled for lysis and enabled us to snap-freeze cells (see below). Cell pellets were resuspended in the optimized lysis buffer (100 mM HEPES pH 7.5, 0.4 M KCl, 1.5 mM MgCl<sub>2</sub>, 1 % Tergitol-type NP-40, 1 mM DTT, 10 % glycerol, 1x Proteinase Inhibitor without EDTA (Roche), 1x PhosStop (Roche)). Note that we increased the molarity of the buffer component (HEPES) from 10 mM to 100 mM to ensure that the pH will stay stable at 7.5 even in the presence of dry ice. Powdered dry ice was added to the extracts, until the samples were completely frozen. The frozen cells were then ground in the pre-cooled coffee grinder. Following this step, the powdered sample was transferred into a 50 ml Falcon tube and incubated at room temperature (under the hood) with open lid until thawing was completed. Samples were incubated on ice for 10 or 30 minutes and centrifuged for 15 min, 4 °C, 13200 rpm.

#### **V.2.2.3 Immunoprecipitation**

When working with RpII33-Flag<sub>2</sub> clones, we pre-incubated magnetic Protein G beads (Dynabeads, ThermoFischer Scientific) with αFlag antibody (M2, Sigma) for 1 h to overnight.

Accordingly, we pre-incubated magnetic Protein G beads (Dynabeads, ThermoFischer Scientific) with  $\alpha$ V5 (BioRad), when performing the IP with extracts from the cell line RpII215-V5 clone 16.

For Strep-IP we used magnetic Strep-Tactin beads (IBA), so that a pre-incubation with a respective antibody was not necessary. In general, 20  $\mu$ l beads were used per 0.5-1 ml of extract, and if necessary, pre-incubated with 2  $\mu$ l of the respective antibody.

(Pre-incubated) beads were then washed using the respective lysis buffer and resuspended in the lysis extracts. For Strep-IPs we added avidin (100  $\mu$ g). After 1-2 hours rolling at 4 °C, the supernatant was removed and beads were washed with IP buffer (25 mM HEPES pH 7.5, 150 mM NaCl, 12.5 mM MgCl<sub>2</sub>, 10 % glycerol, 0.3 mM Tergitol-type NP-40, 1 mM DTT, 1x Proteinase Inhibitor without EDTA (Roche), 1x PhosStop (Roche)). The elution procedure differed depending on the respective IP.

#### **V.2.2.4 Elution**

For  $\alpha$ Flag IPs, we eluted RpII33-Flag<sub>2</sub> either by competition with the 3X flag peptide (Sigma) or by arginine elution. For flag elution, beads were resuspended in elution buffer (110 mM KCl, 10 mM HEPES pH 7.5, 1.5 mM MgCl<sub>2</sub>, 0.3 % Tergitol-type NP-40, 1 mM DTT, 1x PhosStop (Roche), 1x Proteinase Inhibitor without EDTA (Roche), varying concentrations of 3X flag peptide (Sigma) and incubated 30 min on ice or at room temperature.

For the arginine elution (Futatsumori-Sugai et al., 2009), beads were resuspended in a 0.75 M arginine solution (pH 3.5), incubated for 5 min at 37 °C and then neutralized with 0.14x Volume of Tris-HCl (pH 8,0) buffer.

For elution of tandem strep-tagged RpII215, we used 1x Biotin elution solution according to the manufacturer's instructions (IBA).

#### **V.2.2.5 Western blotting**

Cell extracts or samples from IPs (input, supernatant, beads, eluate) were mixed with SDS-loading buffer and applied to a 8 % or 10 % SDS-gel. For Western blotting, membranes were either probed with antibodies against the endogenous RNAPII subunit RpII215, by using the commercial monoclonal antibody 8WG16 (Covance) or the polyclonal antibody Pol3-3 (courtesy of the laboratory of Prof. Dr. Eick) or antibody against the tags that we introduced using our CRIPR/*cas9* approach:  $\alpha$ Flag M2 (Sigma),  $\alpha$ Strep-Tag-II-HRP (IBA) or  $\alpha$ V5-HRP (BioRad).

For assessment of Upf1 mediated decay of the uncut SSA reporter construct, we treated a total of  $3 \times 10^6$  cells with 3 ng/ $\mu$ l dsRNA for knock-down of Upf1, GFP (positive control) or RLuc (knock-down control). We isolated cytoplasmic proteins five days after the knock-down. To this end, cells were harvested, washed twice in 1x PBS and lysed in 1x PBS supplemented with 1 % Tween and 1x Proteinase Inhibitor without EDTA (Roche). A freeze-thaw cycle was conducted. Upon thawing the samples were incubated on ice for 30 min, then centrifuged (16000 x g, 4 °C, 15 min). The supernatant that contained soluble cytoplasmic proteins was mixed with SDS loading buffer. To allow for direct comparison equal amount of protein were loaded on a 12 % SDS-gel. All chemiluminescence images were recorded with a Fuji LAS-3000 mini system or the Amersham Imager 600 (GE healthcare).

#### **V.2.2.6 Recovery of co-immunoprecipitated RNA**

We added SDS to the eluates to a final concentration of 0.5 % or resuspended beads in Proteinase K buffer (0.4 M NaCl, 0.5 % SDS, 50 mM Tris-HCl pH 8.0). Next, we treated samples with 0.8 U Proteinase K (NEB) for 30 min at 56 °C under constant shaking (700 rpm) to release co-immunoprecipitated nucleic acids. RNA was retrieved by adding 3 Volumes of Trizol Reagent (Sigma) to the respective eluate. RNA was then purified using the DirectZol Kit (ZymoResearch) according to the manufacturer's instruction. RNA was used as input for small RNA library generation (see V.2.2.8) or RT-PCR (see V.2.2.7).

#### **V.2.2.7 NET-RT-PCR**

Cell lysis, immunoprecipitation and the respective elution was performed as described above, as was purification of co-immunoprecipitated RNA. RNA was then subjected to a DNase I digestion in 1x DNase I buffer (ThermoFischer Scientific) by adding 1 U of DNase I (ThermoFischer Scientific) to the purified RNA. To inhibit RNA degradation during the incubation time of 30 min at 37 °C we added RiboLock (ThermoFischer Scientific) to a final concentration of 4 U/ $\mu$ l. Subsequently, 0.8 U of Proteinase K (NEB) was added to the mixture and incubated at 65 °C for 15 min. RNA was then purified using the RNA clean and concentrator kit (Zymogen) according to the manufacturer's instructions. For synthesis of cDNA, reverse transcription was performed using the SuperScript III reverse transcriptase kit (Invitrogen) according to the manufacturer's instructions with random hexamer primers. The cDNA was used as a template for locus specific PCR of either RNAPII associated genes (*act5C*, *CG15098*) or the RNAPIII associated locus *7SK*.

### V.2.2.8 Generation of small RNA libraries from co-immunoprecipitated RNA

Small RNA library generation was basically performed as described in (Elmer et al., 2014). Size selection was adapted for NET-Seq library generation and the gel region corresponding to ~ 22-50 bp was excised from the urea-PAGE gel. Size-selected, gel-purified RNA was retrieved by using the ZR small RNA PAGE Recovery Kit (ZymoResearch) according to the manufacturer's instructions. Following the 3' end ligation RNA was again PAGE-purified, size selected (~30-100 bp) and retrieved by using the ZR small RNA PAGE Recovery Kit (ZymoResearch) according to the manufacturer's instructions. Prior to 5' ligation, the purified 3' ligated RNA was treated with T4 polynucleotide kinase (1x T4 RNA Ligase buffer (NEB), 5 Units T4 PNK (ThermoFischer Scientific) for 10 min at 37 °C. T4 PNK was heat-inactivated for 20 min at 65 °C. Sequencing of the libraries was performed on an Illumina HiSeq1500 instrument by LAFUGA (Gene Center, LMU Munich, Germany).

### V.2.2.9 Deep sequencing data analysis

Initial processing of raw reads was performed on the LAFUGA Galaxy Server, using the Galaxy Tools 'Illumina Demultiplex' (Galaxy Version 1.0.0), to assign individual reads to the respective index and 'Clip Adaptor Sequence' (Galaxy Version 1.0.0), to trim off the 3' adaptor sequence. The output file contains the sequences of incorporated RNAs ("clipped reads") and gives insight into the percentage of reads without inserts ("linker-to-linker"). Analysis of clipped reads by 'FASTQC' (Galaxy Version 0.52) was performed to obtain a size distribution profile of the clipped reads.

Clipped reads were then aligned to different precomputed sequence collections from Flybase using the short read aligner 'bowtie' (Langmead et al., 2009), allowing no mismatches. To filter out 2S rRNA reads we generated a reference sequence for the 2S rRNA using the bowtie-build function. For mapping to loci where a DSB was introduced, we also build a custom reference sequence for the *CG15098*, the *Tctp* and *7SK* locus. Downstream analysis was performed with customized perl scripts, e.g. the perl script 'map\_in\_interval'.

## V.2.3 Fly strains and husbandry

### V.2.3.1.1 Fly strains

Table 17 gives an overview on the fly strains used in this study.

**Table 17: Fly strains used in this study.**

genotype	Allele description
Fly strains used for assessment of SSA repair efficiencies	
$260-1; dcr-2^{L811fsX} / CyO$	Flies carry the SSA reporter transgene Heterozygous for the <i>dcr-2</i> null allele
$260-1; dcr-2^{G31R} / CyO$	Flies carry the SSA reporter transgene Heterozygous for the <i>dcr-2</i> point mutated allele
$260-1; dcr-2^{C473Y} / CyO$	Flies carry the SSA reporter transgene Heterozygous for the <i>dcr-2</i> point mutated allele
$259-4; dcr-2^{L811fsX} / CyO$	Flies carry the transgene encoding the I- <i>SceI</i> transgene Heterozygous for the <i>dcr-2</i> null allele
$259-4; dcr-2^{G31R} / CyO$	Flies carry the transgene encoding the I- <i>SceI</i> transgene Heterozygous for the <i>dcr-2</i> null allele
Fly strains additionally used for assessment of CPT sensitivity	
$y, w, eyFLP; FRT42Ddcr-2^{L811fsX};$ $P\{w^+, Dcr-2\}$	Homozygous for the <i>dcr-2</i> null allele Ectopically express <i>dcr-2</i> rescue transgene
$w^{1118}; PBac\{w^{+mC} = WH\}dcr-$ $2^{06544} / CyO$	PiggyBac insertion into the transcription start site Functional RNAi mutant
$y^l, w^{1118} / Dp(1;Y)y^+; Df(2R)ED3385,$ $P\{w^{+mW.Scer\FRT.hs3} =$ $3'.RS5+3.3'\}ED3385 / CyO$	Deficiency strain <i>dcr-2</i> lies within the deficiency region
$w^{1118}; AGO2^{321} / TM3, Sb^1$	Heterozygous for the <i>ago2</i> loss of function allele
$yw; ago2^{V966M}$	Homozygous for the <i>ago2</i> point mutant allele

The generation of transgenic flies that carry either the SSA reporter construct (260-1) or the I-*SceI* nuclease (259-4) on the X chromosome was published in (Schmidts et al., 2016). Using standard genetic approaches, transgenic flies were crossed with *dcr-2* deficient flies carrying the *dcr-2* null allele  $dcr-2^{L811fsX}$ , or one of the *dcr-2* point mutants  $dcr-2^{G31R}$  or  $dcr-2^{C473Y}$  (compare Table 17 for the complete genotype). The mentioned *dcr-2* mutant alleles are described in (Lee et al., 2004).

### V.2.3.1.2 *In vivo* SSA assay

For assessment of SSA efficiencies in the absence of *dcr-2*, we first generated trans-heterozygous flies that carry two different *dcr-2* mutant alleles as well as the SSA reporter. Trans-heterozygous

females were then crossed to males that carry the I-*Scel* transgene and are heterozygous for one of the *dcr-2* mutant alleles. Progeny of this cross was subjected to Western blotting or fluorescence microscopy.

#### **V.2.3.1.2.1 Fluorescence microscopy**

Fluorescence microscopy of fly eyes was conducted using a Leica MZ16FA fluorescence stereomicroscope, equipped with a Leica DFC 320 camera. For all images magnification was set to 99.8x. Acquisition was performed using the Leica DFC Twain camera software with the gain setting set to maximum and an exposure time of 2.9 s. In photoshop the filter ‘Sharpen’ was applied to all raw images. Fluorescence signal intensities within the fly eye was then quantified using ImageJ.

#### **V.2.3.1.2.2 Preparation of fly extracts for Western blotting**

To measure GFP protein levels in the whole fly body, protein extraction from individual flies was performed. Flies were first anesthetized and then homogenized in 75  $\mu$ l lysis buffer (30 mM HEPES, 100 mM KOAc, 2 mM MgOAc, 1 mM DTT, 1 % triton-X and 1x Proteinase Inhibitor without EDTA (Roche)). Extracts were incubated on ice for 30 min and subsequently centrifuged (16000 xg, 4 °C, 15 min). The supernatant was mixed with SDS-loading buffer. We loaded 24  $\mu$ g of total protein on a 12 % SDS-gel and probed the Western blot membrane with the monoclonal B-2 antibody (sc-9996, Santa Cruz) to measure GFP expression. For loading control we measured beta tubulin by probing the membrane with the monoclonal anti beta-Tubulin (E7) antibody from the Developmental Studies Hybridoma Bank (DSHB). All chemiluminescence images were recorded with a Fuji LAS-3000 mini system. Quantification of band intensity was performed using ImageJ.

#### **V.2.3.1.3 *In vivo* camptothecin hypersensitivity assay**

For assessment of CPT sensitivity in a corresponding wild-type, we first backcrossed a *dcr-2* mutant allele with *w<sup>1118</sup>*. All progeny was heterozygous and carried a wild-type chromosome over the balancer *CyO*. These F1 animals were subjected to the CPT assay as described below.

To investigate the role of *dcr-2* in the DSB stress response, we again used mutant strains that carry the SSA reporter transgene. Since the *dcr-2* mutant alleles *dcr-2*[G31R], *dcr-2*[L811fsX] and *dcr-2*[C473Y] were all derived from a mutagenesis screen, we used independent *dcr-2* mutant fly strains that either harbor a deficiency (*Df(2R)ED3385*) or a piggyback insertion into the *dcr-2*

gene (*dcr-2*<sup>06544</sup>) to exclude possible indirect effects from possible background mutations (compare Table 17 for complete genotypes). We set up homozygous or trans-heterozygous crosses to obtain *dcr-2* deficient progeny and corresponding *dcr-2* proficient control animals. Rescue experiments were performed by crossing *dcr-2* mutant strains to flies carrying the *dcr-2* rescue transgene (for complete genotype compare Table 17). In addition, we also assessed the DSB stress response in the absence of *ago2*. A trans-heterozygous cross was set up for strains carrying the *ago2* loss of function allele *AGO2*<sup>321</sup> and the *ago2* point mutant allele *ago2*<sup>V966M</sup> (see Table 17. for the complete genotype).

For all CPT experiments, fly food was coated with different concentrations of CPT (Sigma), the respective solvent control (DMSO) or untreated fly food. Prior to setting up the cross, F0 females were kept on fly food treated with CPT or the DMSO control during four days after eclosion. All subsequent crosses were performed on CPT/DMSO-treated food 7 days after setting up the cross, F0 flies were transferred to freshly CPT/DMSO-coated fly food. RNAi deficient and proficient progeny was counted. The results are listed in Table 17.

---

---

## References

- Abmayr, S.M., Yao, T., Parmely, T., and Workman, J.L.** (2001). Preparation of Nuclear and Cytoplasmic Extracts from Mammalian Cells. In *Current Protocols in Molecular Biology* (John Wiley & Sons, Inc.).
- Aguilera, A., and Gaillard, H.** (2014). Transcription and Recombination: When RNA Meets DNA. *Cold Spring Harbor Perspectives in Biology* 6, a016543.
- Anindya, R., Aygün, O., and Svejstrup, J.Q.** (2007). Damage-Induced Ubiquitylation of Human RNA Polymerase II by the Ubiquitin Ligase Nedd4, but Not Cockayne Syndrome Proteins or BRCA1. *Molecular Cell* 28, 386-397.
- Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P., and Blelloch, R.** (2008). Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev* 22, 2773-2785.
- Behm-Ansmant, I., Gatfield, D., Rehwinkel, J., Hilgers, V., and Izaurralde, E.** (2007). A conserved role for cytoplasmic poly(A)-binding protein 1 (PABPC1) in nonsense-mediated mRNA decay. *Embo J* 26, 1591-1601.
- Bekker-Jensen, S., Lukas, C., Melander, F., Bartek, J., and Lukas, J.** (2005). Dynamic assembly and sustained retention of 53BP1 at the sites of DNA damage are controlled by Mdc1/NFBD1. *The Journal of Cell Biology* 170, 201-211.
- Bologna, N.G., and Voinnet, O.** (2014). The Diversity, Biogenesis, and Activities of Endogenous Silencing Small RNAs in Arabidopsis. *Annual Review of Plant Biology* 65, 473-503.
- Böttcher, R., Hollmann, M., Merk, K., Nitschko, V., Obermaier, C., Philippou-Massier, J., Wieland, I., Gaul, U., and Förstemann, K.** (2014). Efficient chromosomal gene modification with CRISPR/cas9 and PCR-based homologous recombination donors in cultured *Drosophila* cells. *Nucleic acids research*.
- Brabant, F., and Acheson, N.H.** (1995). RNA footprint mapping of RNA polymerase II molecules stalled in the intergenic region of polyomavirus DNA. *Journal of Virology* 69, 4423-4430.

- Brawerman, G.** (1974). Chapter 1 The Isolation of RNA from Mammalian Cells<sup>1</sup>. In *Methods in Cell Biology*, M.P. David, ed. (Academic Press), pp. 1-22.
- Bregman, D.B., Halaban, R., van Gool, A.J., Henning, K.A., Friedberg, E.C., and Warren, S.L.** (1996). UV-induced ubiquitination of RNA polymerase II: a novel modification deficient in Cockayne syndrome cells. *Proceedings of the National Academy of Sciences of the United States of America* *93*, 11586-11590.
- Bunch, H., Lawney, B.P., Lin, Y.-F., Asaithamby, A., Murshid, A., Wang, Y.E., Chen, B.P.C., and Calderwood, S.K.** (2015). Transcriptional elongation requires DNA break-induced signalling. *6*, 10191.
- Ceccaldi, R., Liu, J.C., Amunugama, R., Hajdu, I., Primack, B., Petalcorin, M.I.R., O'Connor, K.W., Konstantinopoulos, P.A., Elledge, S.J., Boulton, S.J., et al.** (2015). Homologous recombination-deficient tumors are hyper-dependent on POLQ-mediated repair. *Nature* *518*, 258-262.
- Celeste, A., Fernandez-Capetillo, O., Kruhlak, M.J., Pilch, D.R., Staudt, D.W., Lee, A., Bonner, R.F., Bonner, W.M., and Nussenzweig, A.** (2003). Histone H2AX phosphorylation is dispensable for the initial recognition of DNA breaks. *Nat Cell Biol* *5*, 675-679.
- Cerosaletti, K., Wright, J., and Concannon, P.** (2006). Active Role for Nibrin in the Kinetics of Atm Activation. *Molecular and Cellular Biology* *26*, 1691-1699.
- Chan, S.H., Yu, A.M., and McVey, M.** (2010). Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in *Drosophila*. *PLoS genetics* *6*, e1001005.
- Chang, H.H.Y., Pannunzio, N.R., Adachi, N., and Lieber, M.R.** (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol* *advance online publication*.
- Chapman, J.R., and Jackson, S.P.** (2008). Phospho-dependent interactions between NBS1 and MDC1 mediate chromatin retention of the MRN complex at sites of DNA damage. *EMBO Reports* *9*, 795-801.
- Chapman, R.D., Conrad, M., and Eick, D.** (2005). Role of the Mammalian RNA Polymerase II C-Terminal Domain (CTD) Nonconsensus Repeats in CTD Stability and Cell Proliferation. *Molecular and Cellular Biology* *25*, 7665-7674.
- Chen, H., Kobayashi, K., Miyao, A., Hirochika, H., Yamaoka, N., and Nishiguchi, M.** (2013). Both OsRecQ1 and OsRDR1 Are Required for the Production of Small RNA in Response to DNA-Damage in Rice. *PLoS ONE* *8*, e55252.
- Choder, M., and Aloni, Y.** (1988). RNA polymerase II allows unwinding and rewinding of the DNA and thus maintains a constant length of the transcription bubble. *Journal of Biological Chemistry* *263*, 12994-13002.

- Churchman, L.S., and Weissman, J.S.** (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* *469*, 368-373.
- Churchman, L.S., and Weissman, J.S.** (2012). Native Elongating Transcript Sequencing (NET-seq). In *Current Protocols in Molecular Biology* (John Wiley & Sons, Inc.).
- Cogoni, C., and Macino, G.** (1997). Isolation of quelling-defective (qde) mutants impaired in posttranscriptional transgene-induced gene silencing in *Neurospora crassa*. *Proceedings of the National Academy of Sciences* *94*, 10233-10238.
- Cogoni, C., and Macino, G.** (1999). Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. *Nature* *399*, 166-169.
- Cong, Y.-S., Wright, W.E., and Shay, J.W.** (2002). Human Telomerase and Its Regulation. *Microbiology and Molecular Biology Reviews* *66*, 407-425.
- Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R., et al.** (2008). An endogenous small interfering RNA pathway in *Drosophila*. *Nature* *453*, 798-802.
- Deshpande, G., Calhoun, G., and Schedl, P.** (2005). *Drosophila argonaute-2* is required early in embryogenesis for the assembly of centric/centromeric heterochromatin, nuclear division, nuclear migration, and germ-cell formation. *Genes Dev* *19*, 1680-1685.
- Di Francesco, A., De Pittà, C., Moret, F., Barbieri, V., Celotti, L., and Mognato, M.** (2013). The DNA-Damage Response to  $\gamma$ -Radiation Is Affected by miR-27a in A549 Cells. *International Journal of Molecular Sciences* *14*, 17881.
- Dumesic, P.A., Natarajan, P., Chen, C., Drinnenberg, I.A., Schiller, B.J., Thompson, J., Moresco, J.J., Yates, J.R., 3rd, Bartel, D.P., and Madhani, H.D.** (2013). Stalled spliceosomes are a signal for RNAi-mediated genome defense. *Cell* *152*, 957-968.
- Dynan, W.S., and Burgess, R.R.** (1979). In vitro transcription by wheat germ ribonucleic acid polymerase II: effects of heparin and role of template integrity. *Biochemistry* *18*, 4581-4588.
- Dynan, W.S., and Burgess, R.R.** (1981). In vitro transcription by wheat germ RNA polymerase II. Initiation of RNA synthesis on relaxed, closed circular template. *Journal of Biological Chemistry* *256*, 5866-5873.
- Eick, D., and Geyer, M.** (2013). The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code. *Chemical Reviews* *113*, 8456-8490.
- Elmer, K., Helfer, S., Mirkovic-Hosle, M., and Förstemann, K.** (2014). Analysis of endo-siRNAs in *Drosophila*. *Methods in molecular biology* *1173*, 33-49.
- Ferrari, F., Plachetka, A., Alekseyenko, Artyom A., Jung, Youngsook L., Oszolak, F., Kharchenko, Peter V., Park, Peter J., and Kuroda, Mitzi I.** (2013). "Jump Start and Gain"

Model for Dosage Compensation in *Drosophila* Based on Direct Sequencing of Nascent Transcripts. *Cell Reports* 5, 629-636.

**Förstemann, K., Horwich, M.D., Wee, L., Tomari, Y., and Zamore, P.D.** (2007). *Drosophila* microRNAs Are Sorted into Functionally Distinct Argonaute Complexes after Production by Dicer-1. *Cell* 130, 287-297.

**Förstemann, K., Tomari, Y., Du, T., Vagin, V.V., Denli, A.M., Bratu, D.P., Klattenhoff, C., Theurkauf, W.E., and Zamore, P.D.** (2005). Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol* 3, e236.

**Francia, S., Michelini, F., Saxena, A., Tang, D., de Hoon, M., Anelli, V., Mione, M., Carninci, P., and di Fagagna, F.d.A.** (2012). Site-specific DICER and DROSHA RNA products control the DNA-damage response. *Nature advance online publication*.

**Francia, S., Michelini, F., Saxena, A., Tang, D., de Hoon, M., Anelli, V., Mione, M., Carninci, P., d'Adda di Fagagna, F.** (2012). Site-specific DICER and DROSHA RNA products control the DNA-damage response. *Nature* doi:10.1037/nature11179.

**Futatsumori-Sugai, M., Abe, R., Watanabe, M., Kudou, M., Yamamoto, T., Ejima, D., Arakawa, T., and Tsumoto, K.** (2009). Utilization of Arg-elution method for FLAG-tag based chromatography. *Protein Expression and Purification* 67, 148-155.

**Gao, M., Wei, W., Li, M.M., Wu, Y.S., Ba, Z., Jin, K.X., Li, M.M., Liao, Y.Q., Adhikari, S., Chong, Z., et al.** (2014). Ago2 facilitates Rad51 recruitment and DNA double-strand break repair by homologous recombination. *Cell Res* 24, 532-541.

**Gascioli, V., Mallory, A.C., Bartel, D.P., and Vaucheret, H.** (2005). Partially redundant functions of *Arabidopsis* DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. *Curr Biol* 15, 1494-1500.

**Gatz, A.M.** (2017). Analysis of early biogenesis steps of double-strand break triggered small RNA response by NET-Seq and strand-specific RT-qPCR. Master Thesis, Förstemann laboratory.

**Giuliani, G., Giuliani, F., Volk, T., and Rabouille, C.** (2014). The *Drosophila* RNA-binding protein HOW controls the stability of dgrasp mRNA in the follicular epithelium. *Nucleic Acids Research* 42, 1970-1986.

**Gu, W., Wind, M., and Reines, D.** (1996). Increased accommodation of nascent RNA in a product site on RNA polymerase II during arrest. *Proceedings of the National Academy of Sciences of the United States of America* 93, 6935-6940.

**Hain, D., Bettencourt, B.R., Okamura, K., Csorba, T., Meyer, W., Jin, Z., Biggerstaff, J., Siomi, H., Hutvagner, G., Lai, E.C., et al.** (2010). Natural Variation of the Amino-Terminal

---

Glutamine-Rich Domain in *Drosophila* Argonaute2 Is Not Associated with Developmental Defects. *PLoS ONE* 5, e15264.

**Hartig, J.V., Esslinger, S., Bottcher, R., Saito, K., and Förstemann, K.** (2009). Endo-siRNAs depend on a new isoform of loquacious and target artificially introduced, high-copy sequences. *Embo J* 28, 2932-2944.

**Hartig, J.V., and Förstemann, K.** (2011). Loqs-PD and R2D2 define independent pathways for RISC generation in *Drosophila*. *Nucleic acids research* 39, 3836-3851.

**Henry, R.W., Mittal, V., Ma, B., Kobayashi, R., and Hernandez, N.** (1998). SNAP19 mediates the assembly of a functional core promoter complex (SNAPc) shared by RNA polymerases II and III. *Genes & Development* 12, 2664-2672.

**Heyer, W.-D., Ehmsen, K.T., and Liu, J.** (2010). Regulation of Homologous Recombination in Eukaryotes. *Annual Review of Genetics* 44, 113-139.

**Iannelli, F., Galbiati, A., Capozzo, I., Nguyen, Q., Magnuson, B., Michelini, F., D'Alessandro, G., Cabrini, M., Roncador, M., Francia, S., et al.** (2017). A damaged genome's transcriptional landscape through multilayered expression profiling around in situ-mapped DNA double-strand breaks. 8, 15656.

**Ira, G., and Haber, J.E.** (2002). Characterization of RAD51-Independent Break-Induced Replication That Acts Preferentially with Short Homologous Sequences. *Molecular and Cellular Biology* 22, 6384-6392.

**Ivanov EL, S.N., Fishman-Lobell J, Haber JE** (1996). Genetic Requirements for the Single-Strand Annealing Pathway of Double-Strand Break Repair in *Saccharomyces Cerevisiae*. *Genetics* 142, 693-704.

**Kent, T., Chandramouly, G., McDevitt, S.M., Ozdemir, A.Y., and Pomerantz, R.T.** (2015). Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase theta. *Nat Struct Mol Biol* 22, 230-237.

**Kim, J., Sturgill, D., Tran, A.D., Sinclair, D.A., and Oberdoerffer, P.** (2016). Controlled DNA double-strand break induction in mice reveals post-damage transcriptome stability. *Nucleic Acids Research* 44, e64-e64.

**Kim, K., Lee, Y.S., and Carthew, R.W.** (2007). Conversion of pre-RISC to holo-RISC by Ago2 during assembly of RNAi complexes. *RNA* 13, 22-29.

**Kruhlak, M., Crouch, E.E., Orlov, M., Montano, C., Gorski, S.A., Nussenzweig, A., Misteli, T., Phair, R.D., and Casellas, R.** (2007). The ATM repair pathway inhibits RNA polymerase I transcription in response to chromosome breaks. *Nature* 447, 730-734.

- Kunzelmann, S., Böttcher, R., Schmidts, I., and Förstemann, K.** (2016). A Comprehensive Toolbox for Genome Editing in Cultured *Drosophila melanogaster* Cells. *G3: Genes|Genomes|Genetics* *6*, 1777-1785.
- Kunzelmann, S., and Förstemann, K.** (2017). Reversible perturbations of gene regulation after genome editing in *Drosophila* cells. *PLOS ONE* *12*, e0180135.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L.** (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* *10*, R25.
- LaRocque, J.R., Jaklevic, B., Su, T.T., and Sekelsky, J.** (2007). *Drosophila* ATR in Double-Strand Break Repair. *Genetics* *175*, 1023-1033.
- Laver, J.D., Marsolais, A.J., Smibert, C.A., and Lipshitz, H.D.** (2015). Regulation and Function of Maternal Gene Products During the Maternal-to-Zygotic Transition in *Drosophila*. *Current Topics in Developmental Biology* *113*, 43-84.
- Lee, H.C., Aalto, A.P., Yang, Q., Chang, S.S., Huang, G., Fisher, D., Cha, J., Poranen, M.M., Bamford, D.H., and Liu, Y.** (2010). The DNA/RNA-dependent RNA polymerase QDE-1 generates aberrant RNA and dsRNA for RNAi in a process requiring replication protein A and a DNA helicase. *PLoS biology* *8*.
- Lee, H.C., Chang, S.S., Choudhary, S., Aalto, A.P., Maiti, M., Bamford, D.H., and Liu, Y.** (2009). qiRNA is a new type of small interfering RNA induced by DNA damage. *Nature* *459*, 274-277.
- Lee, J.-H., and Paull, T.T.** (2004). Direct Activation of the ATM Protein Kinase by the Mre11/Rad50/Nbs1 Complex. *Science* *304*, 93-96.
- Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W.** (2004). Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* *117*, 69-81.
- Lehmann, E., Brueckner, F., and Cramer, P.** (2007). Molecular basis of RNA-dependent RNA polymerase II activity. *Nature* *450*, 445-449.
- Li, C., Vagin, V.V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M.D., Syrzycka, M., Honda, B.M., *et al.*** (2009). Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* *137*, 509-521.
- Liu, B., Liu, M., Wang, J., Zhang, X., Wang, X., Wang, P., Wang, H., Li, W., and Wang, Y.** (2015). DICER-dependent biogenesis of let-7 miRNAs affects human cell response to DNA damage via targeting p21/p27. *Nucleic Acids Res* *43*, 1626-1636.
- Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.E., Smith, D.P., and Wang, X.** (2003). R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science* *301*, 1921-1925.

- Lou, Z., Minter-Dykhouse, K., Franco, S., Gostissa, M., Rivera, M.A., Celeste, A., Manis, J.P., van Deursen, J., Nussenzweig, A., Paull, T.T., et al.** (2006). MDC1 Maintains Genomic Stability by Participating in the Amplification of ATM-Dependent DNA Damage Signals. *Molecular Cell* *21*, 187-200.
- Lukas, C., Melander, F., Stucki, M., Falck, J., Bekker-Jensen, S., Goldberg, M., Lerenthal, Y., Jackson, S.P., Bartek, J., and Lukas, J.** (2004). Mdc1 couples DNA double-strand break recognition by Nbs1 with its H2AX-dependent chromatin retention. *The EMBO Journal* *23*, 2674-2683.
- Madigan, J.P., Chotkowski, H.L., and Glaser, R.L.** (2002). DNA double-strand break-induced phosphorylation of Drosophila histone variant H2Av helps prevent radiation-induced apoptosis. *Nucleic Acids Research* *30*, 3698-3705.
- Maida, Y., Yasukawa, M., Furuuchi, M., Lassmann, T., Possemato, R., Okamoto, N., Kasim, V., Hayashizaki, Y., Hahn, W.C., and Masutomi, K.** (2009). An RNA dependent RNA polymerase formed by hTERT and the RNase MRP RNA. *Nature* *461*, 230-235.
- Mannuss, A., Dukowic-Schulze, S., Suer, S., Hartung, F., Pacher, M., and Puchta, H.** (2010). RAD5A, RECQ4A, and MUS81 Have Specific Functions in Homologous Recombination and Define Different Pathways of DNA Repair in Arabidopsis thaliana. *The Plant Cell* *22*, 3318-3330.
- Martin, N.T., Nakamura, K., Davies, R., Nahas, S.A., Brown, C., Tunuguntla, R., Gatti, R.A., and Hu, H.** (2013). ATM-Dependent MiR-335 Targets CtIP and Modulates the DNA Damage Response. *PLOS Genetics* *9*, e1003505.
- Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., and Zamore, P.D.** (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* *123*, 607-620.
- Matts, J.A., Sytnikova, Y., Chirn, G.-w., Igloi, G.L., and Lau, N.C.** (2014). Small RNA library construction from minute biological samples. *Methods in molecular biology (Clifton, NJ)* *1093*, 123-136.
- Matzke, M.A., and Mosher, R.A.** (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet* *15*, 394-408.
- Mayer, A., and Churchman, L.S.** (2016). Genome-wide profiling of RNA polymerase transcription at nucleotide resolution in human cells with native elongating transcript sequencing. *Nat Protocols* *11*, 813-833.
- Mayer, A., Iulio, J.d., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A., and Churchman, L.S.** (2015). Native Elongating Transcript Sequencing Reveals Human Transcriptional Activity at Nucleotide Resolution. *Cell* *161*, 541-554.

- Mayfield, J.E., Robinson, M.R., Cotham, V.C., Irani, S., Matthews, W.L., Ram, A., Gilmour, D.S., Cannon, J.R., Zhang, Y.J., and Brodbelt, J.S.** (2017). Mapping the Phosphorylation Pattern of *Drosophila melanogaster* RNA Polymerase II Carboxyl-Terminal Domain Using Ultraviolet Photodissociation Mass Spectrometry. *ACS Chemical Biology* *12*, 153-162.
- McVey, M., and Lee, S.E.** (2008). MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends in genetics : TIG* *24*, 529-538.
- McVey, M., Radut, D., and Sekelsky, J.J.** (2004). End-joining repair of double-strand breaks in *Drosophila melanogaster* is largely DNA ligase IV independent. *Genetics* *168*, 2067-2076.
- Mehta, A., and Haber, J.E.** (2014). Sources of DNA double-strand breaks and models of recombinational DNA repair. *Cold Spring Harb Perspect Biol* *6*, a016428.
- Melander, F., Bekker-Jensen, S., Falck, J., Bartek, J., Mailand, N., and Lukas, J.** (2008). Phosphorylation of SDT repeats in the MDC1 N terminus triggers retention of NBS1 at the DNA damage-modified chromatin. *The Journal of Cell Biology* *181*, 213-226.
- Merk, K., Breinig, M., Böttcher, R., Krebs, S., Blum, H., Boutros, M., and Förstemann, K.** (2017). Splicing stimulates siRNA formation at *Drosophila* DNA double-strand breaks. *PLOS Genetics* *13*, e1006861.
- Michalik, K.M., Böttcher, R., and Förstemann, K.** (2012). A small RNA response at DNA ends in *Drosophila*. *Nucleic acids research* *40*, 9596-9603.
- Mirkin, E.V., and Mirkin, S.M.** (2007). Replication Fork Stalling at Natural Impediments. *Microbiology and Molecular Biology Reviews : MMBR* *71*, 13-35.
- Nguyen, V.T., Kiss, T., Michels, A.A., and Bensaude, O.** (2001). 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature* *414*, 322-325.
- Nojima, T., Gomes, T., Carmo-Fonseca, M., and Proudfoot, N.J.** (2016). Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. *Nat Protocols* *11*, 413-428.
- Nojima, T., Gomes, T., Grosso, Ana Rita F., Kimura, H., Dye, Michael J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, Nicholas J.** (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* *161*, 526-540.
- Nolan, T., Braccini, L., Azzalin, G., De Toni, A., Macino, G., and Cogoni, C.** (2005). The post-transcriptional gene silencing machinery functions independently of DNA methylation to repress a LINE1-like retrotransposon in *Neurospora crassa*. *Nucleic Acids Research* *33*, 1564-1573.

- Ohle, C., Tesorero, R., Schermann, G., Dobrev, N., Sinning, I., and Fischer, T.** (2016). Transient RNA-DNA Hybrids Are Required for Efficient Double-Strand Break Repair. *Cell* *167*, 1001-1013.e1007.
- Okamura, K., Chung, W.J., Ruby, J.G., Guo, H., Bartel, D.P., and Lai, E.C.** (2008). The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* *453*, 803-806.
- Okamura, K., Ishizuka, A., Siomi, H., and Siomi, M.C.** (2004). Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev* *18*, 1655-1666.
- Orel, N., Kyryk, A., and Puchta, H.** (2003). Different pathways of homologous recombination are used for the repair of double-strand breaks within tandemly arranged sequences in the plant genome. *The Plant Journal* *35*, 604-612.
- Pankotai, T., Bonhomme, C., Chen, D., and Soutoglou, E.** (2012). DNAPKcs-dependent arrest of RNA polymerase II transcription in the presence of DNA breaks. *Nat Struct Mol Biol* *19*, 276-282.
- Peterlin, B.M., Brogie, J.E., and Price, D.H.** (2012). 7SK snRNA: a noncoding RNA that plays a major role in regulating eukaryotic transcription. *Wiley interdisciplinary reviews RNA* *3*, 92-103.
- Pierce, A.J., Johnson, R.D., Thompson, L.H., and Jasin, M.** (1999). XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes & development* *13*, 2633-2638.
- Portz, B., Lu, F., Gibbs, E.B., Mayfield, J.E., Rachel Mehaffey, M., Zhang, Y.J., Brodbelt, J.S., Showalter, S.A., and Gilmour, D.S.** (2017). Structural heterogeneity in the intrinsically disordered RNA polymerase II C-terminal domain. *Nature Communications* *8*, 15231.
- Potts, P.R., Porteus, M.H., and Yu, H.** (2006). Human SMC5/6 complex promotes sister chromatid homologous recombination by recruiting the SMC1/3 cohesin complex to double-strand breaks. *The EMBO Journal* *25*, 3377-3388.
- Prado, F., and Aguilera, A.** (2005). Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *The EMBO Journal* *24*, 1267-1276.
- Quaresma, A.J.C., Bugai, A., and Barboric, M.** (2016). Cracking the control of RNA polymerase II elongation by 7SK snRNP and P-TEFb. *Nucleic Acids Research* *44*, 7527-7539.
- Raha, D., Wang, Z., Moqtaderi, Z., Wu, L., Zhong, G., Gerstein, M., Struhl, K., and Snyder, M.** (2010). Close association of RNA polymerase II and many transcription factors with Pol III genes. *Proceedings of the National Academy of Sciences of the United States of America* *107*, 3639-3644.

- Renkawitz, J., Lademann, C.A., and Jentsch, S.** (2014). Mechanisms and principles of homology search during recombination. *Nat Rev Mol Cell Biol* *15*, 369-383.
- Romano, N., and Macino, G.** (1992). Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. *Molecular Microbiology* *6*, 3343-3353.
- Schmidts, I., Böttcher, R., Mirkovic-Hösle, M., and Förstemann, K.** (2016). Homology directed repair is unaffected by the absence of siRNAs in *Drosophila melanogaster*. *Nucleic Acids Research* *44*, 8261-8271.
- Schnitzler, G.R.** (2001). Isolation of Histones and Nucleosome Cores from Mammalian Cells. In *Current Protocols in Molecular Biology* (John Wiley & Sons, Inc.).
- Seddon, A.M., Curnow, P., and Booth, P.J.** (2004). Membrane proteins, lipids and detergents: not just a soap opera. *Biochimica et Biophysica Acta (BBA) - Biomembranes* *1666*, 105-117.
- Seitz, H., Ghildiyal, M., and Zamore, P.D.** (2008). Argonaute loading improves the 5' precision of both MicroRNAs and their miRNA strands in flies. *Curr Biol* *18*, 147-151.
- Shanbhag, N.M., Rafalska-Metcalf, I.U., Balane-Bolivar, C., Janicki, S.M., and Greenberg, R.A.** (2010). An ATM-Dependent Transcriptional Silencing Program is Transmitted Through Chromatin in Cis to DNA Double Strand Breaks. *Cell* *141*, 970-981.
- Stark, J.M., Pierce, A.J., Oh, J., Pastink, A., and Jasin, M.** (2004). Genetic Steps of Mammalian Homologous Repair with Distinct Mutagenic Consequences. *Molecular and Cellular Biology* *24*, 9305-9316.
- Steponkus, P.L., Dowgert, M.F., and Gordon-Kamm, W.J.** (1983). Destabilization of the plasma membrane of isolated plant protoplasts during a freeze-thaw cycle: The influence of cold acclimation. *Cryobiology* *20*, 448-465.
- Symington, L.S., and Gautier, J.** (2011). Double-strand break end resection and repair pathway choice. *Annual review of genetics* *45*, 247-271.
- Thomas, A.M., Hui, C., South, A., and McVey, M.** (2013). Common Variants of *Drosophila melanogaster* Cyp6d2 Cause Camptothecin Sensitivity and Synergize With Loss of Brca2. *G3: Genes|Genomes|Genetics* *3*, 91-99.
- Tomari, Y., Du, T., Haley, B., Schwarz, D.S., Bennett, R., Cook, H.A., Koppetsch, B.S., Theurkauf, W.E., and Zamore, P.D.** (2004a). RISC assembly defects in the *Drosophila* RNAi mutant armitage. *Cell* *116*, 831-841.
- Tomari, Y., Du, T., and Zamore, P.D.** (2007). Sorting of *Drosophila* Small Silencing RNAs. *Cell* *130*, 299-308.
- Tomari, Y., Matranga, C., Haley, B., Martinez, N., and Zamore, P.D.** (2004b). A protein sensor for siRNA asymmetry. *Science* *306*, 1377-1380.

- Tresini, M., Marteijn, J.A., and Vermeulen, W.** (2016). Bidirectional coupling of splicing and ATM signaling in response to transcription-blocking DNA damage. *RNA Biology* *13*, 272-278.
- Uziel, T., Lerenthal, Y., Moyal, L., Andegeko, Y., Mittelman, L., and Shiloh, Y.** (2003). Requirement of the MRN complex for ATM activation by DNA damage. *The EMBO Journal* *22*, 5612-5621.
- van Gent, D.C., Hoeijmakers, J.H.J., and Kanaar, R.** (2001). Chromosomal stability and the DNA double-stranded break connection. *Nat Rev Genet* *2*, 196-206.
- Wagner, S.D., Yakovchuk, P., Gilman, B., Ponicsan, S.L., Drullinger, L.F., Kugel, J.F., and Goodrich, J.A.** (2013). RNA polymerase II acts as an RNA-dependent RNA polymerase to extend and destabilize a non-coding RNA. *The EMBO Journal* *32*, 781-790.
- Wang, Q., and Goldstein, M.** (2016). Small RNAs Recruit Chromatin-Modifying Enzymes MMSET and Tip60 to Reconfigure Damaged DNA upon Double-Strand Break and Facilitate Repair. *Cancer Research* *76*, 1904-1915.
- Weber, Christopher M., Ramachandran, S., and Henikoff, S.** (2014). Nucleosomes Are Context-Specific, H2A.Z-Modulated Barriers to RNA Polymerase. *Molecular Cell* *53*, 819-830.
- Wei, W., Ba, Z., Gao, M., Wu, Y., Ma, Y., Amiard, S., White, C.I., Danielsen, J.M., Yang, Y.G., and Qi, Y.** (2012). A Role for Small RNAs in DNA Double-Strand Break Repair. *Cell* *149*, 101-112.
- White, E., Schlackow, M., Kamieniarz-Gdula, K., Proudfoot, N.J., and Gullerova, M.** (2014). Human nuclear Dicer restricts the deleterious accumulation of endogenous double-stranded RNA. *Nat Struct Mol Biol* *21*, 552-559.
- Wickersheim, M.L., and Blumenstiel, J.P.** (2013). Terminator oligo blocking efficiently eliminates rRNA from Drosophila small RNA sequencing libraries. *BioTechniques* *55*, 269-272.
- Wilson, T.E., and Lieber, M.R.** (1999). Efficient Processing of DNA Ends during Yeast Nonhomologous End Joining: EVIDENCE FOR A DNA POLYMERASE  $\beta$  (POL4)-DEPENDENT PATHWAY. *Journal of Biological Chemistry* *274*, 23599-23609.
- Xie, A., Kwok, A., and Scully, R.** (2009). Role of mammalian Mre11 in classical and alternative non-homologous end joining. *Nature structural & molecular biology* *16*, 814-818.
- Yan, D., Ng, W.L., Zhang, X., Wang, P., Zhang, Z., Mo, Y.-Y., Mao, H., Hao, C., Olson, J.J., Curran, W.J., et al.** (2010). Targeting DNA-PKcs and ATM with miR-101 Sensitizes Tumors to Radiation. *PLOS ONE* *5*, e11397.
- Yang, Q., Ye, Q.A., and Liu, Y.** (2015). Mechanism of siRNA production from repetitive DNA. *Genes Dev* *29*, 526-537.
- Yang, Y.-G., and Qi, Y.** (2015). RNA-directed repair of DNA double-strand breaks. *DNA Repair* *32*, 82-85.

- Yang, Z., Zhu, Q., Luo, K., and Zhou, Q.** (2001). The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature* *414*, 317-322.
- Yi, X., Tesmer, V.M., Savre-Train, I., Shay, J.W., and Wright, W.E.** (1999). Both Transcriptional and Posttranscriptional Mechanisms Regulate Human Telomerase Template RNA Levels. *Molecular and Cellular Biology* *19*, 3989-3997.
- Yu, A.M., and McVey, M.** (2010). Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. *Nucleic acids research* *38*, 5706-5717.
- Zhang, Z., Chang, S.S., Xue, Z., Zhang, H., Li, S., and Liu, Y.** (2013). Homologous recombination as a mechanism to recognize repetitive DNA sequences in an RNAi pathway. *Genes & development* *27*, 145-150.
- Zhang, Z., Yang, Q., Sun, G., Chen, S., He, Q., Li, S., and Liu, Y.** (2014). Histone H3K56 Acetylation Is Required for Quelling-induced Small RNA Production through Its Role in Homologous Recombination. *The Journal of Biological Chemistry* *289*, 9365-9371.

---

## List of Figures

Fig. 1: Common steps in homologous recombination.	5
Fig. 2: The DGUS.US-1 reporter for SSA.	9
Fig. 3: A GFP based SSA reporter for <i>Drosophila melanogaster</i> .	11
Fig. 4: The comparison of small RNA libraries after control treatment or transfection of the I-SceI nuclease shows that introduction of a break in the reporter leads to generation of siRNAs against the break.	12
Fig. 5: Schematic overview of the short tract gene conversion reporter that was established by the Jasin lab (Pierce et al., 1999) and its potential repair products.	12
Fig. 6: Exploiting our C-terminal tagging approach to assess Rad51-dependent HDR repair.	14
Fig. 7: C-terminal tagging of <i>RpII33</i> .	22
Fig. 8: C-terminal tagging of <i>RpII215</i> .	24
Fig. 9: The ‘standard’ laboratory protocol was modified to test the influence of different disruption strategies on RNAPII release from the nucleus.	25
Fig. 10: Investigating the influence of different detergents on RNAPII extraction from the nucleus.	26
Fig. 11: Cellular fractionation and comparison of the influence of different detergents on RNAPII extraction from the nucleus.	27
Fig. 12: RNAPII is tightly associated to chromatin and pelleted by centrifugation of nuclear fractions.	28
Fig. 13: Chromatin-wash of the nuclear fraction using a high KCl buffer facilitates RNAPII release from chromatin.	29
Fig. 14: Sonication facilitates release of the hyperphosphorylated RpII215 from the nucleus.	30
Fig. 15: RpII215 release from chromatin is reproducibly facilitated by hypertonic lysis with a KCl containing buffer and sonication.	31
Fig. 16: Comparison of lysis by the optimized protocol (high KCl, sonication) and protein extraction by grinding.	32
Fig. 17: Immunoprecipitation of RpII33-Flag <sub>2</sub> using Protein G beads that were pre-incubated with $\alpha$ Flag (M2).	33
Fig. 18: Comparison of elution methods to elute immunopurified RpII33-Flag <sub>2</sub> .	34
Fig. 19: Immunoprecipitation and specific elution of RpII215-Strep <sub>2</sub> using Strep-Tactin Type2HC beads and 1x Biotin-elution buffer.	35
Fig. 20: Specificity of immunoprecipitation tested by NET-RT-PCR.	37
Fig. 21: T7 endonuclease assay to detect error-prone DSB repair for the respective loci.	39
Fig. 22: Distribution of sequence lengths over all sequences.	41
Fig. 23: Sense and antisense reads mapping to <i>CG15098</i> for three different conditions.	45

Fig. 24: Sense and antisense reads mapping to <i>7SK</i> for three different conditions (cut in <i>7SK</i> , cut in <i>Tcp</i> and uncut control).	50
Fig. 25: Knock-down of the NMD factor Upf1 leads to de-repression of GFP expression from the long reporter construct.	52
Fig. 26: HDR efficiencies of the SSA reporter are unaffected by knock-down of siRNA factors.	53
Fig. 27: Depletion of certain small RNA biogenesis factors leads to cell cycle perturbations.	55
Fig. 28: Introducing a <i>lig4</i> deletion in the SSA reporter cell line.	57
Fig. 29: Schematic overview of a representative cross for the <i>in vivo</i> SSA assay.	59
Fig. 30: I- <i>SceI</i> expressing animals are mosaic and allow assessing the balance between SSA and end-joining repair pathway choice.	59
Fig. 31: Measurement of mean GFP fluorescence within the fly eyes of <i>dcr-2</i> deficient animals and heterozygous control.	60
Fig. 32: Analysis of mean GFP protein expression of <i>dcr-2</i> proficient and deficient animals assessed by Western blot.	61
Fig. 33: Assessing the influence of HR donor topology on the dependence of HDR repair on Rad51.	63
Fig. 34: HDR efficiencies of genome editing are unaffected by knock-down of siRNA factors.	63
Fig. 35: The <i>dcr-2</i> <sup>06544</sup> allele has impaired function in transgenic RNA interference.	67
Fig. 36: PCR-based assay to detect <i>cyp6d2</i> splicing defective (SD) or no transcript (NT) variants.	68

---

## List of Tables

Table 1: Conditions tested for specific elution of RpII33-Flag <sub>2</sub> using the 3X Flag-peptide.	33
Table 2: Assessment of the technical quality of IS03.	40
Table 3: Contribution of 2S rRNA reads to all clipped reads.	41
Table 4: Percentage of clipped reads of a read length $\geq 17$ nt after <i>in silico</i> removal of 2S rRNA mapping to the indicated sequence collections from Flybase.	42
Table 5: Number of clipped reads of the indicated read length mapping to <i>CG15098</i> .	45
Table 6: Number of clipped reads of the indicated read length that map to <i>Tctp</i> .	47
Table 7: Number of clipped reads of the indicated read length (size selection was performed <i>in silico</i> using the Galaxy tool Filter-FastQ) that map to <i>7SK</i> .	50
Table 8: Assessing the general DSB stress response of <i>dcr-2</i> and <i>ago2</i> mutant flies upon treatment with camptothecin.	66
Table 9: Oligos used for generation of the sgRNA directed against the C-terminus of RpII33.	79
Table 10: Oligos used for overlap-extension PCR to generate sgRNA templates that are transcribed <i>in vivo</i> by RNAPIII.	80
Table 11: Oligos used for generation of HR donor templates with short homology arms (60 nt).	81
Table 12: Overview on plasmids that were used as templates for HR donor template generation.	81
Table 13: Overview of oligos designed for construction of pIS1	82
Table 14: Sequences for generation of dsRNA.	83
Table 15: Overview on cell lines that were used in this study.	83
Table 16: Primers used for amplification of the indicated loci for a T7 endonuclease assay.	87
Table 17: Fly strains used in this study.	93

---

## Acknowledgments

An erster Stelle möchte ich mich ganz herzlich bei meinem Doktorvater Professor Klaus Förstemann bedanken, für die Chance an zwei spannenden Projekten zum Thema DSB-induzierte siRNAs zu arbeiten. Ich bin dir sehr dankbar, dass wir trotz „Negativresultaten“ das Projekt zur biologischen Rolle der siRNAs bis zur Publikation weiterverfolgt haben. Danke für dein Vertrauen, deine Hilfe und für dein offenes Ohr und Verständnis in allen Belangen. Danke auch für die Vorab-Korrektur dieser Arbeit.

Herzlichen Dank auch an Dietmar Martin für die Übernahme des Zweitgutachtens. Ihm und allen Mitgliedern meiner Prüfungskommission vielen Dank für die investierte Zeit. Meinem TAC Komitee, Mario Halic und Katja Lammens, ganz besonderen Dank für ihre Zeit und all die zahl- und hilfreichen Anmerkungen zu meinem Projekt.

Ich möchte allen aktuellen und ehemaligen Mitgliedern der AG Förstemann ganz herzlich für die tolle Arbeitsatmosphäre danken, die schönen Feiern und den Zusammenhalt innerhalb der Gruppe, die meine Zeit am Genzentrum so unendlich bereichert haben. Ich werde die gemeinsamen Kaffeepausen sehr vermissen.

Liebe Karin, ich danke dir für deine langjährige Freundschaft, für den Unterschlupf, den du mir zu mancher Gelegenheit gewährt hast, für die Hilfe beim Umzug, die Tanz-Nachhilfe und die lustigen gemeinsamen Bastel- und Backaktionen.

Liebe Romy, du standest mir über die Jahre immer mit Rat und Tat zur Seite, sowohl im Labor als auch privat. Unvergessen z.B. unsere erste gemeinsame FACS Session, bei der der Plate Reader so überhaupt nicht mitspielen wollte und wir well für well mitgefiebert haben bis in die späten Abendstunden. Danke auch für das große Umstyling von 2015 :) Danke, dass du immer für mich da warst, um (dumme :) Fragen zu beantworten, Mut zu machen und den Rücken zu stärken.

Lieber Stefan: danke für die tolle Nachbarschaft :), für das wilde Hypothesen-Aufstellen, dass du dir selbst im größten Stress noch Zeit genommen hast und für deine Anmerkungen zu dieser Arbeit.

Lieber Volker, danke für all die wissenschaftlichen und nicht so wissenschaftlichen Diskussionen und für deinen wunderbaren Humor.

Liebe Tiana, vielen Dank für eine tolle gemeinsame Zeit im Labor und auf dem Retreat. Danke für's Zuhören, Aufmuntern und Ablenken. Und danke für die Last-Minute Korrektur. :)

Alan, Stephan, Filip und Anna einen ganz herzlichen Dank für euer Engagement und eure tolle Arbeit. Liebe Anna シ, どうもありがとう für deine tolle, aufgeschlossene, positive und liebe Art, deinen Einsatz (お疲れ様でした!) und unsere Japanisch-Sessions. 楽しかったです。 Und nicht zuletzt für die Vorab-Korrektur dieser Arbeit.

Lieber Fabian danke für die gemeinsamen Unternehmungen, die Hilfe beim Umzug, all die witzigen Gespräche und dafür, dass du manches in ein etwas entspannteres Licht rückst :)

Ganz besonders möchte ich mich bei meinen Eltern für ihre Liebe, ihre Geduld und all die Unterstützung bedanken. Danke, dass ihr mir Studium und Promotion ermöglicht habt. Danke, dass ihr immer an mich glaubt und immer für mich da seid. Danke auch für euer Interesse an meinen jeweiligen Forschungsprojekten über all die Jahre hinweg. Auch an meinen Bruder für die vielen Aufmunterungen ganz herzlichen Dank.

Und nicht zuletzt, möchte ich mich auf das allerherzlichste bei Bene bedanken. こころのそこからあいしてるよ。 Danke, dass du mich unterstützt, immer für mich da bist und an mich glaubst, mich motivierst, mich zum Lachen bringst und unendlich viel Geduld mit mir hast.