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

The spliceosome-mediated siRNA biogenesis at DNA double-strand breaks in *Drosophila melanogaster*

Karin Martina Merk

aus
Mutlangen, Deutschland

2023

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, 23.03.2023

……..
Karin Merk

Dissertation eingereicht am 28.03.2023

1. Gutachter: Prof. Dr. Klaus Förstemann

2. Gutachter: Prof. Dr. Roland Beckmann

Mündliche Prüfung am 15.06.2023

Summary

Small interfering RNAs (siRNAs) arise from a variety of exogenous and endogenous triggers and exhibit important regulatory and protective functions in genome surveillance and gene regulation. From a long double-stranded RNA (dsRNA) precursor of different origin, perfectly complementary 21nt small RNAs are created via dicing and loaded onto an effector protein to yield functional siRNAs. In *Neurospora crassa, Arabidopsis thaliana, Drosophila melanogaster,* rice and human cells, DNA double-strand breaks (DSBs) lead to the formation of small RNAs at the damage site.

As previously described, the siRNA response to DSBs in *D. melanogaster* depends on transcriptional activity at the damaged locus and can be induced by transfection of a linearized plasmid causing the repression of a reporter sequence *in trans.* To answer the question how a DNA double-strand break leads to the formation of a double-stranded RNA precursor and thus the formation of locus-derived siRNAs, I validated the hits from a genome-wide reporter-based RNAi screen in *D. melanogaster* cells for factors promoting or inhibiting this siRNA response. Besides proteins involved in either DSB recognition/processing, or the siRNA biogenesis pathway, spliceosomal components were identified to promote DSB-induced siRNA production. In contrast, mRNA processing and polyadenylation of the nascent transcripts seem to reduce siRNA formation.

In additional experiments, I could then exclude an indirect effect of the knockdown of certain splicing factors on the protein levels of alternatively spliced siRNA biogenesis pathway key players such as Ago2 and Dcr-2 – their levels remain unchanged. Furthermore, the steady-state levels of spliced/unspliced pre-mRNA transcripts from selected non-siRNA factor related genes are largely independent of the knockdown of individual identified spliceosomal components. This suggests a direct role of the spliceosome in translating the DSB into an siRNA response rather than general effects of the performed knockdowns.

Through introducing site-specific DNA double-strand breaks at selected endogenous loci in *Drosophila* cells via CRISPR-Cas9-mediated cleavage followed by deep sequencing of the resulting small RNAs, I showed further that DNA breaks in intronless genes or upstream of a gene's first intron do not trigger siRNA production. Contrarily, cleavage at a minimum distance downstream of the first intron, thus implying the co-transcriptionally acting spliceosome to be present on the nascent transcript at the damaged locus, leads indeed to transcription-dependent small RNA formation with the siRNAs matching the area between the transcription start site (TSS) and the induced DNA break. The thereby observed siRNAs are equally distributed across introns and exons as well as intron/exon junctions. This suggests that the unspliced pre-mRNA contributes the sense strand to the dsRNA precursor and either the unspliced pre-mRNA or the DNA must serve as template for the synthesis of the antisense strand. Furthermore, the small RNAs covering intronic and exonic regions of the cleaved locus, as well as the

highly abundant components of the pre-catalytic spliceosome within the validated factors promoting siRNA production at DSBs, indicate that the spliceosome is stalled in a pre-catalytic state when encountering a double-strand break. Although I could not directly show any physical interaction of a potentially important kinase involved in splicing with a prospective substrate, the obtained results clearly suggest that the stalled spliceosome triggers the biogenesis of an antisense strand complementary to the unspliced pre-mRNA at the break, thus stimulating the DSB-induced siRNA response.

In addition to DNA double-strand breaks, other features such as high-copy transgenes, structured sgRNAs synthesized from transfected PCR templates and a natural intron transplanted to a different locus trigger an siRNA response visible in small RNA deep-sequencing experiments. Comparable to the DSB-induced siRNAs, the production of siRNAs at high copy transgenes is also stimulated by splicing factors, as shown via validation screens with additional reporters. The spliceosome-dependent siRNA signal investigated and presented in this thesis is thus not limited to DNA double-strand breaks but rather describes a more widely occurring genome defense mechanism.

Table of contents

1 Introduction

Among other triggers, endogenous small interfering RNAs (endo-siRNAs) are formed in several organisms when DNA double-strand breaks (DSBs) occur in actively transcribed intron-containing genes. The following introduction will give insights into the current understanding of sources, recognition and repair of DNA double-strand breaks, transcription and co-transcriptional splicing of protein-coding genes, the triggers and biogenesis of small non-coding RNAs and the interplay of these processes in the context of DNA damage.

1.1 DNA double-strand breaks challenge genome integrity

The integrity and continuity of the genomic information encoded in the DNA of cells is constantly challenged by various intrinsic and extrinsic factors. Different forms of DNA damage caused by a number of agents and processes can be repaired via specific pathways to re-establish functionality (Hoeijmakers 2009; Mehta and Haber 2014). Due to the disruption of both strands of the DNA molecule, double-strand breaks are the most severe form of DNA damage for genome integrity as the sequence information cannot be directly recovered from the complementary strand. These lesions can result from certain exogenous factors such as chemicals and therapeutics, ionizing radiation and (natural) UV-light but also from a number of endogenous sources (Mehta and Haber 2014; Hoeijmakers 2009). For example, DNA double-strand breaks are caused by reactive oxygen species emerging from cellular processes or by stalled replication forks resulting from collisions with obstacles such as altered DNA (nucleotides or chromatin structure), the transcription machinery or DNA binding molecules (Mirkin and Mirkin 2007; Aguilera and Gaillard 2014). Furthermore, processes like stalled transcription (Hamperl and Cimprich 2014) or transcription-coupled repair (Sollier et al. 2014) can facilitate the formation of R-loops and thus increase genomic instability and DNA damage. However, DNA double-strand breaks are also essential or beneficial for instance for topoisomerase II-dependent chromatin remodeling (Morimoto et al. 2019), meiotic recombination (Székvölgyi, Ohta, and Nicolas 2015) or the formation of T and B cell receptors (van Gent, Hoeijmakers, and Kanaar 2001) and are thus deliberately introduced into the genome during these processes.

As DNA damage occurs frequently but can lead to severe malfunctions and diseases, it is crucial for the cells to recognize the defects and facilitate their repair or induce apoptosis. For the repair of the highly mutagenic DNA double-strand breaks, there are basically two distinct mechanisms: non-homologous end-joining (NHEJ) pathways and homology-directed repair (HR) (Symington and Gautier 2011; Wyman and Kanaar 2006). In addition to the actual repair, DNA damage also affects a number of processes at the break, such as chromatin remodeling and transcription. Furthermore, DNA damage influences, as well as is influenced by, general processes like cell cycle regulation or apoptosis signaling.

1.2 The protein response to DNA double-strand breaks

1.2.1 Double-strand break recognition and ATM/ATR signaling

As a first step in DNA double-strand break sensing and repair, the so-called MRN complex consisting of Mre11, Rad50 and Nbs1 (nbs in *Drosophila*) recognizes and binds the ends of the broken DNA strands. The bound complex interacts with and thus recruits the signaling kinases ataxia telangiectasia mutated (ATM) and ataxia telangiectasia and Rad3-related protein (ATR) to the site of DNA damage (Lee and Paull 2004; Cerosaletti, Wright, and Concannon 2006).

This so-called primary recruitment leads to the activation of signaling kinases (Uziel et al. 2003). The activated mammalian ATR and ATM kinases regulate cell-cycle progression in response to DNA damage (Abraham 2001) and trigger a downstream signaling cascade by phosphorylating more than 700 cellular proteins (Matsuoka et al. 2007). One key target of the ATM and ATR kinases in the response to (ionizing radiation-induced) DNA damage in mammalian cells is the histone variant H2Ax which is phosphorylated at a specific serine residue. In *D. melanogaster*, the only existing H2A variant H2Av is also phosphorylated in response to DNA damage (Madigan, Chotkowski, and Glaser 2002). While the initial recruitment of the MRN complex to the double-strand breaks occurs rapidly and independent of the formation of γ-H2Ax, the phosphorylation triggers the localization of additional MRN complexes as well as ATM/ATR kinases to the site of DNA damage (Celeste et al. 2003). Phosphorylated Mdc1 hereby acts as an adaptor between the phosphorylated H2Ax and ATM (Lou et al. 2006). Another protein which is itself phosphorylated by ATM but also promotes ATM signaling is 53BP1. Although the factor by itself rapidly and transiently binds unphosphorylated H2Ax, the increased binding of 53BP1 to γ-H2Ax is again facilitated by phosphorylated Mdc1 (Bekker-Jensen et al. 2005). A third factor which can bind to both the unphosphorylated H2Ax and more sustainably to γ-H2Ax via interaction with the phosphorylated Mdc1 is Nbs1 (Lukas et al. 2004). While the binding and recruitment of 53BP1 and other proteins such as BRCA1 to γ-H2Ax foci is facilitated by Mdc1 upon phosphorylation by ATM, the binding and recruitment of additional Nbs1 and thus MRN complexes to these foci requires phosphorylation of Mdc1 by Casein kinase 2 (Chapman and Jackson 2008; Melander et al. 2008).

The initial protein response to DNA double-strand breaks is thus amplified and extended to the area around the recognized break via an ATM/ATR-mediated signaling cascade. As the two signaling kinases are conserved, there is both an ATM (tefu) and ATR (mei-41) orthologue in *D. melanogaster*. However, the main functions of the *Drosophila* kinases differ from the mammalian system: the cell-cycle response (de Vries et al. 2005; Bayer et al. 2018) is predominantly facilitated by the ATR-orthologue mei-41, whereas the ATM-homologous tefu protein acts in telomere stabilization (Oikemus et al. 2006) and p53-dependent apoptosis (LaRocque et al. 2007).

In addition to the recruitment of DNA repair factors to the damage site, phosphorylation by ATM/ATR also causes or regulates the ubiquitination, SUMOylation or methylation of histones as well as other modifications of chromatin-remodeling factors or additional (DNA-binding) proteins thus influencing the accessibility of the DNA for either the DNA repair machinery or transcription-related enzymes. Furthermore, the recruited or modified proteins can control cell-cycle progression or alter protein interactions and thus influence DNA repair (Lai et al. 2013; Lukas, Lukas, and Bartek 2011).

1.2.2 Processing of the DNA ends and repair pathway choice

Subsequent to their recognition, DNA double-strand breaks are repaired via different repair pathways in a cell-cycle-dependent manner. Both, the non-homologous end-joining pathways where the DNA ends are basically re-ligated, and the homology-dependent repair pathways involving DNA resection and synthesis at the break, require accessible or even specifically processed DNA ends (Symington and Gautier 2011).

“Clean” DNA ends can be rapidly bound and processed by the classical non-homologous end-joining machinery, including the Ku70/Ku80 factors, DNA-dependent protein kinase catalytic subunit (DNA-PKcs, however there are no PKcs currently known to be present in *D. melanogaster*) and Ligase IV (Ceccaldi, Rondinelli, and D'Andrea 2016; Liao, Tammaro, and Yan 2016; Doré et al. 2004). In contrast, ends containing damaged nucleotides or blocked DNA (e.g. through covalently bound proteins) need early processing by the nuclease Mre11 in complex with Rad50 and Nbs1 (MRN complex) (Cejka 2015; Liao, Tammaro, and Yan 2016). Homologues of this complex exist in both prokaryotes (without Nbs) and eukaryotes (with Nbs). As a first step, the DNA ends are bound and kept in close proximity by the Mre11-Rad50 complex (Wyman and Kanaar 2006). During the so-called short-range resection, the MRN complex facilitates an endo-nucleolytic cleavage of the DNA typically 15-25bp away from an obstacle which can be of different nature. In addition, the MRN complex executes a 3' to 5' exonuclease activity towards the double-strand break starting at a previously induced endo-nucleolytic cleavage site thus creating a 3' overhang of up to 200-300 nucleotides (Symington and Gautier 2011). These various functions of the MRN complex are achieved by (different) conformational changes in the Rad50-Mre11 heterotetramer upon binding of DNA (Kashammer et al. 2019). In the ATP-dependent cleavage mechanism CtIP, a protein conserved across all eukaryotic kingdoms (Uanschou et al. 2007), acts as a co-factor of Mre11 (Anand et al. 2016; Liao, Tammaro, and Yan 2016). In particular the endonuclease activity of Mre11 is both facilitated and regulated through the interaction with phosphorylated CtIP (Anand et al. 2016).

As a second step, either exonuclease 1 (Exo1) or the Sgs1-Dna2 helicase nuclease are recruited to the double-strand break by the MRN-complex. The DNA ends are further resected (5' to 3' direction) to create extensive 3' ssDNA overhangs either from clean ends or as part of bidirectional resection together with the MRN complex (Garcia et al. 2011; Cejka 2015). For the bidirectional resection, an endo-nucleolytic cleavage is introduced by the MRN complex up to 300nt away from the DNA lesion. The

MRN-complex then further removes the nucleotides towards the DNA lesion (3' to 5' exonuclease activity), whereas the processive exonucleases Exo1 or Sgs1-Dna2 extend the 3' overhang via 5' to 3' resection of the 5' strand (Garcia et al. 2011). It has further been suggested that the 3' overhang DNA could be protected by forming a DNA-RNA-hybrid with antisense RNA, however the mechanism of antisense transcription at the DSB is still under discussion. While some data in human cells indicate that RNA Pol III might be recruited to the double-strand break by the MRN complex and the processing of the DNA end by MRN/CtIP is required for Pol III transcription initiation (Liu et al. 2021), other experiments in *Drosophila* show that the antisense transcription at DNA double-strand breaks is performed by RNA Pol II (Böttcher et al. 2021).

Mre11 also has a high affinity to 3' overhang ssDNA and catalyzes DNA unwinding independent of its nuclease activity as shown in *Saccharomyces cerevisiae* (Ghodke and Muniyappa 2013). Furthermore, results from *Schizosaccharomyces pombe* demonstrate, that the Mre11 nuclease activity and the CtIP homologue are necessary to release either the Ku70/Ku80 complex or the MRN complex from DNA ends thus enabling binding of RPA and DNA repair (Langerak et al. 2011).

The resection and hence the pathway choice is highly regulated and influenced by the cell cycle state in which the damage occurs. While the fast non-homologous end-joining repair predominantly occurs in G0/G1 phase, the presence of the sister chromatid in late S and G2 phase allows for homologous recombination (Ceccaldi, Rondinelli, and D'Andrea 2016; Cejka 2015). The phosphorylation of different substrates involved in end resection, such as CtIP or Exo1, by cyclin-dependent kinases influences the repair pathway choice. For example in late S/G2 phase, the chromatin remodeler CSB and BRCA1 interact with the MRN-complex and CtIP and promote end resection, thus triggering homologous recombination (Batenburg et al. 2019).

In addition to the nature of the DNA ends and the cell-cycle state in which the DNA damage occurs, the chromatin context of the double-strand break plays an important role in the repair pathway choice (Aymard et al. 2014). Thus, double-strand breaks that occur in actively transcribed genes are preferably resected and repaired via homologous recombination compared to double-strand breaks occurring in intergenic or silent loci.

1.2.3 Different repair mechanisms for DNA double-strand breaks

Dependent on the nature of the DNA ends at the break, the cell-cycle state and the extent of resection, double-strand breaks are repaired via non-homologous end-joining (Ku70/Ku80-bound clean ends), micro-homology mediated end-joining (MMEJ) or homologous recombination pathways (extensively resected ends) (Symington and Gautier 2011; Ceccaldi, Rondinelli, and D'Andrea 2016; Wyman and Kanaar 2006).

Non-homologous end-joining (NHEJ) can take place in any cell-cycle state and occurs most frequently in G0/G1 phase when no sister chromatid is available for other less error-prone repair pathways. For NHEJ, the DNA ends are bound and protected from extensive resection by the Ku70/Ku80 complex which also recruits DNA-PKcs (Mahaney, Meek, and Lees-Miller 2009). Based on the end structure at the lesion, there are several sub-pathways of NHEJ (Chang et al. 2017). Blunt ends as well as "clean" ends with homology overhangs (e.g. created with restriction enzymes) can be directly and error-free ligated by DNA ligase IV in complex with XRCC4/XLF and Ku70/Ku80. Incompatible ends are either resected by Artemis-DNA-PKcs with or without using micro-homologies of only a few nucleotides and/or extended via nucleotide synthesis by DNA polymerases Polμ and Polλ prior to ligation. If necessary for ligation, the 5'/3' ends are modified by kinases/phosphatases to obtain the mandatory 3' hydroxyl and 5' phosphate groups (Wyman and Kanaar 2006). The resection and/or nucleotide synthesis prior to ligation most likely leads to (small) insertions or deletions in the DNA sequence (Chang et al. 2017). However, additional less error-prone non-homologous end-joining pathways can be observed for example for actively transcribed genes (Chakraborty et al. 2016).

The micro-homology mediated end-joining (MMEJ) pathway is an alternative end-joining pathway which neither requires the classical NHEJ proteins Ku70/Ku80 and DNA ligase IV, nor Rad51. In contrast to NHEJ, MMEJ requires limited resection of the DNA ends by the MRN complex and CtIP to reveal short (8-22bp) homologies close to the break (Symington and Gautier 2011; McVey and Lee 2008). The homology-regions on both sides of the break are then annealed and the resulting flaps are removed by Rad1 and Rad10 endonucleases. DNA polymerase θ, in *D. melanogaster* encoded by the *mus308* gene, then fills the gaps prior to ligation (Chan, Yu, and McVey 2010; Kent et al. 2015). In the synthesis-dependent MMEJ pathway variant, the micro-homology is created de novo by polymerase θ (Yu and McVey 2010). Finally, a ligase other than DNA ligase IV (e.g. ligase I or ligase III) completes the MMEJ, thus introducing mostly short sequence deletions adjacent to the original site of the break (McVey and Lee 2008). MMEJ can occur throughout the cell-cycle. However, both the initiation of NHEJ and the extensive resection of the DNA ends to induce homologous recombination prevent MMEJ (Symington and Gautier 2011).

When DNA double-strand breaks occur in late S/G2 phase when sister chromatids are available as templates, the lesions can be repaired via homologous recombination (HR). After extensive 5' to 3' resection which is a prerequisite for homologous recombination, the ssDNA overhangs are covered and thus protected by replication protein A (RPA). RPA is then exchanged for the recombinase Rad51, a process mediated by BRCA2 and Rad52, which forms nucleoprotein filaments with the ssDNA (Renkawitz, Lademann, and Jentsch 2014). During the following homology search, the filament pairs with the donor sequence of the undamaged sister chromatid in a strand exchange thereby forming a so-called displacement loop (D-loop). The D-loop is then extended by polymerase δ with the sliding clamp PCNA loaded by RFC1-5 (Wright, Shah, and Heyer 2018). D-loop extension and subsequent ligation

give rise to a double Holliday junction (dHJ) intermediate which can be resolved in different ways (Wyman and Kanaar 2006). Resolution by a helicase and topoisomerase results in non-crossover products, whereas the resolvases can yield both crossover and non-crossover products (Wyman and Kanaar 2006; Renkawitz, Lademann, and Jentsch 2014). Alternatively, D-loops can also be resolved without dHJ-formation via synthesis-dependent strand-annealing (SDSA), resulting in a non-crossover product, or cause break-induced replication (BIR) (Mehta and Haber 2014; Wright, Shah, and Heyer 2018). All these homologous recombination pathways lead to basically error-free repair.

The single-strand annealing pathway (SSA) uses homologous sequences of repeated DNA regions on both sides of the double-strand break to anneal them in a Rad51-independent manner (Mehta and Haber 2014). These homologies are made accessible through extensive resection at the damage site and the ssDNA ends are bound by Rad52. The non-homologous flaps created through annealing of the homologous repeat sequences are cleaved by Rad1-Rad10 and the annealed strands are ligated by DNA ligase 1 (Ceccaldi, Rondinelli, and D'Andrea 2016). In contrast to other homologous recombination pathways, SSA results in large deletions as the region between the repeated sequences and one of the repeats is removed during repair.

1.3 Transcription and co-transcriptional activities at DNA double-strand breaks

1.3.1 Transcription and DNA damage

In eukaryotic cells, the DNA dependent RNA polymerases I, II and III transcribe different classes of genes. RNA polymerase I synthesizes the 28S, 18S and 5.8S rRNAs which are obtained from a common precursor through modification and processing. Protein-coding genes are transcribed by RNA polymerase II (Pol II) and processed into mRNAs. Furthermore, various non-coding small nuclear and small cytosolic RNAs (snRNAs/scRNAs) are also transcribed by Pol II. Other small non-coding RNAs of various functions are transcribed by RNA polymerase III (Pol III) as well as all tRNA genes and the 5S rRNA (Dieci et al. 2007). In addition, Pol III can also transcribe cytosolic DNA from pathogens, thus creating specific RNAs which trigger an immune response (Chiu, Macmillan, and Chen 2009).

All eukaryotic RNA polymerases consist of several subunits and require additional transcription factors for promotor recognition, transcription initiation and elongation (Neish et al. 1998; Reuter and Strasser 2016). At the transcription bubble, the double-stranded DNA is unwound to access the template strand for RNA synthesis. The nascent transcript then forms a DNA:RNA hybrid with the coding strand and the non-template strand is displaced in a loop. The single stranded DNA (ssDNA) of the non-template strand can thus be bound by ssDNA binding proteins such as Sub1 (initiation) or RPA (elongation) (Sikorski et al. 2011). These so-called R-loops consisting of a DNA:RNA hybrid and a displaced ssDNA loop, can form at different loci, preferentially in GC-rich regions e.g. close to promotors or in 3' UTRs during transcription (nascent transcript acts *in cis*), but also *in trans* independent of transcription (Sollier and

Cimprich 2015). During transcription, R-loop structures can contribute to either initiation or termination (Skourti-Stathaki, Kamieniarz-Gdula, and Proudfoot 2014) in relation to their localization. As transcription progresses along the DNA, the local R-loop structures are gradually dissolved. However, if transcription is stalled or paused, the formation of extensive R-loop structures can cause genomic instability for example through causing collisions between the stalled transcription and the replication machineries (Hamperl and Cimprich 2014) or by inducing chromatin condensation (Castellano-Pozo et al. 2013).

Initiation, elongation and termination of Pol II transcription, as well as co-transcriptional RNA processing, is regulated via the carboxyterminal domain (CTD) of Rpb1, the largest subunit of RNA polymerase II (Hsin and Manley 2012). The CTD consists of heptapeptide repeats (52 repeats in vertebrates, 39 in *D. melanogaster*) leading to the globally conserved but locally heterogeneous and flexible structure of this domain (Portz et al. 2017). While the CTD remains unphosphorylated during initiation the heptapeptides of the CTD are constantly modified, mostly phosphorylated, in distinct patterns throughout the further steps of transcription. The induced structural changes facilitate the targeted binding of proteins involved in various processes such as chromatin remodeling, transcription termination, 5' capping, splicing and 3' processing of the mRNA precursor thus interlinking the different processes (Mayer et al. 2012; Meinhart and Cramer 2004; Hsin and Manley 2012).

While DNA damage on the one hand poses an obstacle for transcription that requires repair or bypassing mechanisms to resume the synthesis of the nascent RNA (Pankotai and Soutoglou 2013), transcription can on the other hand also play a role in DNA repair (Sollier and Cimprich 2015; Puget, Miller, and Legube 2019). When an RNA polymerase encounters a site of DNA damage during transcription, the actions at this location must be orchestrated and regulated. Specific factors such as CSA/CSB/TFIIH are recruited to the site and transcription-coupled nucleotide excision repair (TC-NER) is instigated. The damaged nucleotides are removed and the gap is filled and re-ligated. Accessibility for the repair machinery is ensured by either backtracking or removal and proteasomal degradation of the stalled polymerase (Gaillard and Aguilera 2013; Vermeulen and Fousteri 2013). Furthermore, chromatin remodeling is induced to enable polymerase backtracking and to prevent transcription (Pankotai and Soutoglou 2013; Vermeulen and Fousteri 2013). At DNA double-strand breaks, the proteasomal degradation of the stalled RNA polymerase is triggered through the activity of DNA protein kinase (DNAPK). In absence of DNAPK, the double-strand break can be bypassed by the RNA polymerase as it is the case for certain non-bulky DNA lesions in an error-prone fashion (Pankotai et al. 2012; Pankotai and Soutoglou 2013). In addition to inducing the transcription-coupled repair pathway for damaged nucleotides, the DNA:RNA hybrid R-loop structures at DNA double-strand breaks of actively transcribed loci could even play a role in DSB repair by providing the RNA molecule as a template for DNA repair (Keskin et al. 2014; Chakraborty et al. 2016) or via regulating the accessibility of ssDNA at DSBs in repetitive regions for end resection factors (Ohle et al. 2016). It has been observed, that the formation of

DNA:RNA hybrids is induced at DSBs of human cells by Drosha and that both NHEJ and HR repair efficiencies are reduced in the absence of R-loops (Lu et al. 2018).

Besides impacting DNA repair, transcription can also be a cause of DNA damage or recombination when R-loops in actively transcribed regions are not resolved but persist or are processed. When the replication machinery collides heads-on with a paused RNA polymerase or the positively supercoiled DNA between the two machineries, the stalling of replication can lead to DNA double-strand breaks (Prado and Aguilera 2005; Aguilera and Gaillard 2014). A co-directional collision of the replication machinery with a persisting R-loop behind a stalled RNA polymerase may also cause replication arrest potentially followed by a replication restart either after recombination or by using the R-loop structure as Primer (Aguilera and Gaillard 2014). Furthermore, the processing of DNA damage by the TC-NER endonucleases at an R-loop formation can result in DNA double-strand breaks (Sollier et al. 2014).

In contrast to transcription itself, various RNA processing steps which are mostly started or performed co-transcriptionally reduce R-loop formation (Wickramasinghe and Venkitaraman 2016; Hamperl and Cimprich 2014). The binding of different proteins to the nascent RNA prevents annealing with the DNA to form an R-loop. The same applies for the formation of a secondary RNA structure as observed for certain classes of RNAs. When the nascent transcript is spliced, the resulting RNA molecule is not complementary to the DNA template anymore and can thus per se not re-hybridize with the DNA to form R-loop structures. Finally, packaging of the nascent mRNA into ribonucleoprotein complexes (mRNPs) and association with the mRNA-export complex renders the RNA unavailable for R-loop formation. R-loops can also actively be resolved by either the RNAse H family nucleases via unwinding or by specific DNA:RNA helicases.

1.3.2 Splicing and genome stability

In eukaryotes, the different steps of pre-mRNA processing occur co-transcriptionally and are coordinated via distinct phosphorylation patterns of the Pol II CTD. Factors involved in these processes are recruited to the CTD and the respective processing step starts when the corresponding region (e.g. the intronic sequence for pre-mRNA splicing) has been transcribed. Thus, for the majority of mRNAs containing non-protein coding intronic regions, the introns at the 5'-end of the already capped nascent pre-mRNA are removed by the splicing machinery while the RNA is still being elongated at the 3'-end (Khodor et al. 2011). While small introns of less than 70nt are extremely rare and inefficiently spliced in mammals, more than half of the *Drosophila* introns are shorter than 80nt and short introns can be spliced out well (Mount et al. 1992). The consensus sequences around the conserved 5' splice site (AG|GU) and the 3' splice site (CAG|X) as well as the branchpoint within the intron (CUXAX) are highly similar between yeast, *Drosophila* and mammals (Mount et al. 1992).

The different steps of the splicing reaction are catalyzed by the snRNPs U1/U2/U4/U5/U6 each consisting of a specific snRNA, Sm proteins (Lsm proteins in U6 snRNP) and additional splicing factors. In the spliceosome, the snRNPs form large macromolecular complexes (>150 components in human) with additional splicing factors and protein complexes, with the respective composition and structure varying in each step of the splicing reaction (Valadkhan and Jaladat 2010; Hoskins and Moore 2012; Will and Luhrmann 2011). Both the structure and the composition of the spliceosome in the different complexes are highly conserved between *Drosophila* and human (Herold et al. 2009).

As a first step prior to the splicing reaction, the U1 snRNP binds to the 5' splice site, and the U2 snRNP with ATP and additional factors binds to and remodels the structure at the branchpoint, thus exposing the conserved adenosine (A complex). The U4, U5 and U6 snRNP form a tri-RNP-complex which enters the spliceosome by binding to U1 (B complex). In the next step, the U6 snRNP dissolves from U4 and pairs with U2 thus removing U1 from the 5' splice site. To activate the B complex upon removal of U1 and U4, the Prp19/Prp19-related protein complex enters the spliceosome to yield the activated B complex (B^{act}) in which U5 interacts with both the 5' and the 3' splice sites. This activation step also requires the RES complex (Bao et al. 2017). In the following B* complex, the branch point adenosine performs a nucleophilic attack on the 5' splice site resulting in the lariat structure of the C complex. In the second splicing step, the hydroxyl-group of the 5' splice site attacks the 3' splice site yielding the free spliced RNA and the remaining lariat. The U2, U5 and U6 snRNPs and other protein components are then recycled and the EJC/TREX complex is deposited on the spliced RNA (Herold et al. 2009; Wang, Murigneux, and Le Hir 2014).

In higher eukaryotes, most genes contain more than one intron thus allowing for several possible combinations of 5' splice sites and 3' splice sites within a pre-mRNA. In some cases, many different spliced mRNAs are created at the same time in the same tissue, whereas other transcript isoforms are selectively created at specific time points, in specific tissues or as response to certain stimuli, for example DNA damage signals (Brown et al. 2014; Marengo and Wassarman 2008). The alternate usage of certain splice sites is regulated via the binding of splicing factors, mainly RNA binding proteins such as SR proteins or hnRNPs (Brooks et al. 2015; Zhou and Fu 2013). The presence and function of these RNA binding factors is again regulated for example via phosphorylation by specific kinases or signal-induced degradation rendering alternative splicing a highly dynamic and finely tuned mechanism (Katzenberger, Marengo, and Wassarman 2009; Zhou and Fu 2013).

Similar to DNA damage and transcription/co-transcriptional activities in general, there is also a two-sided interconnection between splicing and the DNA damage response. Proteins involved in splicing or splicing regulation are either directly targeted for example by DNA damage response (DDR) signaling kinases and other modifying enzymes, subject to up-/downregulation of their protein levels, or impacted in their function by changes in the chromatin structure in response to DNA damage (Mikolaskova et al. 2018;

Lenzken, Loffreda, and Barabino 2013; Nilsson, Wu, and Schwartz 2018; Shkreta and Chabot 2015). On the other hand, the up-/down-regulated (alternative) splicing of several key proteins and additional factors indirectly regulates the corresponding processes of DDR signaling, end resection and DNA repair via modulated protein levels (Pederiva and Farnebo 2018; Adamson et al. 2012; Prados-Carvajal et al. 2018; Shkreta and Chabot 2015).

A special case of bidirectional coupling can be observed when transcription-blocking DNA damage (not DNA double-strand breaks) occurs at an actively spliced gene. The (late-stage) spliceosome is displaced from the transcript (possibly to enable backtracking of the RNA polymerase) thus rendering the RNA available for R-loop formation with the unwound DNA. The R-loop formation then induces non-canonical ATM-signaling which in turn leads to removal of spliceosomes and changes in alternative splicing at distant genomic regions to induce a diverse DDR (Tresini et al. 2015; Tresini, Marteijn, and Vermeulen 2016).

In general, both the binding of splicing factors/the spliceosome and the splicing reaction itself reduce R-loop formation by preventing the re-annealing of the nascent pre-mRNA with the complementary DNA template strand (Wickramasinghe and Venkitaraman 2016). The reduced R-loop formation upon introduction of an intron into an intronless gene in yeast suggests further, that introns might protect cells from genomic instability caused by R-loop formation during transcription (Bonnet et al. 2017). Furthermore, in *S. cerevisiae*, the binding of the spliceosome to intronless genes leads to spliceosome-mediated decay of the transcripts (Volanakis et al. 2013).

Besides co-transcriptional splicing of protein-coding mRNAs, splicing is also implicated to be crucial for small RNA-mediated silencing in different organisms. In *Cryptococcus neoformans,* transposon transcripts are identified via stalling of the spliceosome after the first catalytic step and subjected to RNAi by the Spliceosome-Coupled And Nuclear RNAi (SCANR) complex (Dumesic et al. 2013). In *D. melanogaster* the biogenesis of Piwi-interacting RNAs (piRNAs) is regulated by selectively targeting the non-spliced transcripts from piRNA clusters into the piRNA biogenesis pathway thus distinguishing them from (spliced) mRNAs (Zhang, Wang, et al. 2014). In fission yeast, splicing factors also seem to facilitate RNAi-mediated heterochromatin silencing (Bayne et al. 2008; Kallgren et al. 2014). However, comparable to the core piRNA effector Piwi in *D. melanogaster* (Malone et al. 2014), some main RNAi effector proteins in *S. pombe* require extensive splicing so that indirect effects of splicing via altered protein levels of RNAi proteins are also possible.

In addition to their role in splicing, a number of splicing factors also have further independent functions. The Prp19 complex which is required for activation of the B complex in splicing is also involved in genome maintenance, recruitment of ubiquitylated proteins to the proteasome, transcription elongation and genome maintenance (Chanarat and Sträßer 2013). In DNA damage response, the Prp19-complex binds to ssDNA-bound RPA and acts as ubiquitin ligase thus triggering ATRIP binding and ATR-kinase

signaling (Marechal et al. 2014). Another protein with additional functions in genome surveillance is the spliceosomal core protein SmD1 which interacts with core proteins of both the siRNA and the microRNA RNA-induced silencing complex (RISC) in *Drosophila,* suggesting a role for SmD1 in non-coding RNA mediated gene silencing independent from its role in splicing (Xiong et al. 2013; Xiong et al. 2015).

1.4 Small non-coding RNAs and DNA double-strand breaks

1.4.1 Different classes and pathways of small RNAs

The integrity of the genome is constantly challenged by endogenous and exogenous factors in the cells and gene expression needs to be regulated in response to these threads. Amongst diverse functions of different non-coding RNAs in various processes, small non-coding RNAs are involved in post-transcriptional regulation of genomic loci (Rother and Meister 2011). As first observed in petunia (van der Krol et al. 1990) and then described as RNA interference (RNAi) in the model organism *Caenorhabditis elegans* (Lee, Feinbaum, and Ambros 1993; Fire et al. 1998), small non-coding RNAs of antisense orientation can down-regulate the translation of mRNAs in many eukaryotes.

To facilitate RNA-induced silencing, long double-stranded RNA precursors (dsRNA) of different origins are cleaved into small dsRNAs by Dicer-proteins (Meister and Tuschl 2004; Lee et al. 2004; Wilson and Doudna 2013; Rother and Meister 2011; Carthew and Sontheimer 2009). The thus obtained small dsRNAs of defined length and different specific characteristics for each pathway are then loaded onto pathway-specific Argonaute-proteins to target the complementary RNA molecules as functional RNA-induced silencing complex (RISC) (Meister and Tuschl 2004; Okamura et al. 2004; Wilson and Doudna 2013; Czech and Hannon 2011). Independent of or in addition to their functions in RNAi, both Ago2 (Taliaferro et al. 2013; Gao et al. 2014; Carissimi et al. 2015) and Dicer (Castel et al. 2014; Kurzynska-Kokorniak et al. 2015; Wang et al. 2015) have been shown to have additional functions for example in transcription, splicing and DNA-repair.

In *D. melanogaster*, three different classes of small RNAs with specific triggers, mostly distinct processing machineries (Okamura et al. 2004; Yang et al. 2014; Zhou et al. 2008; Forstemann et al. 2007; Lee et al. 2004) and various functions are involved in RNA-induced silencing: piRNAs, microRNAs and siRNAs. Piwi-interacting RNAs (piRNAs) ensure genome stability in the germline whereas microRNAs (miRNAs) regulate gene expression. Small interfering RNAs (siRNAs) are processed from long dsRNA of exogenous (endo-siRNAs) or endogenous (exo-siRNAs) sources such as dsRNA viruses, hairpin-structured RNA, antisense transcription at high-copy loci and DNA double-strand breaks. Other eukaryotes also create different small dsRNA species. However, in contrast to *D. melanogaster,* different small RNAs such as miRNAs and siRNAs are processed by the same Dicer protein (Lee et al. 2013) and loaded into the same Argonaute protein for example in humans, thus rendering *D. melanogaster* a highly suitable organism to selectively investigate a certain class of small RNAs.

Of the three classes of small RNAs the biogenesis mechanism of the germline-specific piRNAs differs most from the others. The ~25-30nt piRNAs originate from specific heterochromatic regions called master loci which contain multiple transposon fragments thus providing a template for sequences that need to be repressed in the cell (Hartig, Tomari, and Förstemann 2007; Brennecke et al. 2007; Malone et al. 2009). The transcripts from these piRNA-clusters are amplified in a so-called ping-pong mechanism by the Argonaute proteins Ago3 (sense piRNA) and Aub (antisense piRNA). In addition to Aub, the antisense piRNA can also be loaded onto the Piwi Argonaute protein and prevent the activation and translocation of transposons.

MicroRNAs basically have two origins, RNA polymerase II transcripts originating from specific loci which form a hairpin structure (pri-miRNA) or intronic regions functioning as mirtrons after splicing of the pre-mRNA (Ghildiyal and Zamore 2009). The hairpin-structured pri-miRNAs are processed in the nucleus into the 60-70nt partly mismatched hairpin pre-miRNAs by the RNAse III enzyme Drosha building up the microprocessor complex with the dsRNA binding protein Pasha (Denli et al. 2004), and then exported to the cytoplasm. There, these pre-miRNAs as well as the pre-miRNAs obtained from mirtrons upon debranching (Okamura et al. 2007) are cleaved into 22nt dsRNAs of miRNA and miRNA* by Dicer-1 (Dcr-1) and the dsRBP isoform LoqsPB (Forstemann et al. 2005; Saito et al. 2005). The miRNA/miRNA* is then loaded into Ago1, the miRNA* is expelled and the miRNA forms the functional RNA-induced silencing complex (RISC) (Forstemann et al. 2007). MicroRNAs are not perfectly complementary to their target sequence and will therefore not induce cleavage but bind to their target mRNA via base-pairing of a short seed-sequence mostly in the 3' UTR of the mRNA thus inhibiting translation and inducing deadenylation. While most miRNAs are loaded onto Ago1, and the miRNA and siRNA pathways are mostly separate in *Drosophila,* distinct miRNAs displaying perfect complementarity to their target are sorted into Ago2 instead of Ago1 (Forstemann et al. 2007).

1.4.2 Biogenesis and triggers of small interfering RNAs

The third class of small non-coding dsRNAs, the small interfering RNAs, can be divided into exo-siRNAs and endo-siRNAs based on the source of the dsRNA precursor molecule. The biogenesis pathway from long dsRNA precursor to functionally loaded siRNAs is systematically shown for endogenous dsRNA sources in Figure 1-1. For both groups of siRNAs the long dsRNA precursor molecules are cut into 21nt dsRNAs of perfect complementarity with two nucleotide 3' overhangs and a 5' phosphate by Dcr-2 in complex with a dsRNA-binding protein (Ghildiyal and Zamore 2009). Next, the 21nt dsRNAs are loaded into Ago2 by a RISC-loading complex consisting of Dcr-2, chaperones and again a specific dsRBP (R2D2 for exo-siRNAs) (Kawamata and Tomari 2010; Meister 2013). During loading, the guide and passenger strands of the dsRNA are defined via sensing of the duplex formation energy of each end (Tomari et al. 2004). As a second step, separation of the guide and passenger strand is facilitated via endonucleolytic cleavage of the passenger strand by C3PO thus yielding the functional

RISC (Matranga et al. 2005). Guided by the complementarity of the included small RNA, the RISC slices the complementary transcribed RNAs or viral RNA leading to their subsequent degradation.

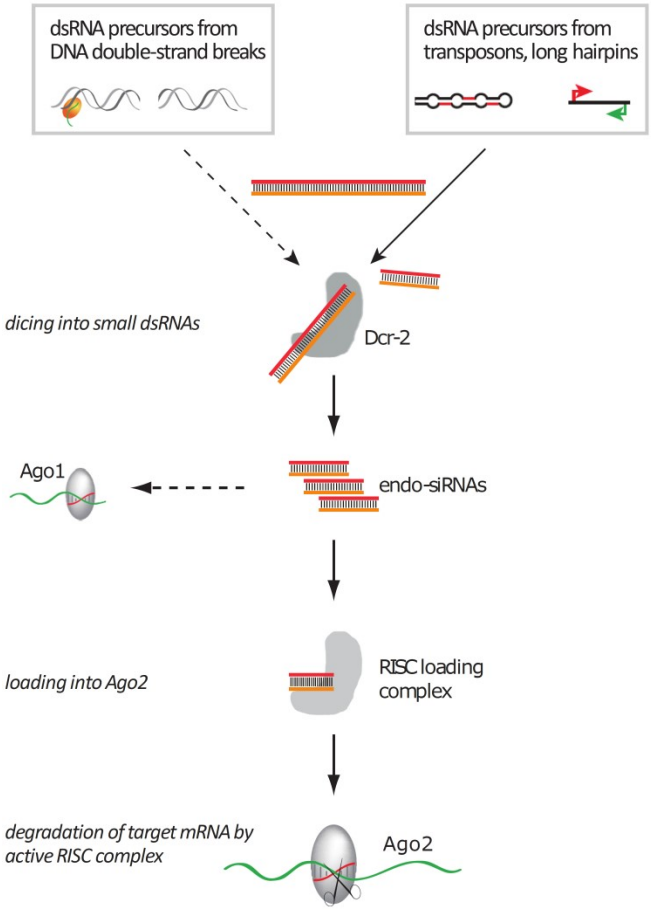

Figure 1-1: Schematic of the endo-siRNA pathway in Drosophila. The long dsRNA precursor resulting from long hairpin or antisense transcription at repetitive sequences or DNA double-strand breaks is processed by Dcr-2 into 21nt small dsRNAs. They are loaded mainly into Ago2 by a Dcr-2 containing RISC loading complex thus yielding the functional RISC complex for target transcript degradation. **Adapted from (Hartig and Forstemann 2011).**

Although the pathways for exo-siRNA and endo-siRNA biogenesis share the same mechanism and main effector proteins Dcr-2 and Ago2, the dsRBPs acting as co-factors of Dcr-2 in both the dicing and the RISC-loading steps differ. In general, R2D2 acts as cofactor for Dcr-2 in exo-siRNA biogenesis whereas the loquacious isoform Loqs-PD participates in endo-siRNA formation. However, both dsRBPs can also function redundantly in the exo- and the endo-siRNA pathways (Hartig et al. 2009; Hartig and Forstemann 2011; Mirkovic-Hosle and Forstemann 2014). Some endo-siRNAs can also be cross-loaded into Ago1, the effector protein of microRNAs to a limited extend (Ameres et al. 2011).

The exo-siRNA pathway is triggered by various viral dsRNA structures including replication intermediates of DNA and RNA viruses, genomic dsRNA, and unique RNA hairpin structures which are processed into functional siRNAs (Sabin et al. 2013; Li et al. 2013). Furthermore, external dsRNAs or siRNA-pools can deliberately be introduced into cells to facilitate targeted RNAi.

While transposable elements in the germline are repressed by piRNAs, endo-siRNAs have been shown to implement RNA mediated silencing in somatic cells in *Drosophila* (Chung et al. 2008) and high copy-number genes can lead to an endo-siRNA response (Cruz and Houseley 2014). Endogenous long dsRNAs of different origin can be entered into the endo-siRNA pathway as depicted in Figure 1-1. Convergent transcription, transcription of inverted repeats, long perfectly complementary hairpin structures and antisense transcription can give rise to long dsRNAs and thus RNAi induction (Czech et al. 2008; Okamura, Balla, et al. 2008; Okamura, Chung, et al. 2008; Ghildiyal and Zamore 2009; Okamura and Lai 2008; Russo, Harrington, and Steiniger 2016). As the underlying mechanisms for the production of the dsRNA precursor take place in the nucleus, the processing into siRNAs is, however, performed by Dcr-2 in the cytoplasm, it has been shown that the dsRBP blanks can shuttle the dsRNAs between those compartments (Nitschko et al. 2020).

1.4.3 DNA double-strand break induced small RNAs and DSB repair

In addition to transposable elements, DNA double-strand breaks have been identified as a trigger for endo-siRNA formation in various organisms. The firstly observed DSB-triggered small RNA species are the so-called quelling-induced RNAs (qiRNAs) which are formed at the only repetitive (rDNA) locus (or any locus artificially inserted in tandem repeats) upon provoking replicative stress in *Neurospora crassa* (Lee et al. 2009). These 20-21nt qiRNA are processed by the same factors as the ~25nt small RNAs involved in silencing of transgenes and transposon defense (Nolan et al. 2005; Romano and Macino 1992). This quelling mechanism in *N. crassa* involves a helicase, an enzyme with both, DNA- and RNA-dependent RNA polymerase activity, an Argonaute protein and Dicers (Cogoni and Macino 1997; Cogoni and Macino 1999). In *N. crassa* there is a bidirectional link between the DDR and qiRNAs: the recruitment of Rad51 to the DSBs is a pre-requisite for qiRNA formation and the qiRNA response supports homologous recombination (Yang, Ye, and Liu 2015; Zhang, Yang, et al. 2014).

Besides the fungus *N. crassa,* break-derived small RNAs are also observed in plants such as *Arabidopsis thaliana* (Wei et al. 2012) and rice (Chen et al. 2013), human cells (Francia et al. 2012; Wei et al. 2012) and *D. melanogaster* (Michalik, Bottcher, and Forstemann 2012).

In rice, the helicase OsRecQ1 and the RNA-dependent RNA polymerase OsRdR1 are required for small RNA formation, and there is non-coding RNA formation at the rDNA locus (Chen et al. 2013) comparable to the RNA observed in *N. crassa*. The dsRNA precursor for the *A. thaliana* small RNAs are synthesized by RNA pol IV and an RNA-dependent RNA polymerase (Wei et al. 2012). There are, however, multiple small RNA biogenesis factors with partly redundant functions, such as Dicer like proteins DCL 2-4 thus making it difficult to investigate the influence of specific factors on small RNA processing and potentially DNA repair.

For cultured human cells, reduced HR repair rates have been reported upon knockdown of DICER, DROSHA and AGO2 (Wei et al. 2012; Gao et al. 2014) and the repair-associated γ-H2Ax foci persist longer in absence of break-derived small RNAs (Francia et al. 2012). The small RNAs are not required for initial recruitment of the MRN-complex to the DSB, but are necessary for the further amplification of the damage response (Francia et al. 2016). Furthermore, the biosynthesis of long RNAs acting as precursor and binding partner for the break-derived small RNAs is performed by RNA polymerase II potentially using the DNA at the break as template for RNA synthesis in both directions (Michelini et al. 2017). The mechanism by which the small RNAs contribute to amplifying the DDR signal and thus to DNA repair is still unclear. The same is true for the question what triggers the antisense transcription and thus the creation of small RNAs in response to DSBs. As the same effector proteins are involved in both miRNA and break-derived small RNA synthesis in human, any deletion of these effector proteins could affect either pathway and thus indirectly affect DNA repair.

In *D. melanogaster*, the break-derived small RNAs induced by transfection of a linearized reporter plasmid thus imitating a DNA double-strand break are classical endo-siRNAs processed by Dcr-2 and loaded into Ago2 (Michalik, Bottcher, and Forstemann 2012). They cover the reporter region between the transcription start site and the restriction site of the linearized plasmid and induce the repression of a reporter signal *in trans*. As shown in both, *Drosophila* cells and flies, DSB repair is unaffected in *dcr-2* mutants who cannot generate endo-siRNAs (but have a functional miRNA biogenesis pathway) (Schmidts et al. 2016). This implies that in *D. melanogaster,* the break-derived endo-siRNAs have no impact on DNA repair but rather function in regulation of gene expression or transcript degradation.

In contrast to *N. crassa* and *A. thaliana*, there is no RNA-dependent RNA polymerase in *D. melanogaster* which could use an mRNA transcript as template for antisense transcription at the double-strand break to enable endo-siRNA generation. Experiments in our group performed both in parallel and after the here presented work (Böttcher et al. 2021) identify the DNA-dependent RNA polymerase II as the protein performing the antisense transcription at the DNA double-strand break.

1.4.4 A genome-wide dual-luciferase RNAi screen for factors involved in the small RNA response to DNA double-strand breaks

A genome-wide RNAi screen performed and presented in my Master thesis (Merk (2013); data published in (Merk et al. 2017)) gives a first overview on possible facilitators (and antagonists) of double-strand break induced siRNA formation in *D. melanogaster.* The screen in *Drosophila* S2 cells was performed by combining a genome-wide RNAi approach with a reporter assay as shown in Figure 1-2. In this reporter assay, a linearized plasmid carrying an inverted *Renilla* luciferase (Rluc) sequence is co-transfected with a circular full-length *Renilla* luciferase expression vector as reporter. A third plasmid carrying a firefly luciferase (Fluc) reporter gene is used to monitor transfection efficiency and cell viability. Both the *Renilla* and firefly luciferase reporters contain an intron in the 5' UTR preceding the CDS. The linearized

plasmid, imitating a DNA double-strand break, gives rise to endo-siRNAs, which then repress the *Renilla* luciferase reporter *in trans* (Michalik, Bottcher, and Forstemann 2012). A de-repression of the reporter (increased Rluc signal) upon knockdown of a specific gene with respect to the unresponsive majority of factors or defined negative controls indicates a role of this factor in DSB-induced siRNA production (positive hits). In contrast, the knockdown of factors that prevent the siRNA-response leads to even stronger repression of the Rluc signal (negative hits).

The performance and the dynamic range of the assay in the genome-wide screen are evaluated via the included controls as shown in the Master thesis and publication (Merk 2013; Merk et al. 2017). A strong reporter response is observed for the knockdown of the essential RNAi factors Ago2 and Dcr-2, whereas the negative controls dsGFP and dsRed influence neither Fluc nor Rluc expression. With respect to factors preventing siRNA generation, the effect of a direct Rluc knockdown is much weaker compared to the knockdown of Ago2 or Dcr-2. However, "negative hits" can also be identified. Furthermore, dsAgo1 (a key player of the microRNA pathway) and dsThread (and apoptosis inhibitor) are used as controls to identify and distinguish between cells with a reporter response despite decreased viability (comparable to Ago1) and dead cells (as observed upon Thread knockdown).

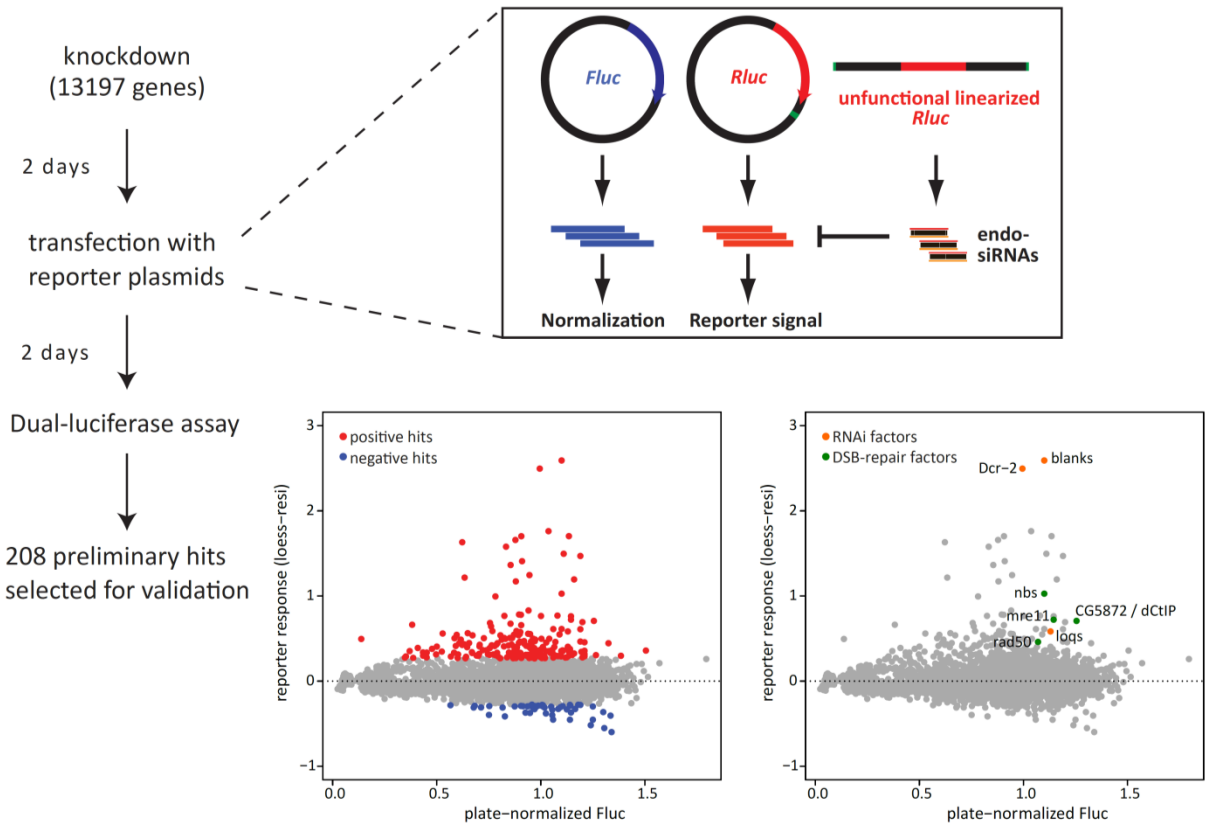

Figure 1-2: Overview on genome-wide RNAi screen for factors which facilitate or reduce DSB-induced siRNA formation in *Drosophila* cells: reporter, workflow and initial (unvalidated) results.
For the genome-wide screen, 13197 different genes in *Drosophila* DMel cells were subjected to RNAi (Day 1). Two days after knockdown, the cells were transfected with a mix of pRB1 (*Renilla*-luciferase reporter plasmid), pRB2 (firefly-luciferase normalization plasmid) and linearized pRB3 (plasmid with inverse *Renilla*-reporter digested with EcoRI-HF generating an siRNA response). *Renilla* and firefly luciferase activities were consecutively measured on Day 5 and normalized on the plate median. LOESS-residuals were calculated from the mean normalized Rluc values of two biological replicates for each target

gene and plotted against the corresponding plate-normalized Fluc signal thus quantifying the reporter response. Using these LOESS-residuals, a z-score was calculated. LOESS-residuals vs. normalized Fluc signal from the genome-wide screen are depicted for non-control samples only. The individual target genes are colored according to the LOESS-residual-based z-scores with positive hits (z-score ≥4) in red, negative hits (z-score ≤-4) in blue and non-responsive samples in grey. 208 preliminary hits were selected for validation. Known RNAi factors (indicated in orange) and components of the DSB-repair machinery (indicated in green), are among the positive hits causing a de-repression of the siRNA-reporter. **Data obtained in Master thesis (Merk 2013), published in (Merk et al. 2017).**

The reporter response, depicted as calculated LOESS-residuals (loess-resi) with respect to the Fluc values as described in the Master thesis (Merk 2013), is shown in Figure 1-2 for each knockdown-construct. Factors with robust z-scores ≥4 (positive) and ≤-4 (negative) calculated from these LOESS-residuals (Birmingham et al. 2009), after removal of retracted gene models and genes not expressed in S2 cells, are considered preliminary hits and the target genes including the corresponding z-scores are listed in Appendix 2: Detailed results from validation screens. The 142 positive and 66 negative preliminary hits are colored in red/blue in Figure 1-2 in contrast to the majority of non-responsive factors shown in grey. Among these initial positive hits are known components of the endo-siRNA pathway (Dcr-2, loquacious) and other factors connected to RNAi such as blanks, as well as the MRN complex (nbs, rad50, mre11) and further double-strand break repair factors (CG5872/CtIP) as indicated in Figure 1-2.

1.5 Aim of the study

DNA double-strand breaks induce the formation of small non-coding RNAs at the site of damage in various organisms such as *Neurospora crassa, Arabidopsis thaliana*, human cells, rice and *Drosophila melanogaster*. While some hypothesis regarding the proteins involved in the biogenesis of the small RNAs and the impact of these small RNAs on DNA repair have been investigated since the discovery of these endogenous small RNAs, the mechanism of how the DNA damage is transferred into a functional endo-siRNA response remained still unclear.

The aim of this project was thus to gain further insight into the processes occurring at the DNA double-strand break and come up with a mechanism of how the presence of a DNA double-strand break triggers the formation of break-induced endo-siRNAs.

- As a first step towards identifying contributing and counteracting factors and processes, I validated the preliminary hits from the genome-wide RNAi screen for factors involved in the small RNA response to DNA double-strand breaks with independently designed dsRNAs.
- I further narrowed down their function as either specific for DSB-induced siRNA formation or as part of a more general defense reaction to diverse triggers including high-copy inserts by applying the validation dsRNA constructs to different reporter cell lines.
- Next I investigated what splicing steps the obtained splicing factors indicate as prerequisite for endo-siRNA formation and checked whether a selection of factors impact the protein level of key siRNA effector proteins.

- To confirm that not only the plasmid-based reporter but also actual double-strand breaks in the genome trigger the transcription- and splicing-dependent endo-siRNA biogenesis, I used the CRISPR-Cas9 system to introduce cuts at defined positions with regard to the intron-exon structure and different transcription levels of genomic loci and analyzed the resulting small RNA response via small RNA deep-sequencing. I also investigated the highly abundant small RNA reads mapping to the CRISPR U6-sgRNA.
- Furthermore, I examined the small RNA distribution along the intronic and exonic regions of the targeted gene to see whether the unspliced or spliced transcript is processed into small RNAs. To monitor the splicing efficiency upon knockdown of selected splicing factors, I performed spliced/unspliced loci-specific qRT-PCR on selected genes.
- I also attempted to investigate the changed siRNA-response at the CRISPR-induced break upon introduction of an (endogenous) intron into an intronless gene. However, I instead discovered that the introduction of an intron into an intronless gene per se triggers a profound endo-siRNA response.
- Further investigations focused on the lammer kinase Doa which led to a strong de-repression of both the DSB-reporter and the high-copy reporter in the RNAi experiments. Unfortunately, neither the obtained results from knockdown and deep-sequencing nor the co-immunoprecipitation experiments with a potential interaction partner yielded any conclusive results or starting point for further investigations regarding the role of Doa in endo-siRNA formation.

Taken together, in the here presented project I used a variety of different methods to investigate what factors are involved in the endo-siRNA formation in response to DNA double-strand breaks (and beyond) and specifically worked on identifying the mechanism by which the spliceosome translates DNA damage into a trigger for endo-siRNA formation.

2 Results

2.1 Which factors or processes promote siRNA production at DNA double-strand breaks?

DNA double-strand breaks lead to small RNA production in various organisms such as *Neurospora crassa*, *Arabidopsis thaliana*, vertebrates and cultured *Drosophila melanogaster* cells. While a mechanism involving RNA pol IV and an RNA-dependent RNA polymerase (RdRP) has been proposed for DSB-induced dsRNA and thus small RNA production in *N.crassa*, it still remains elusive how these small double-stranded RNAs are synthesized in *D. melanogaster* or vertebrates which lack both RNA pol IV and RdRP. It is moreover not known, how the presence of a DNA double-strand break is transferred into a signal for small RNA production.

To identify factors or processes which promote or inhibit the siRNA formation at DNA double-strand breaks, I performed validation screens for the 142 positive and 66 negative preliminary hits of the genome-wide screen (results from Master thesis, see Figure 1-2 and list in Appendix 2). Validation was carried out in two consecutive steps. First, I re-screened the selected preliminary hits with dsRNA constructs equivalent to those of the genome-wide HD2 screening library (validation I) for a third biological replicate. The validation experiments were performed in similar Drosophila S2 cells using linearized pRB4 (truncated *Renilla* luciferase instead of the inverted *Rluc* insert used for the genome-wide screen) as endo-siRNA trigger. Furthermore, I designed two independent, non-overlapping dsRNAs for each candidate (DesignA and DesignB) and repeated the validation screening of the selected candidates with each of those designs (validation II). Those candidates for which the de-repression (positive hits) or further repression (negative hits) observed during the genome-wide screen could be reproduced with at least two out of three dsRNAs (design from genome-wide screen and independent designs A and B) were considered validated (unless excluded for other reasons), leading to a set of 88 factors that promote DSB-induced siRNA formation and 37 candidates for the repression of DSB-derived siRNA generation. The comprehensive data demonstrating the validation procedure is summarized in Appendix 2: Detailed results from validation screens.

Figure 2-1 shows that for the majority of the initially positive hits, the knockdown with the initial dsRNA construct (validation I) results in a clear de-repression of the reporter in comparison to the negative controls thus validating the screening-results. Correspondingly, the knockdown of most of the initially negative hits again causes further repression of the reporter plasmid. However, a number of positive and negative hits show an Rluc/Fluc dependence comparable to the negative controls and can thus not be validated with the original dsRNA, as it can be expected for the third “replicate” of a primary screen.

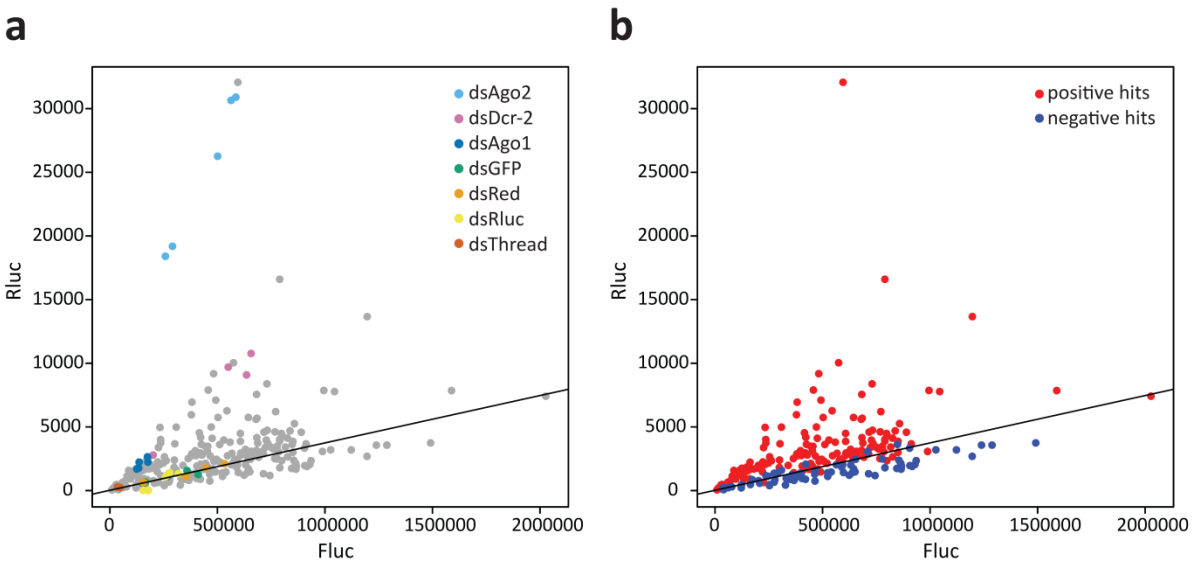

Figure 2-1: Validation I with original dsRNA constructs confirms results from genome-wide screen. *Drosophila* S2 cells were treated with dsRNAs constructs identical to those of the HD2-screening library targeting the preliminary positive and negative hits selected for validation. Rluc and Fluc values of the siRNA-reporter (pRB1, pRB2, linearized pRB4) were measured 2 days after transfection. The solid line illustrates the Rluc vs. Fluc dependence of the non-responsive negative controls dsRed and dsGFP depicted in orange and green. a) Positive and negative controls were included in the validation screen and exhibit the expected reporter response. b) Non-control samples are depicted and colored according to the initial reporter response in the genome-wide screen in red (preliminary positive hits) and blue (preliminary negative hits).

The second validation step with two additional, independent dsRNAs for each of the initially selected 208 preliminary candidates is shown in Figure 2-2. Upon averaging both designs and two replicates each, the data including the positive and negative controls shows the expected de-repression and repression comparable to the screening data. In contrast to the genome-wide screen, the vast majority of non-responsive factors are not included anymore (see Figure 2-2a), thus rendering a different normalization method necessary. As shown in Figure 2-2b, the Rluc/Fluc dependence of the included negative controls dsRed and dsGFP, visualized by the regression line, correlates well with cells without dsRNA treatment, hence substantiating the control-based normalization. In order to compare the different dsRNA designs and to acquire a quantity describing the reporter response, Rluc/Fluc ratios were calculated. However, in addition to positive hits, dead cells are also characterized by high Rluc/Fluc ratios. This can be seen from both the dsThread data points indicated in Figure 2-2c and the comparison of this plot to Figure 2-2d where a viability threshold was applied. Thus, "dead cells" can not directly be distinguished from positive hits leading to a strong de-repression of the reporter. To avoid the exclusion of potentially interesting candidates through a strict viability threshold, no such threshold was applied. Instead, the graphical analysis was restricted to the interesting dynamic range indicated by Figure 2-2d, and dsRNA constructs with a strong viability effect (dead cells) were subsequently marked in the analyzed dataset.

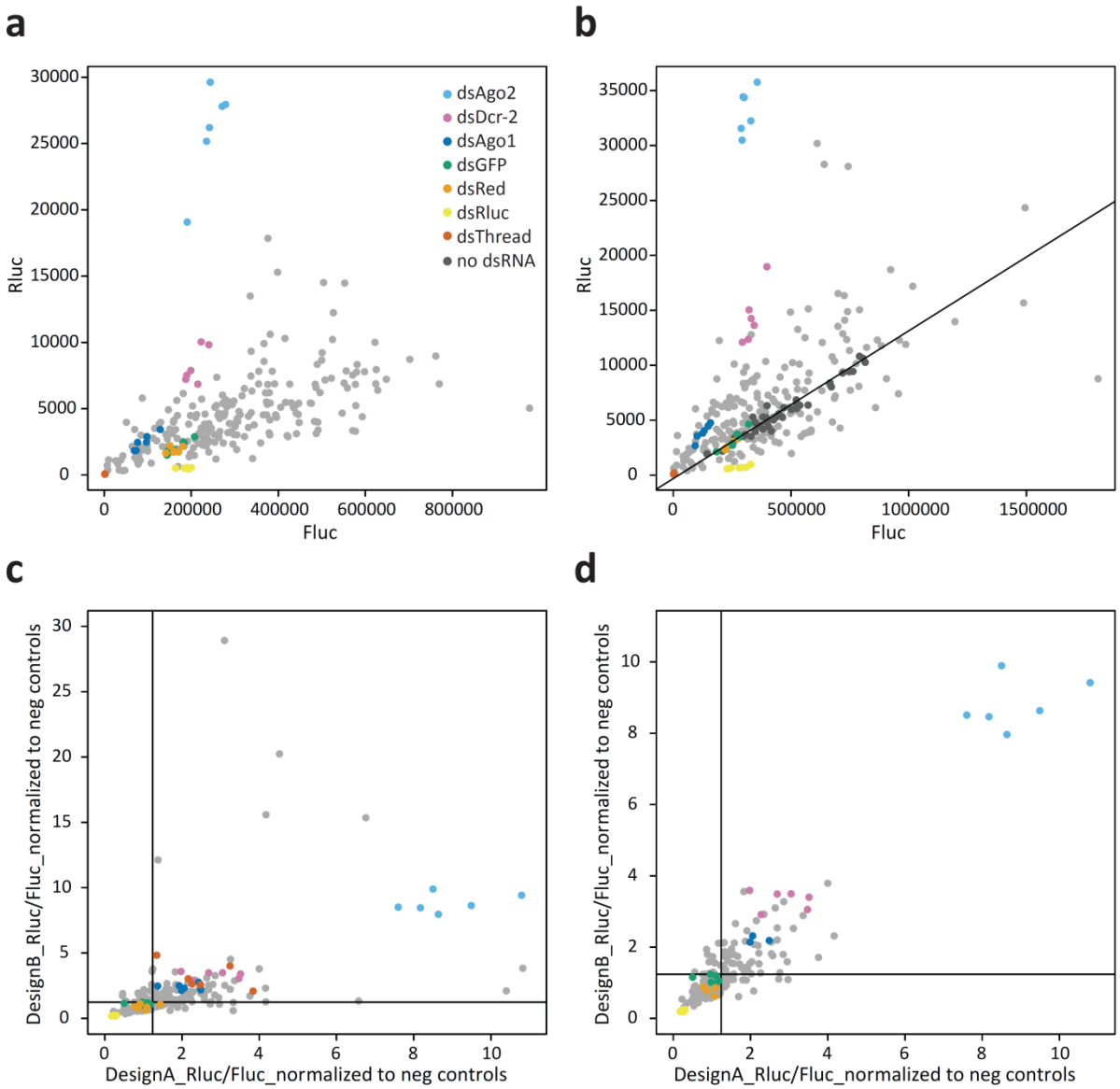

Figure 2-2: Results from validation II were normalized on negative controls and require monitoring of viability effects. *Drosophila* S2 cells were treated with dsRNA DesignA or DesignB (2 replicates each) and transfected with pRB1, pRB2 and linearized pRB4 reporter plasmids. a) Rluc vs. Fluc values are plotted as means of DesignA and DesignB with two replicates each prior to normalization. The included controls display the expected influence on the reporter signal and a clear Rluc/Fluc dependence is visible. b) Rluc vs. Fluc values are depicted for replicate 1 of DesignB. The solid line shows the correlation of Rluc and Fluc values for the negative controls dsGFP and dsRed marked in green and orange. The non-treated cells in black exhibit the same dependence as the negative controls. c) Rluc/Fluc ratios of the obtained data were calculated and normalized on the negative controls of the respective replicate. The replicates for both dsRNA designs were averaged separately and the resulting values are depicted without a viability cutoff. The solid lines indicate the validation thresholds of each dsRNA, values above this threshold were considered positive hits in validation II. The factors above this threshold include the viability controls dsThread colored in red. d) Viability cutoffs (Fluc ≥ 100000 or Rluc ≥ 4000 for replicate1 and Fluc ≥ 50000 or Rluc ≥ 1500 for replicate2) were applied. Rluc/Fluc ratios averaged across both replicates are displayed for DesignB vs. DesignA.

The reporter response induced by the two independent dsRNA designs A and B is compared to the results of the original screen indicated by color in Figure 2-3a, showing that besides a number of factors which could not be validated with either design, a clear de-repression above the indicated validation cutoff or further repression of the reporter upon knockdown of a certain factor can mostly be observed for both dsRNAs. In addition, a small number of initial hits can be validated with one but not the other design. As described above, these results obtained with designs A and B were then compared to those from

validation I resulting in the validated hit-list presented in Appendix 2: Detailed results from validation screens and published in (Merk et al. 2017).

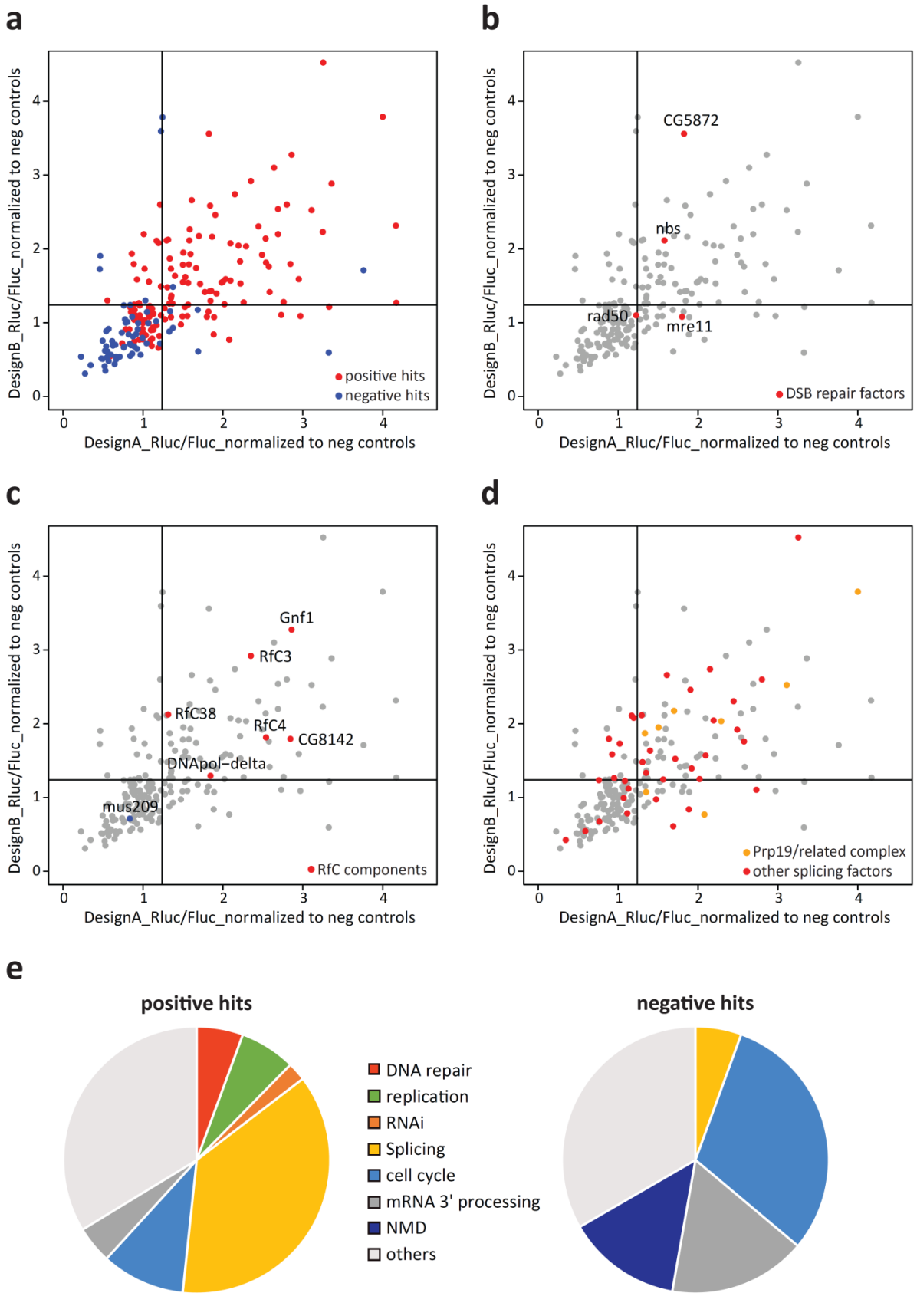

Figure 2-3: Specific complexes involved in DSB-repair, as well as numerous splicing factors promote siRNA formation at DNA double-strand breaks and could be validated with independent dsRNAs. Rluc/Fluc of the DSB-induced siRNA reporter (pRB1, pRB2, linearized pRB4) were calculated for each target upon knockdown with either of the two independent dsRNA designs A and B and normalized on the negative controls dsRed and dsGFP. The mean values of two biological replicates without any general viability cutoff are shown for both designs and each non-control target. All graphs are restricted to the relevant range indicated by viability considerations. The respective validation cutoffs for positive hits ($1+\sigma_{neg\ controls}$) are displayed as solid lines. a) Targets are colored according to their reporter influence in the genome-wide screen in red (initially positive hits) and blue (initially negative hits). b) Components of the MRN double-strand break repair complex are indicated in red. c) All subunits of the clamp loader RfC as well as the target DNA-polymerase δ are among the validated positive hits and indicated in red, while the sliding clamp itself is marked in blue. d) Factors associated with splicing (red, inferred from GO-term analysis), and in particular components of the Prp19/-related complex (orange), are marked within the validation data. e) Pie-charts of selected enriched GO-terms based on gene ontology analysis of the positive and negative hits validated with at least 2 out of 3 independent dsRNAs (analysis performed with GOrilla). **Adapted from (Merk et al. 2017).**

The gene ontology analysis of the validated candidates shown in Figure 2-3e provides a first overview of biological processes involved in DSB-induced siRNA formation. As expected, known RNAi factors, as well as certain DNA repair and a specific set of replication factors (GO-term enrichment of "DNA metabolic process" with a p-value of 4.25×10^{-7}) are among the validated positive hits. Surprisingly, 37 % of all genes promoting the siRNA formation at the break are connected to splicing, although only ~1 % of the initially screened target genes are associated with this GO-term (GO-term "RNA splicing", p-value 7.06×10^{-28}). Processes that repress DSB-induced siRNA production include 3' end processing activities such as mRNA cleavage and polyadenylation ("mRNA 3' end processing" 1.96×10^{-12}) and mRNA quality control mechanisms ("nonsense-mediated decay" 1.11×10^{-7}). Correct transcript termination or removal of potentially compromised transcripts therefore seems to diminish the triggers for siRNA generation.

2.1.1 Processing of DNA ends stimulates and precedes siRNA generation

To initiate DNA double-strand break repair, the ends of the broken strands are recognized and bound by the Mre11-Rad50-Nbs1 (MRN) complex. All three components of this complex are among the initial positive hits and while the de-repression upon knockdown of rad50 is not strong enough to confirm the screening result, both, mre11 and nbs, could be validated with independent dsRNAs as shown in Figure 2-3b. Besides DNA damage signaling following the recognition of the breaks, the MRN complex also facilitates 5' to 3' resection of the ends thus promoting homology-directed repair. As a next step, the 3' overhang is further extended by CtIP/Sae2 (Panier and Durocher 2013), which then triggers the homologous recombination repair pathway. CG5872, most likely the *Drosophila* homologue of CtIP as indicated by bioinformatics analysis and observed function (Uanschou et al. 2007), was among the strong initial candidates from the genome-wide screen and could clearly be validated as a positive hit (Figure 2-3b). During homologous recombination repair, the generated 3' single-stranded section anneals with a homology template in a Rad51-mediated manner and DNA synthesis is then performed by the replicative DNA polymerases δ and ε. In addition to DNA-polδ itself, all subunits of the replication factor C (RfC) clearly promote siRNA formation at DSBs as shown in Figure 2-3c. During replication, RfC facilitates the loading of the sliding clamp mus209, the *Drosophila* homologue of PCNA (processivity clamp proliferating cell nuclear antigen) onto the DNA thus enabling long-range DNA synthesis during replication. In contrast to the clamp loader RfC, the processivity factor PCNA itself, which can be

diversely modified and plays an important regulatory role during DNA repair, was initially among the negative hits, but could not be confirmed through the validation process.

Taken together, the presence of various factors associated with different (early) steps of homology-directed repair combined with the absence of NHEJ factors such as the Ku70/Ku80 complex suggest that the initiation of homology-directed repair promotes the production of DSB-induced siRNAs. Furthermore, the results demonstrate that the recognition and initial processing of the double-strand break clearly precedes siRNA formation.

2.1.2 Splicing factors stimulate siRNA production

While DNA repair factors are somewhat expected candidates for a process initiating at a double-strand break, the striking overrepresentation of splicing components within the positive hits is rather unanticipated. However, many splicing factors are recovered in the screen as depicted in Figure 2-3d, thus suggesting a contribution of the mRNA splicing machinery or process to the formation of DSB-induced siRNAs. The initially identified splicing factors thereby cover a wide quantitative range of reporter de-repression and show a validation rate comparable to other hits. Among the splicing factors, components of the Prp19/-related complex are particularly abundant within the detected hits as indicated in Figure 2-3d.

2.1.3 Surveillance mechanism possibly serves in genome defense beyond DSBs

Similarly to DNA double-strand breaks, transposons can be triggers for small RNA formation. During transposon recognition in the fungus *C. neoformans*, stalled spliceosomes are involved in triggering small RNA production (Dumesic et al. 2013). To investigate whether the observed involvement of splicing factors in siRNA formation in *D. melanogaster* is limited to DSB-induced small RNAs or rather a common feature shared with other genome defense mechanisms such as transposon recognition, the initial candidates were tested for their ability to promote siRNA-mediated repression of a genomically integrated high-copy transgene. As depicted in Figure 2-4a, a cell-line with a genomically integrated high-copy *Renilla* luciferase leading to a strong siRNA response thus imitating a transposon situation, and a low-copy firefly luciferase which does not induce siRNA formation for normalization, was used for screening. This modified version of the previously described high-copy endo-siRNA reporter (Hartig et al. 2009) allowed a direct comparison to the analogous DSB-reporter since both systems are based on the same reporter genes. The results summarized in Figure 2-4b and further shown in Appendix 2 indicate that splicing factors as well as a number of other validated candidates from the screen are involved in both, the siRNA production in response to DNA double-strand breaks and high-copy loci such as transposons since the respective knockdown leads to a de-repression of both reporter systems. However, another group of candidates including the detected DNA repair factors (MRN-complex, CtIP) and the RfC complex are specifically involved in the siRNA formation at DSBs and the respective knockdowns have no effect on

the high-copy reporter. Hence, the particular mechanisms leading to small RNA formation in response to DNA double-strand breaks or high-copy transgenes appear to be specific for the source of genomic instability. However, perturbed mRNA splicing may serve as a common trigger for endo-siRNA biogenesis in both genome defense mechanisms.

To verify the results obtained with the dual-luciferase high-copy reporter imitating a transposon situation, the analogous GFP-based high-copy reporter (Hartig et al. 2009) was used to re-screen the candidates. The comparison of both reporters in Figure 2-4c points out that the dual-luciferase reporter is more sensitive than its GFP-based counterpart, as the dynamic range of this reporter is higher and numerous knockdowns lead to a detectable de-repression of the dual-luciferase but not the GFP reporter. In contrast, only few knockdowns selectively de-repress the GFP reporter. For example, the knockdown of *pavarotti (pav)*, supposedly a microtubule motor protein involved in mitosis, causes the formation of large cells. While the normalization on firefly luciferase in the dual-luciferase reporter system can at least partly account for that effect and prevent a disproportionally high reporter signal, the increased size of the cells leads to an accordingly high GFP-signal and thus false positive result of the GFP-based reporter in the FACS readout. However, apart from these technically induced differences between the reporters, several knockdowns trigger a detectable de-repression of both high-copy reporters. Next to blanks, a protein known to play a role in RNAi (Gerbasi et al. 2011; Nitschko et al. 2020), several validated positive candidates involved in splicing such as l(1)10Bb, Hrb27C, CG10754, CG7185 are among those causing the strongest effect on both reporters, underlining the possible role of splicing factors in transposon-induced siRNA formation.

In order to identify factors for which the RNAi knockdown strongly influences the reporter signal independent from siRNA production, the candidates were also counter-screened with a GFP-based low-copy reporter cell line which does not induce endo-siRNA production. As shown in Figure 2-4d, a number of candidates, mostly factors connected to mitosis such as pav, ial, Incenp, feo and Bub-R1, indeed lead to an increased GFP signal upon knockdown. Interestingly, in addition to these proteins connected to mitosis, the knockdown of the ATP-dependent RNA helicase belle, which was recovered and validated as a negative hit thus preventing DSB-induced siRNA formation, induced an increased GFP signal of the low-copy reporter. This factor has been described as an RNAi pathway factor (Ulvila et al. 2006) and can supposedly interact with both Ago2 and Ago1 in an RNA-dependent manner. As it was shown to promote RNA interference without obviously changing siRNA levels, it was considered to act downstream of dsRNA processing and loading (Zhou et al. 2008). The screening results thus also suggest that belle has an impact on DSB-induced siRNA formation; however, its role remains to be investigated as the knockdown seems to contrarily affect different reporter setups.

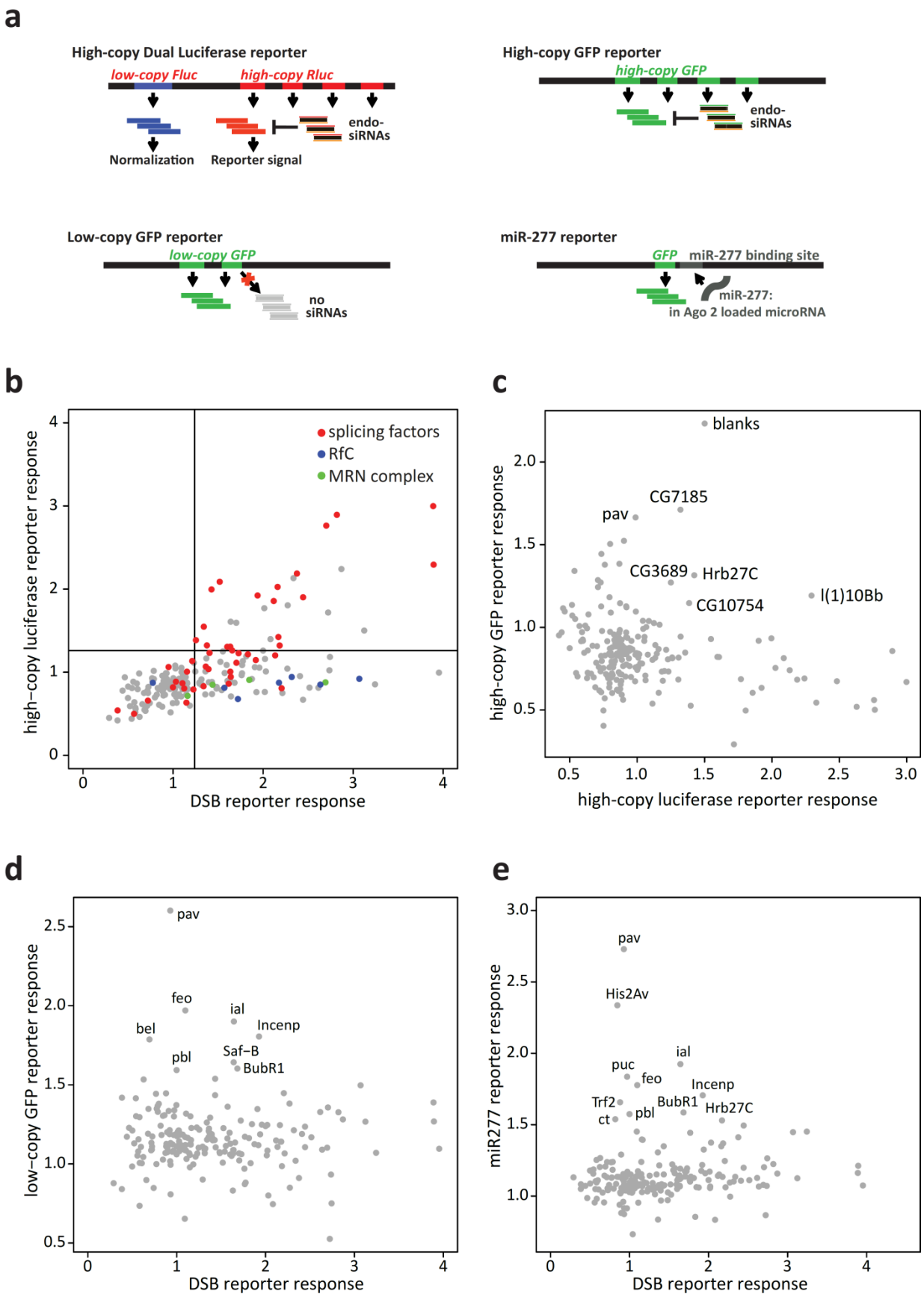

Figure 2-4: Validated hits from genome-wide screen with DSB-induced siRNA reporter have different effects on high-copy- , low-copy-, and miR277-based reporter cells. The initial positive and negative candidates from the genome-wide screen were further investigated by applying the independent dsRNAs to 4 additional reporter cell lines. All reporters were subjected to

the respective knockdowns, and luminescence (high-copy dual luciferase reporter) or fluorescence (GFP-based reporters) was measured 4 days after knockdown. a) Schematic functionality of the utilized reporter cell lines. b) Direct comparison of the reporter response of the dual luciferase high-copy reporter (Klon4) and DSB-reporters. For both reporters the Rluc/Fluc ratios of 4 replicates (2 biological replicates each for dsRNA designs A and B) were normalized on the included negative controls dsRed and dsGFP and averaged. The solid lines mark the respective validation thresholds ($1+\sigma_{neg\ controls}$; 1.24 for DSB and 1.26 for high-copy reporter). Splicing factors are marked in red, DNA repair factors in green and the RfC complex including PCNA and DNA-polymerase δ in blue. c) Direct comparison of luciferase- and GFP-based high-copy reporters. The mean fluorescence of 10000 GFP-based high-copy reporter cells (63N1) was detected in the FL1 channel and normalized on the average values of the negative controls dsRed and dsRluc. For both reporters a total of 4 replicates including two independent dsRNA knockdowns were averaged to obtain the depicted results. d) Comparison of a GFP-based low-copy reporter (63-33) with the DSB-induced siRNA reporter. The low-copy reporter was measured and normalized in analogy to the GFP-based high-copy reporter in c). e) Comparison of the GFP-based miR-277-reporter, carrying binding sites for the Ago2-loaded microRNA in the 3' UTR of the reporter gene, and the DSB-reporter. Measurements and normalization of the miR-277 reporter were performed as explained for the 63N1-reporter in c). **Part b) published in (Merk et al. 2017).**

In addition to double-strand break, high-copy and low copy reporters, the initial hits were also screened with a cell line containing a GFP reporter with two perfect matches to miR-277 in its 3'-UTR (Forstemann et al. 2007). This microRNA is processed by Dcr-1, loaded into Ago2 by Dcr-2/R2D2 and can then repress the GFP reporter. Thus the knockdowns that either non-specifically influence all reporters (the expression of both *Renilla* luciferase and GFP is driven by the same promotor in the respective reporters) or hit core RNAi factors that are necessary for miR-277 functionality, lead to detectable reporter de-repression. Figure 2-4e illustrates that the same mitosis factors which influence the supposedly non-responsive low-copy GFP reporter also lead to increased fluorescence signals for the miR-277 reporter. This finding suggests that knockdowns of these factors cause changes in the cells which influence the reporter signal, though not by specifically influencing the investigated process, thus identifying these candidates as either false positives or non-reliable hits. Therefore, technically validated candidates which show a strong response for both the low-copy GFP reporter (≥ 1.6) and the miR-277 reporter (≥ 1.6) are marked as false positives in Appendix 2: Detailed results from validation screens and not counted as validated hits (4 false positives).

In addition to those factors which influence both, the low-copy and the miR-277 reporter, some knockdowns selectively trigger a response of the miR-277 but not the low-copy reporter. In contrast to His2Av, Trf2, puc and ct, which could not be validated with the DSB reporter, the knockdown of Hrb27C, a protein involved in regulation of mRNA splicing (Hammond et al. 1997; Blanchette et al. 2009; Brooks et al. 2015) but also translation enhancement (Nelson et al. 2007) and localization-dependent translational repression (Yano et al.), leads to a de-repression of the miR-277 reporter in addition to both high-copy reporters and the DSB reporter but not the low-copy reporter. This factor might therefore play a role in different small RNA pathways. Most knockdowns including the vast majority of splicing factors, however, do not influence the miR-277 reporter. Hence, they seem to play a role specifically in siRNA formation rather than having a general influence on the key players of the core small RNA machinery.

In summary, the counter-screening with different reporters demonstrates that most of the validated hits are either specific for the production of DSB-induced siRNA (e.g. DSB repair factors) or even part of a

more general siRNA genome defense mechanism which is induced by both, DNA double-strand breaks and high-copy loci such as transposons (e.g. splicing factors).

2.2 What role do splicing factors play in the observed small RNA response?

The strong enrichment of factors connected to the process of splicing within the validated hits of the genome-wide screen indicates that either the presence of these splicing factors or the mRNA splicing process itself is important for siRNA formation in response to both, DNA double-strand breaks and high-copy loci. To further clarify the potential contribution of the splicing machinery to small RNA biogenesis, the here identified splicing factors were examined in more detail with respect to their role in the splicing process. Furthermore, the abundance of core RNAi proteins was studied upon knockdown of selected splicing factors to evaluate the possibility of an indirect effect via modulated protein levels of the functional RNAi machinery.

2.2.1 Which splicing factors are identified and where in the splicing cycle are they involved?

During the splicing process several spliceosome subcomplexes are distinguishable, which are dedicated to the different steps/reactions of the splicing cycle. To answer the question in which subcomplexes the factors identified as important for small RNA formation play a role, the validation results were mapped onto the the splicing process (mapping performed by K. Förstemann in (Merk et al. 2017)).

As illustrated in Figure 2-5a, the recovered and validated positive hits are not limited to a certain spliceosome complex but cover all the phases of the splicing process with the highest abundance within the Prp19- and Prp19-related complexes. The single splicing factor among the negative hits is tho2, a component of the Tho/Trex complex which is deposited onto the mRNA after splicing is completed.

In addition, Figure 2-5c depicts the complete ensemble of factors included in the different spliceosomal complexes and their corresponding results from the genome-wide screen. This mapping illustrates that most complexes assembled throughout the splicing process consist of factors with both, no or weak, and strong impact on the DSB-reporter upon knockdown. However, the Prp19- and Prp19-related complexes comprise the highest enrichment of validated positive hits (8 out of 16 within this complex vs. 26 out of 138 within all spliceosome components, χ^2-test: $p = 0.037$) as well as the only splicing factor leading to a de-repression of the DSB-reporter comparable to the core RNAi factor Dcr-2. This specific complex which enables the progression of the pre-catalytic spliceosome to the first catalytic step thus appears to be particularly important for the induction of the siRNA response to DNA double-strand breaks.

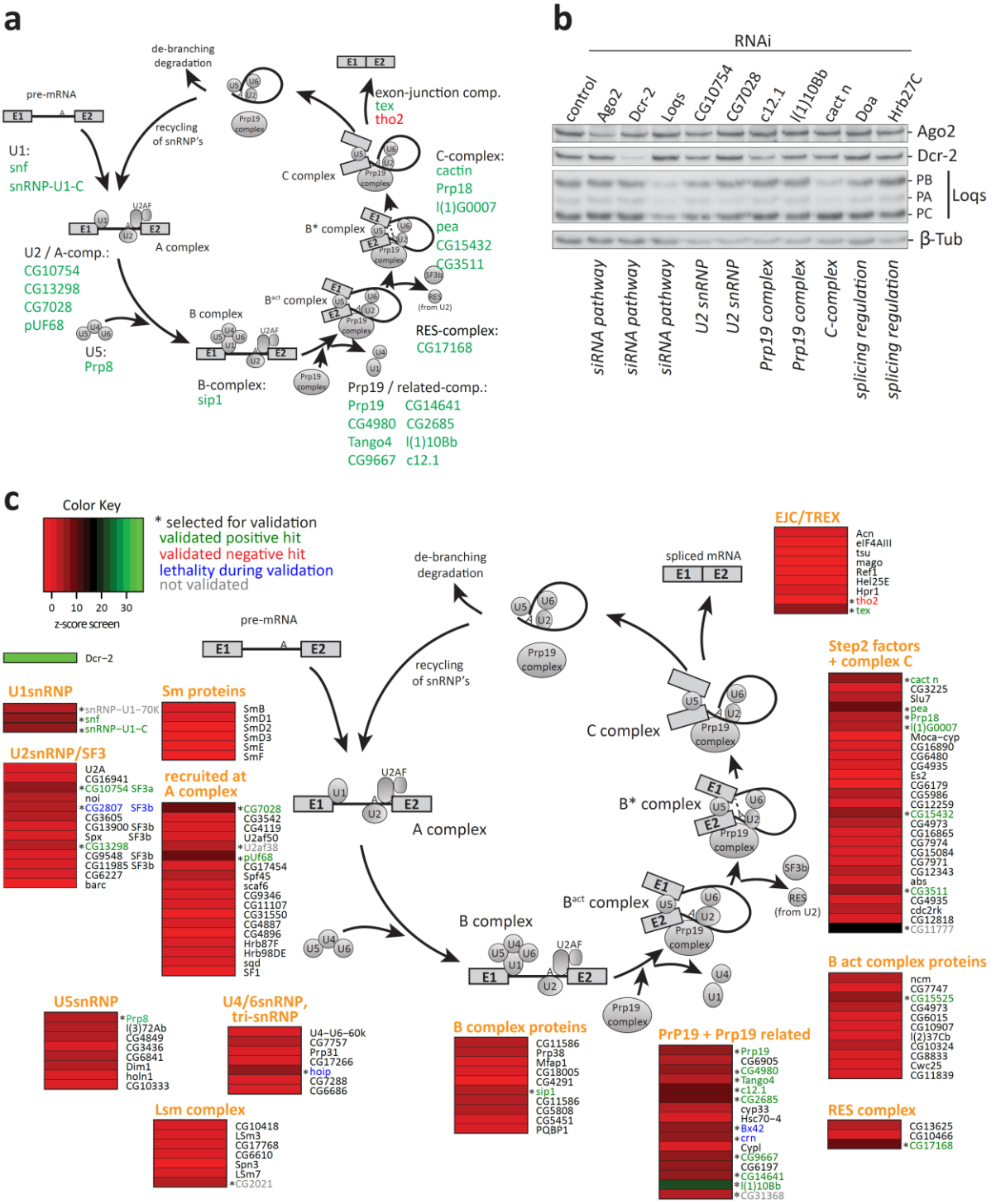

Figure 2-5: Components of the Prp19/Prp19-related complex are particularly abundant among the splicing factors recovered and validated from the genome-wide screen and the knockdowns of selected splicing factors do not compromise the protein levels of core RNAi components. a) Validated positive (green) and negative (red) hits from the genome-wide screen are mapped onto the various complexes of the splicing reaction. b) Western Blot analysis shows that knockdowns of the indicated splicing factors promoting siRNA formation in the genome-wide screen do not reduce the protein levels of key components of the RNAi machinery. Knockdown conditions in S2 cells were chosen in analogy to the (validation) screen and protein extracts were prepared 4 days after knockdown. 20 µg of total protein were subsequently analyzed with the indicated antibodies including a ß-tubulin loading control. c) Comprehensive overview of the complexes assembled during the different steps of the splicing reaction and their influence on the DSB-reporter. The impact of each factor on DSB-induced siRNA formation, represented by the z-score from the genome-wide screen, is illustrated with individual heatmaps using the depicted color key. All factors are marked according to their selection for validation (*at gene symbol) and names of those selected factors are color-coded according to validation result (green: positive hit; red: negative hit; blue: lethality; grey: not validated). **Adapted from (Merk et al. 2017).**

2.2.2 Does the splicing machinery act indirectly via changes in protein levels of RNAi factors?

One possibility how splicing could influence the formation of siRNAs in response to both high-copy loci and DNA double-strand breaks is by affecting the protein levels of the core RNAi factors since they all depend on alternative splicing of the corresponding pre-mRNAs. This potential effect has previously been observed for RNAi-mediated heterochromatin formation in *S. pombe* where mis-spliced introns within key RNAi factors in splicing component mutants significantly contribute to defects in heterochromatin assembly (Kallgren et al. 2014). However, as described in the previous chapter, the mir-277 reporter, which also depends on functional Dcr-2 and Ago2 proteins, is not generally influenced by the knockdown of splicing factors, suggesting that the strong effects on both the high-copy and DSB reporter are unlikely to be explained by altered splicing efficiencies of the core RNAi components. Furthermore, for several factors from different spliceosome complexes as well as regulators of alternative splicing which display a strong, validated effect on DSB-induced siRNA formation, protein levels of Ago2, Dcr-2 and loquacious were monitored upon knockdown of the respective splicing component. The Western blot in Figure 2-5b demonstrates that no consistent protein level changes can be observed for any of these three key players of RNAi, thus indicating a direct role of splicing in the endo-siRNAs response to DSBs and transposons rather than an indirect effect due to splicing defects in RNAi components.

2.3 Can single directed DSBs in the genome induce splicing-dependent siRNA production?

As shown in the previous chapters, transfection of a linearized reporter plasmid imitating a DNA double-strand break induces a splicing-dependent endo-siRNA response in *Drosophila* cells. In order to directly observe this response and to confirm the importance of splicing for siRNA generation, DNA double-strand breaks were introduced at defined positions with respect to the intron/exon structure of selected genes by applying the CRISPR-Cas9 technology. The consequential endo-siRNA response was then investigated via small RNA deep-sequencing.

2.3.1 DSBs can be introduced at specific positions via CRISPR-Cas9 in S2 cells

Three different target genes were chosen and CRISPR-Cas9-mediated cleavage was executed at different positions within these genes relative to both, splice sites and transcription start site as schematically illustrated in Figure 2-6. Neither of those target genes affected the DSB-reporter in the genome-wide screen, nor did the knockdown of any of them cause detectable changes in cell viability. Loci-specific effects on siRNA production or cells due to the DSB-induced disruption of the gene are therefore not expected for any of the selected genes. To gain comprehensive insight into the factors which influence the strength of the siRNA response, the selected target genes differed in expression levels (*CG15098* and *Tctp* strongly expressed, CG18273 weakly expressed) and intron structure (see schemes in Figure 2-6).

The ability to introduce DNA double-strand breaks at the targeted position/locus was individually assessed for each of the transfected CRISPR single guide RNA (sgRNA) U6-PCR templates and each

experiment via the T7 endonuclease assay. The results in Figure 2-6 demonstrate that directed DSB were induced by every sgRNA design. Although an absolute quantification of cleavage efficiency is not possible, no systematic differences between the individual genes or with respect to certain cutting positions can be observed.

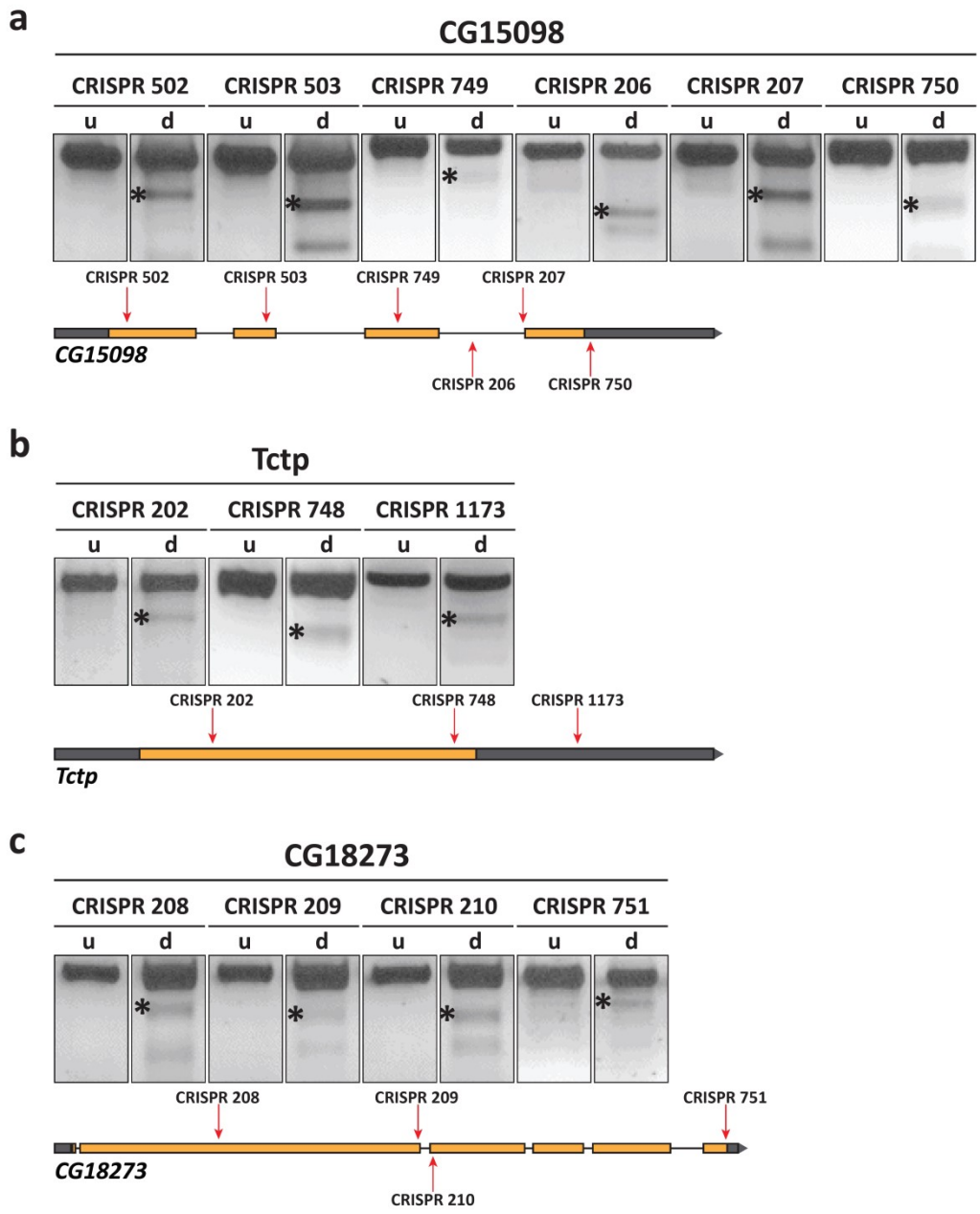

Figure 2-6: T7 endonuclease assay confirms CRISPR-Cas9-mediated introduction of double-strand breaks at various positions in selected target genes. To assess the site-specific introduction of DNA double-strand breaks in Cas9-expressing Drosophila cells (5-3 cells), genomic DNA was isolated from samples also used for preparation of deep-sequencing libraries and subjected to analysis via T7 endonuclease assay. The cartoons illustrate the target positions of the respective sgRNAs and agarose gels of the undigested (u) and digested (d) PCRs-products are depicted for each cleavage site. The largest expected cleavage product of the T7 endonuclease verifying the CRISPR-induced cleavage and mutagenesis at the investigated target position is marked with an asterisk. Cleavage success is evaluated at different positions within the target genes a) *CG15098*, b) *Tctp* and c) *CG18273*. **Adapted from (Merk et al. 2017).**

2.3.2 A small RNA response is elicited if a DSB is introduced downstream of the first intron

To confirm the induction of small RNA production observed for reporter plasmids with strong intron-containing promotors upstream of the digested reporter gene (Michalik, Bottcher, and Forstemann 2012) and to further investigate the influence of the cleavage position with respect to the intron-exon-structure of the gene on the siRNA response, the highly expressed, intron-containing gene *CG15098* was endogenously cleaved by Cas9 which was targeted with specific sgRNAs to defined cleavage positions within the gene.

Small RNAs were then sequenced and mapped back to the target locus. The resulting distribution depicted in Figure 2-7 shows that cleavage upstream of the first intron does, apart from a large number of small RNAs mapping to the cleavage position and covering the homology region of the CRISPR sgRNA, not induce small RNA production along the gene. A cut in the second exon, only 82nt downstream from the intron-exon-junction of the first short intron, leads to rather weak small RNA production barely above the background-level of the uncut sample. In contrast, cleavage further downstream results in a strong siRNA response with sense and antisense reads covering the whole region upstream of the cut and extending as far as the transcription start site (TSS). This suggests that whereas the newly synthesized transcripts could contribute the sense strand to the dsRNA precursor of the small dsRNAs, antisense transcription also needs to occur and extend all the way from the induced break to the TSS. The region downstream of the break, though, does not exhibit any detectable increase in small RNAs compared to the uncut sample. This response can be observed for all cutting positions downstream of the second exon regardless if the exact cleavage position is located within intronic or exonic regions and distant or proximal to splice sites. Even a cut in the 3' UTR, shortly downstream of the stop codon, results in small RNAs covering the >1.1 kb distance to the TSS.

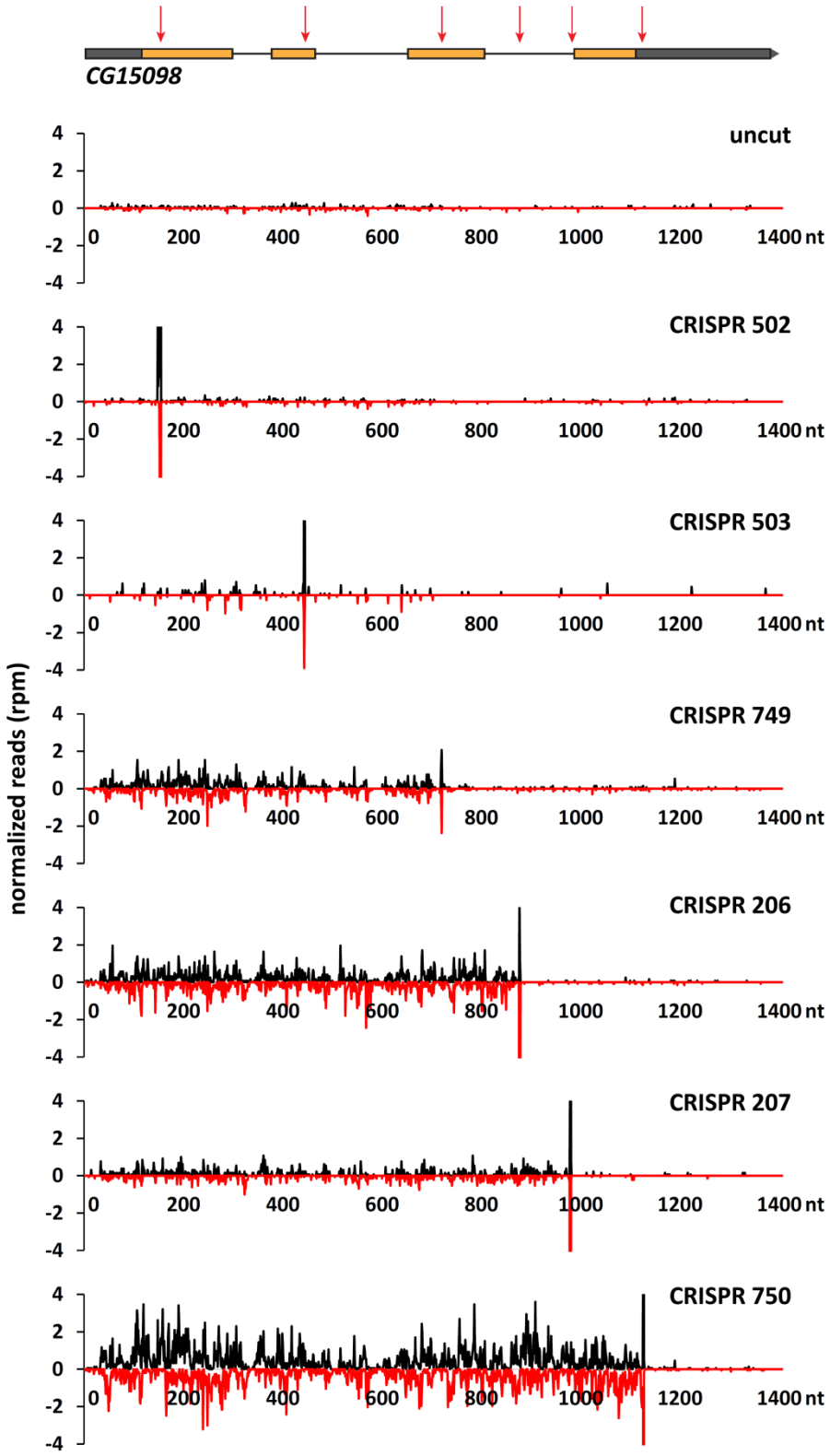

Figure 2-7: CRISPR-mediated introduction of a DSB downstream of the first intron in the highly expressed gene *CG15098* induces the formation of small RNAs mapping to the transcribed region upstream of the cleavage site. Cas9-expressing *Drosophila* cells (5-3) were transfected with U6-templates for site-specific sgRNAs. Total RNA was isolated 4 days after transfection and small RNA libraries were prepared and analyzed via deep-sequencing. The obtained sequences were selected for 19-25nt reads and mapped to the target gene. The number of reads mapping to each position of *CG15098* was determined in single-nucleotide resolution and normalized on the total number of microRNA- and transposon-mapping reads to account for the sequencing depth of the specific library. Sense (5' end of small RNAs mapped) and antisense reads (3' end of small RNAs mapped) are depicted in black and red for the individual CRISPRs and the schematic drawing at the top illustrates the different cleavage positions with respect to gene structure. **Adapted from (Merk et al. 2017).**

2.3.3 Characteristic siRNA-response is reproducibly induced by Cas9-mediated cleavage

In order to evaluate the reproducibility of the small RNA response, Figure 2-8a/b displays two biological replicates each, for different cutting positions within the target gene CG15098. Although the replicates show variations with respect to the individual quantity of the small RNA response, the qualitative observations regarding the requirement of an intron sufficiently upstream of the introduced break are entirely reproduced. Furthermore, the variations between both, replicates and cutting positions, show that as long as the intron-requirement is met there is no direct correlation between the exact cutting position in terms of both intron-exon structure and distance to the TSS, and the strength of the response.

In addition to comparable number of sense and antisense reads, as exhibited for the mapped small RNAs upstream of the induced break, functional endo-siRNAs are characterized by a typical length of 21nt. Figure 2-8c shows the length distribution of all small RNAs mapping to the region between the TSS and the cleavage position of CRISPR 206 (reads mapping to the sgRNA not included). The narrow size distribution with a clear maximum at 21nt together with the equal amounts of sense and antisense reads and the ability of Ago2 and Dcr-2 dependent small RNAs from linearized plasmids to repress a reporter *in trans* clearly suggest that the small RNAs generated at DNA double-strand breaks are indeed functional endo-siRNAs.

To determine whether the investigated endo-siRNA response is in fact induced by Cas9-mediated double-strand breaks as opposed to the sole presence of a sgRNA complementary to a certain genomic locus, the template for CRISPR 206 was transfected into S2 cells lacking the Cas9 nuclease. The graph in Figure 2-8d illustrates that in the absence of the Cas9 nuclease the presence of an sgRNA does not induce the biogenesis of the endo-siRNAs covering the transcribed region upstream of the break. The small RNAs generated in absence of Cas9 are restricted to the sgRNA sequence. Therefore, the initial trigger for the small RNA response is either the double-strand break itself or the presence of the nuclease at the cleavage site. The latter could, however, not explain the potential of a purified, linearized plasmid to produce functional siRNAs.

Taken together, genomically introduced DNA double-strand breaks lead to a reproducible endo-siRNA response if they occur downstream of a position where co-transcriptional splicing takes place.

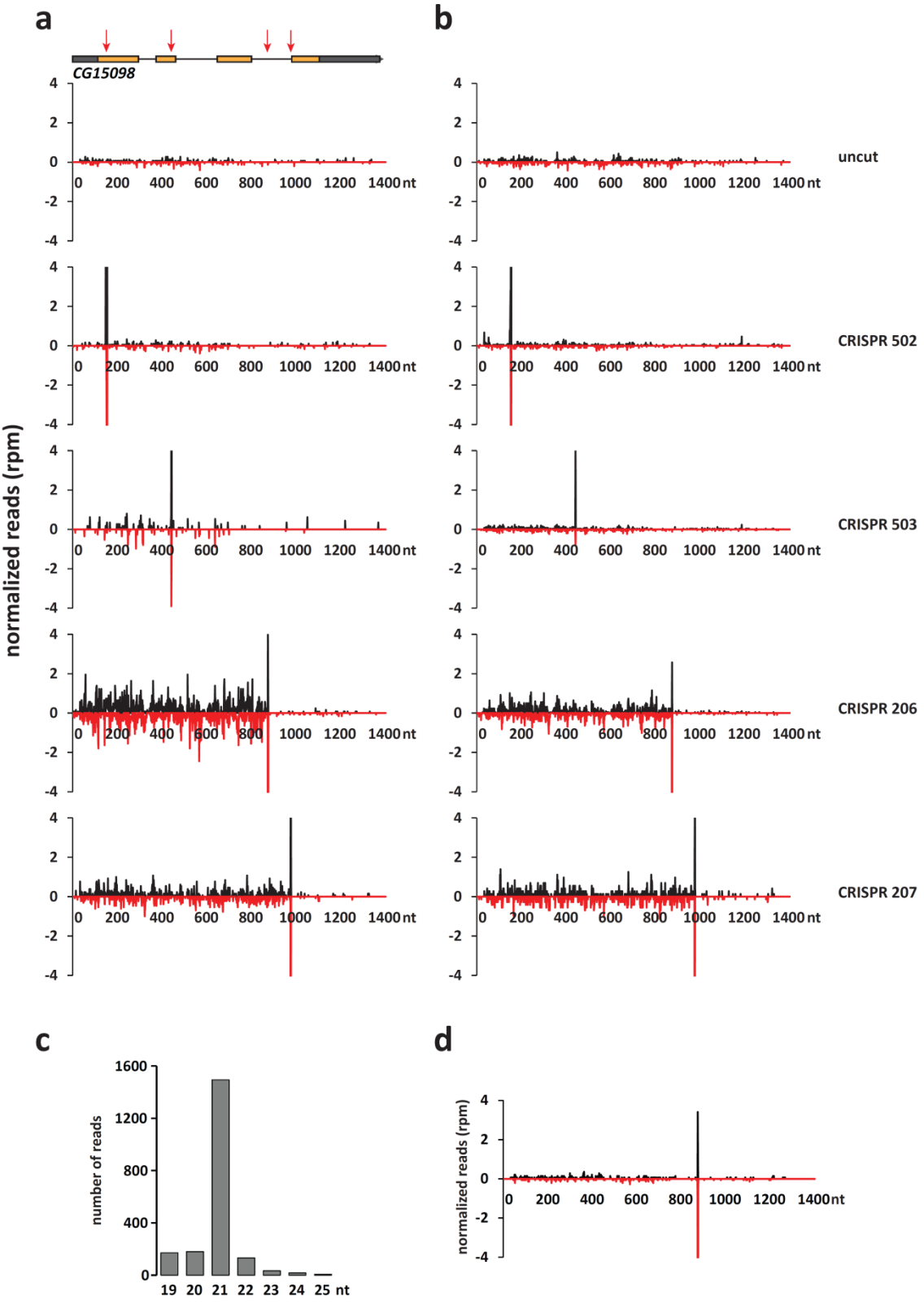

Figure 2-8: The small RNA response to CRISPR/Cas9-induced double-strand breaks is reproducible and consists of 21nt siRNAs which cannot be triggered by the transfection of a U6-sgRNA template in absence of the nuclease. Two independent biological replicates of the indicated cleavage and deep-sequencing experiments are shown in a) and b). Sense/antisense reads are colored in black/red. The traces indicate the position of the 5'-end for sense and of the 3'-end for antisense-oriented small RNAs. c) The length distribution of the sense and antisense reads mapping to the region upstream of the homology region of the sgRNA was determined for the first replicate (small RNA mapping in a)) cleaved with CRISPR 206. d) S2 cells without Cas9 expression were transfected with a U6-template for the CRISPR 206 sgRNA and analyzed in analogy to the Cas9-cleaved samples. **In parts adapted from (Merk et al. 2017).**

2.3.4 DSB-induction results in reproducible siRNA coverage pattern independent of the cleavage position

Instead of uniformly covering the region between the transcription start site and the double-strand break, the induced siRNAs seem to display an astonishingly reproducible pattern while covering both intronic and exonic regions of the targeted locus. Therefore, Figure 2-9 compares the small RNA coverage of two replicates cleaved at the same position as well as two samples cut at different positions within the same gene regarding the exact mapping position of each small RNA molecule. The direct overlays shown in Figure 2-9a and b indicate that for both, the identical and the different cleavage positions, the individual reads seem to preferably start at exactly the same positions in different samples. To confirm this finding, the different distributions were also added up for both combinations. Randomly distributed reads should then equal out each other thus resulting in a flat, smooth read coverage. Yet, the resulting distributions shown in Figure 2-9c and d illustrate that for both the equal and the different cutting positions, the patterns add up to an even more pronounced pattern thus indicating that the exact siRNA coverage is not random but highly reproducible and does not depend on the cleavage position.

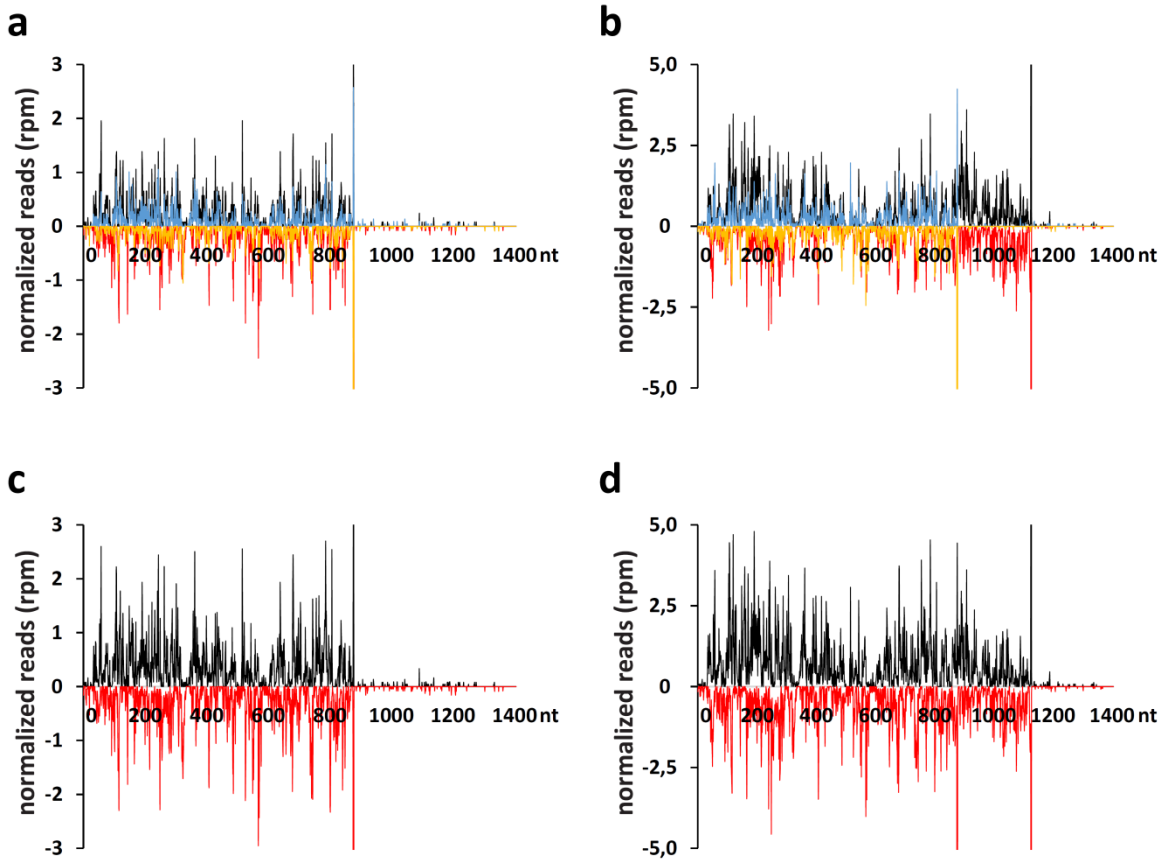

Figure 2-9: The exact siRNA coverage pattern upstream of the induced break is highly reproducible for both, biological replicates and samples with different cleavage positions. The compared libraries were mapped and normalized individually as described. The traces indicate the position of the 5'-end for sense (black/blue) and of the 3'-end for antisense-oriented small RNAs (red/yellow). a) Overlay of siRNA coverage patterns for two biological replicates of CRISPR 206. b) Overlay of siRNA coverage pattern for CRISPR 206 (blue/yellow) and CRISPR 750 (black/red). c) Normalized small RNA reads of two biological replicates, each transfected with a CRISPR 206-template, were added up to obtain the resulting siRNA coverage. d) Normalized small RNA reads of two normalized libraries one of each cut with CRISPR 206 and CRISPR 750 respectively were added up to result in the displayed siRNA coverage. **Part a) adapted from (Merk et al. 2017)**

2.3.5 CRISPR-induced DSBs at different positions do not noticeably change levels of miRNAs or transposon-mapping siRNAs in general

Beside the investigated DSB-induced endo-siRNAs, other much more abundant groups of small RNAs, such as microRNAs and transposon-repressing siRNAs, are present within the cells. As changes in their levels can immensely influence the cells and as those small RNAs are also included in the deep-sequencing libraries and frequently used for normalization, both the miRNAs and transposon-mapping small RNAs were analyzed with respect to changes induced by the induction of Cas9-mediated DSBs. Figure 2-10a shows a comparison of the individual abundance of each miRNA for different cleavage sites in *CG15098*. The correlation between different CRISPRs, different replicates and cut/uncut libraries is at least 0.99 for all samples, indicating that the genomic cleavage of *CG15098* does not induce changes in intracellular miRNAs. The same is true for transposon-mapping small RNAs as depicted in Figure 2-10c thus enabling normalization of the DSB-induced siRNA on the much more abundant microRNAs and transposon-matching small RNAs to account for the sequencing depth of the individual libraries.

The variation of small RNAs mapping to individual mRNA transcripts is much more noticeable for different cuts, as well as different biological replicates of the same cleavage position, compared to miRNAs or transposon-mapping reads. Therefore, mean normalized counts of two biological replicates were used to compare the transcript levels of different samples as depicted in Figure 2-10b. Although the correlation is not as good as for microRNAs and transposons-mapping reads, most transcript-mapping small RNAs remain unchanged upon cleavage of *CG15098*. As indicated in Figure 2-10d, the most prominent change in transcript-mapping reads can be attributed to *CG15098* itself, as both the DSB-induced siRNAs, but also the much more abundant CRISPR-homology mapping small RNAs, are only present in the cleaved sample. For other outliers either increased or decreased upon cleavage, no common single transcript and no particular shared function or location within the genome can be determined.

Taken together, the analysis of other, highly abundant small RNAs indicates, that the introduction of double-strand breaks at a required minimum distance downstream of an intron leads to the biogenesis of siRNAs mapping to the cleaved loci without having a noticeable effect on other groups of small RNAs.

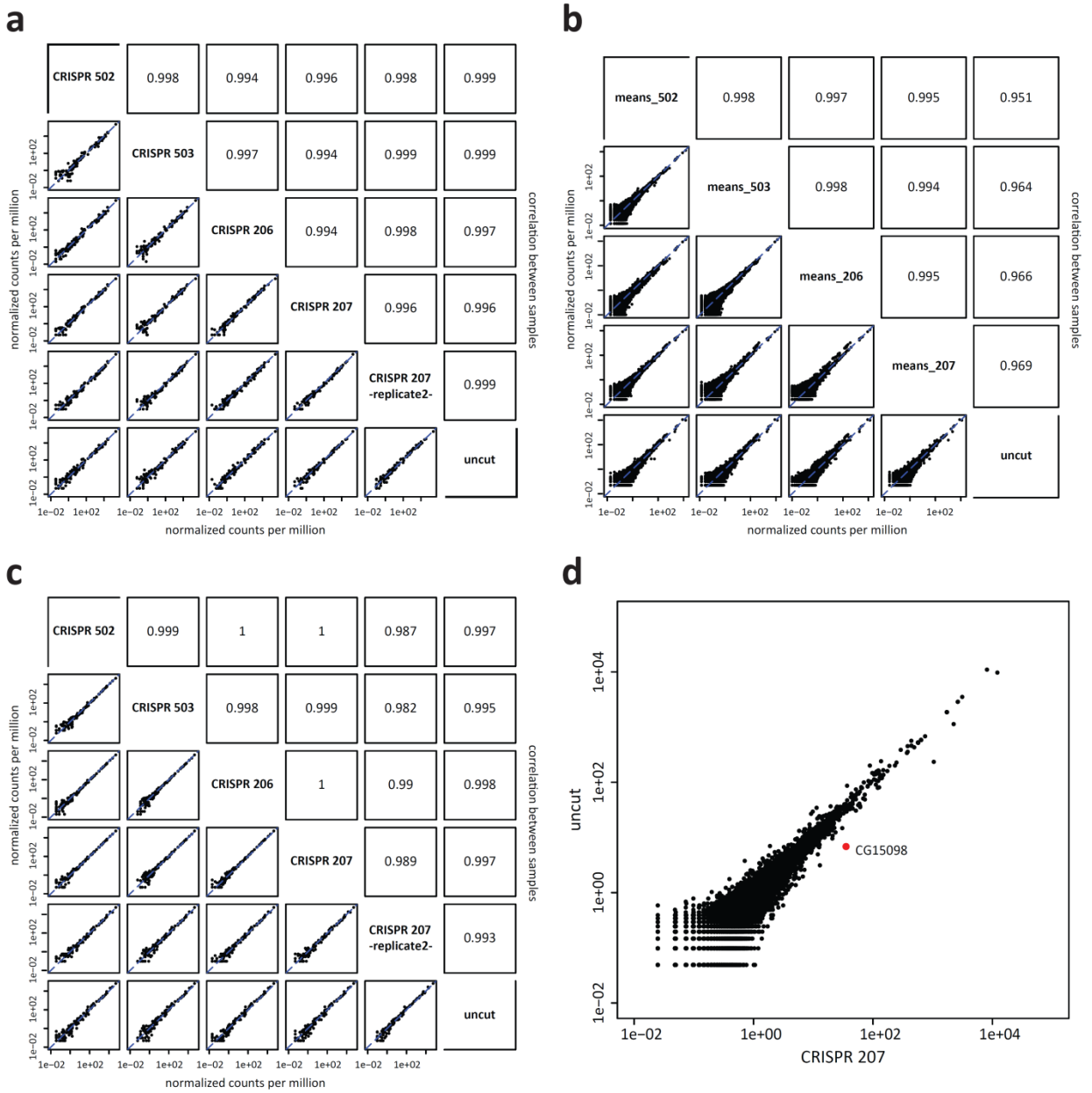

Figure 2-10: MicroRNAs, transposon-mapping small RNAs and mRNA transcripts are mostly unaffected by CRISPR-mediated introduction of DSBs. a) The abundance of each microRNA was determined for the indicated deep-sequencing libraries and normalized to display the counted reads per 10^6 reads mapping to *D. melanogaster*. The results from an uncut library as well as libraries of 4 different cutting positions within the *CG15098* gene and an additional biological replicate of CRISPR207 are compared to each other and the high correlation is indicated by the listed R-values. b) Two biological replicates are averaged for each of four different cutting positions in *CG15098* and the reads mapping to each individual transcripts are counted and normalized to the sequencing depth (counted reads/10^6 *D. melanogaster* reads). c) The numbers of small RNA reads mapping to each transposon are normalized to counts/10^6 *D. melanogaster*-mapping reads of the indicated samples and compared to each other. d) The normalized mean counts mapping to individual transcripts of the uncut and the CRISPR207-cut samples are compared and the targeted gene *CG15098* is marked in red. The indicated number of reads mapping to *CG15098* includes the large number of small RNAs matching the homology-region of the respective sgRNA.

2.4 Is this response dependent on the intron-structure and expression level of a gene?

The diverse levels of siRNA production for altered cleavage positions with respect to the first intron in the highly expressed *CG15098* gene, as well as the high abundance of splicing factors within the positive hits of the screen, suggest a connection between the presence of an intron, and thus the splicing

machinery, and break-induced siRNA biogenesis. This intron-dependence and the influence of transcription levels and the transcribing polymerase was next explored by extending the CRISPR cleavage sites to intronless, weakly expressed and Pol III-transcribed genes.

2.4.1 Cleavage of an intronless gene does not lead to a comparable siRNA response

To verify the idea that the presence of an intron upstream of the break, rather than the mere positioning of the break at a certain minimum distance to the TSS, is indeed a prerequisite for the siRNA response, the experiment was extended to *Tctp*, a highly expressed gene with no intron in neither promotor nor coding sequence. For none of the three selected cutting positions any siRNA-response comparable to the intron-containing gene was obtained as shown in Figure 2-11. Even the cleavage in the 3' UTR, more than 800nt downstream of the transcription start site and thus at a distance resulting in a strong siRNA response in the intron-containing gene, leads to siRNA generation only marginally above background level.

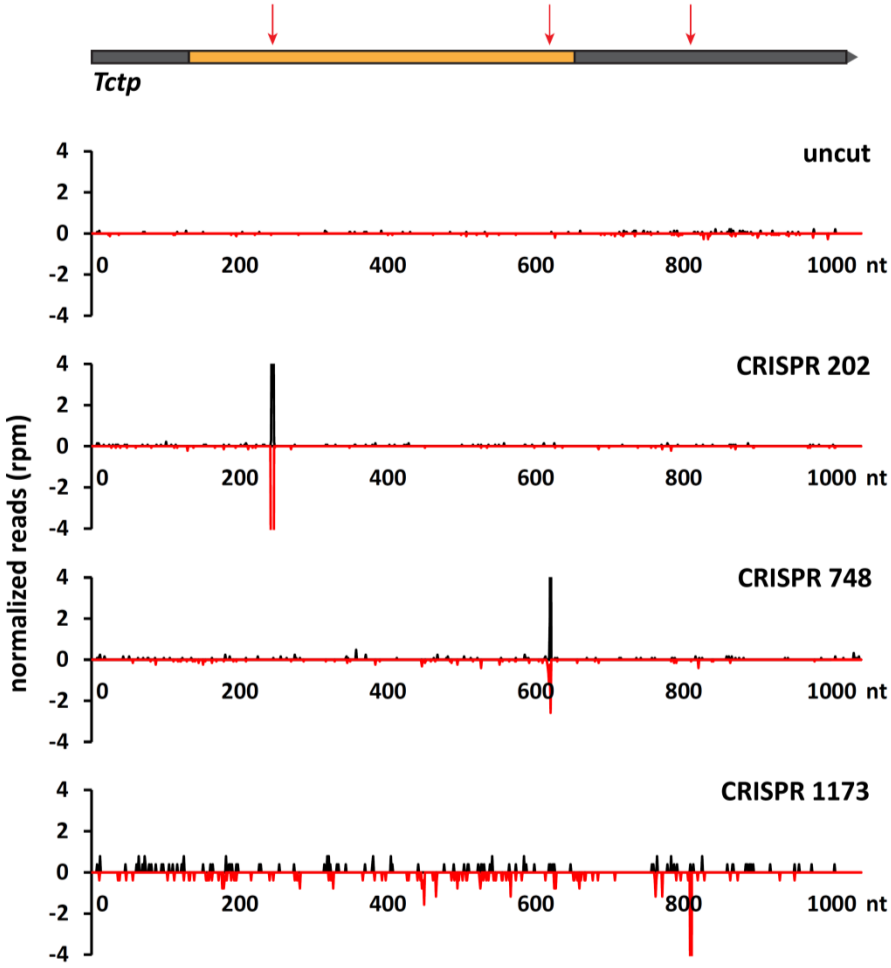

Figure 2-11: CRISPR-mediated introduction of a DSB in the highly expressed intronless gene *Tctp* does not induce small RNA formation as observed for the intron-containing gene *CG15098*. Cas9-expressing *Drosophila* cells (5-3) were transfected with site-specific sgRNA templates. Total RNA was isolated 4 days after transfection and small RNA libraries were prepared and analyzed via deep-sequencing. The obtained sequences were selected for 19-25nt reads and mapped to the target gene. The number of reads mapping to each position of *Tctp* was determined at single-nucleotide resolution and normalized on the total number of microRNA- and transposon-mapping reads to account for the sequencing depth of the specific library. The traces indicate the position of the 5'-end for sense (black) and of the 3'-end for antisense-oriented small RNAs (red) for the individual CRISPRs and the schematic drawing at the top illustrates the different cleavage positions within the gene. **Adapted from (Merk et al. 2017).**

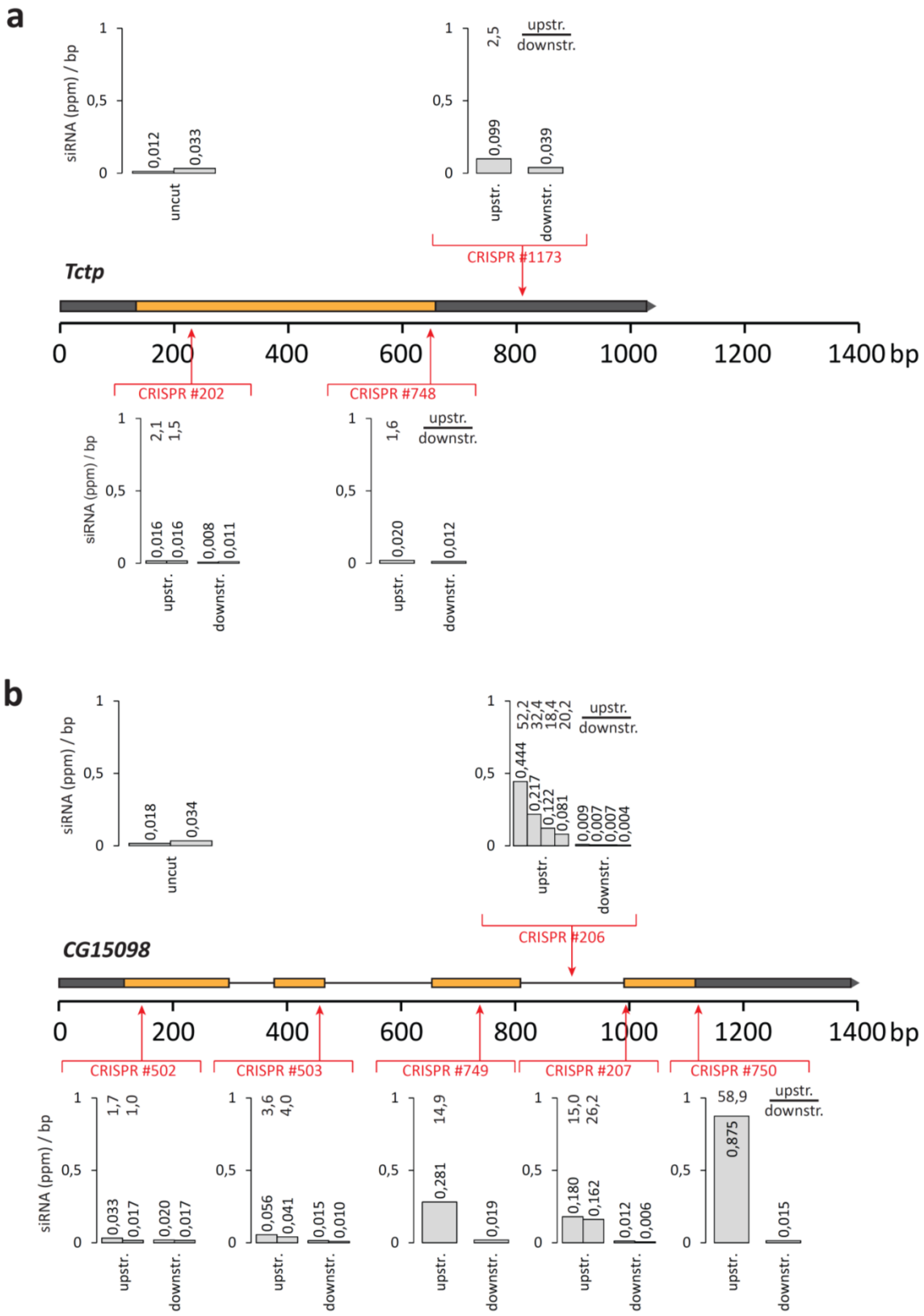

Figure 2-12: Quantification of the siRNA-response to CRISPR-induced double-strand breaks reveals a correlation between gene structure of the cleaved target and the efficacy of small RNA formation. The obtained small RNA reads mapping to the targeted gene were normalized on the total of microRNAs and transposon-mapping reads of the respective sample. The normalized reads were then used to determine the reads per basepair average for the region between the transcription start site and the respective cleavage position ("upstr.", the homology region of the sgRNA was excluded) and the region downstream of the cut to the annotated 3'-end of the transcript ("downstr."). The indicated upstream/downstream ratios quantify the siRNA response and allow for a comparison between both, different cutting positions and different target genes. a) Genomic cleavage within the intronless gene *Tctp* hardly induces any siRNA response. b) Cuts in the intron-containing gene *CG15098* lead

to a strong siRNA response if the cleavage takes place at a required minimum distance downstream of an intron. **Adapted from (Merk et al. 2017). Slight differences in the here presented numbers compared to the publication result from re-mapping of the data to an updated/corrected version of the gene model.**

With the aim to quantify and directly compare the siRNA response for different positions and genes, the matched reads between TSS and break, and those in the region succeeding the cleavage site were each normalized to the number of nucleotides on the respective side of the break with the exception of the cleavage position covered by sgRNA-derived reads. This enables a comparison of read densities per base pair upstream and downstream of the break, with more efficient formation of DSB-induced siRNAs leading to higher upstream/downstream ratios. The thus obtained quantitative measures of break-derived small RNA production are depicted in Figure 2-12 for both highly expressed genes. For the intronless gene *Tctp*, the ratios for all induced breaks range between 1.5 and 2.5 thus indicating that hardly any additional siRNAs are produced upon cleavage comparable to the cleavage positions upstream (1.0/1.7) and very close to the first intron (3.6/4.0) in the intron-containing gene *CG15098.* The comparison of both target genes regarding the siRNA production for cleavage positions around 800 nt downstream of the TSS (ratios for *CG15098*: 14.9 and 28 to 52 vs. *Tctp* 2.5) shows, that the presence of an intron upstream of the break is important for the strength of the response, rather than the mere distance to the start site. Furthermore, the introduced double-stand breaks in *CG15098* at all cleavage sites at a required minimum distance downstream of the first intron tremendously induce the formation of break-derived siRNAs (ratios between 14.9 and 58.9) with exact numbers varying between replicates and no obvious increase for cuts further downstream within the gene.

2.4.2 Cuts in less expressed genes lead to less siRNAs but comparable induction

To confirm the results attained for the strongly expressed *CG15098* gene and to investigate the influence of the gene expression level on the strength of the break-induced small RNA response, the moderately expressed, intron-containing gene *CG18273* was chosen as third target gene for Cas9-mediated genomic cleavage. As shown in Figure 2-13, the siRNA production in response to the introduced DNA double-strand break can be triggered in the less expressed gene as well. Interestingly, in addition to the siRNA response, the locus shows moderate, cleavage-independent small RNA production covering the last 1500bp of the gene even in the uncut sample. Although no overlapping transcription has been detected in this region (Graveley et al. 2011), *CG18273* and its neighboring gene *CG3176* are convergently transcribed and the annotated transcripts are separated by only 8bp. The first selected cleavage position in *CG18273* is located in the second exon about 950nt downstream of the first very short 60bp-intron as indicated in the corresponding trace in Figure 2-13 (sgRNA-derived reads were discarded during size-selection due to the shorter homology region of the sgRNA). The clear siRNA-response for this cleavage position demonstrates that one short intron is sufficient for the induction of the siRNA response as long as the cut is located at a required minimum distance downstream of the region where splicing occurs. In addition, the last investigated cleavage position close to the stop codon and thus 4686nt downstream of the TSS still triggers an endo-siRNA production that covers the area upstream of the break all the way to

the beginning of the transcript. This suggests that the mRNA transcript contributes to siRNA generation, and shows that the break-derived small RNA response can even cover a window of several kb.

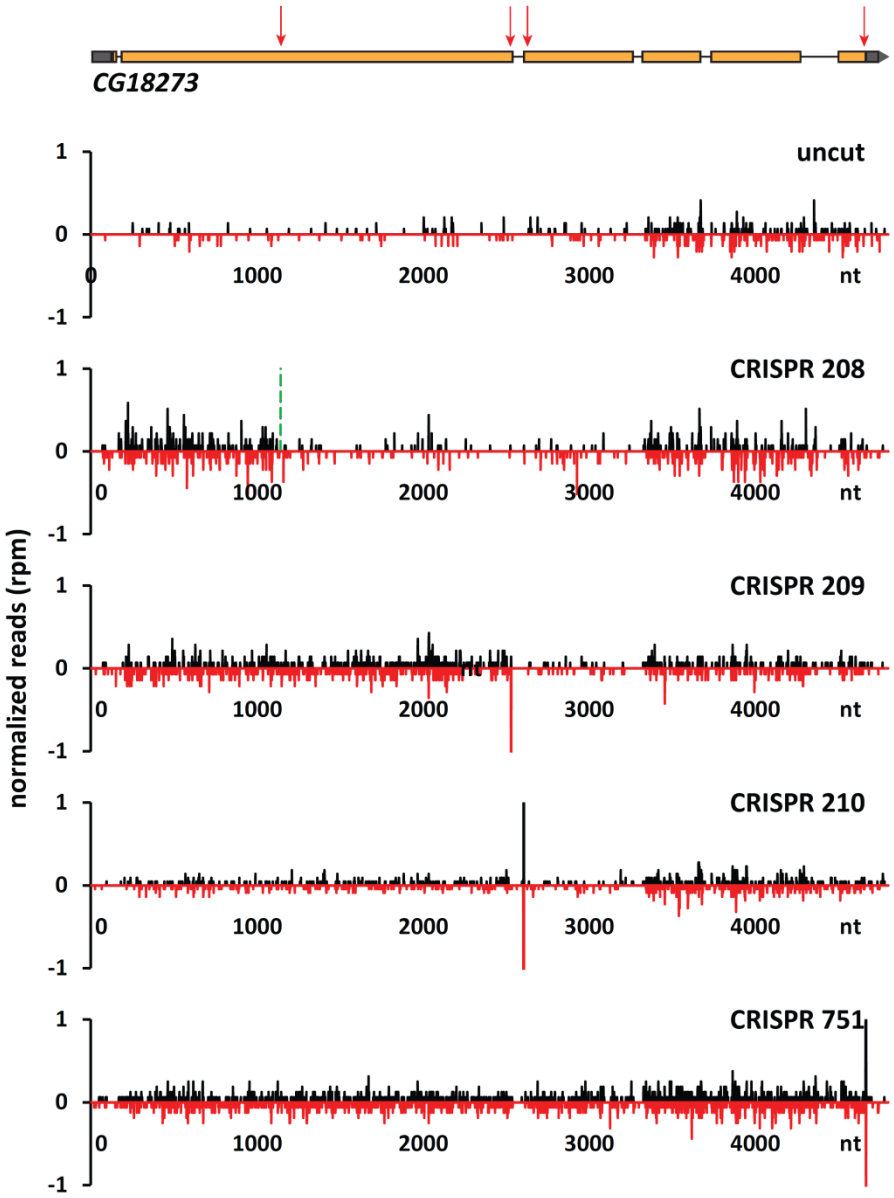

Figure 2-13: Genomic cleavage within the weakly expressed intron-containing gene *CG18273* clearly induces an siRNA response. Cas9-expressing *Drosophila* cells (5-3) were transfected with site-specific sgRNA templates. Total RNA was isolated 4 days after transfection and small RNA libraries were prepared and analyzed via deep-sequencing. The obtained sequences were selected for 19-25nt reads and mapped to the target gene. The number of reads mapping to each position of *CG18273* was determined at single-nucleotide resolution and normalized on the total number of microRNA- and transposon-mapping reads to account for the sequencing depth of the specific library. The traces indicate the position of the 5'-end for sense (black) and of the 3'-end for antisense-oriented small RNAs (red) for the individual CRISPRs and the schematic drawing at the top illustrates the different cleavage positions within the gene. The dashed line in the diagram for CRISPR 208 indicates the cutting position as the reads mapping to the shorter homology region of this sgRNA were filtered out during size selection. **Data published in (Merk et al. 2017).**

In order to quantify the siRNA response to the introduced break, the break-derived siRNA had to be distinguished from the cleavage-independent siRNAs covering the 3'-region of the gene. Therefore, the background siRNA levels of the uncut control were subtracted from each cleaved sample prior to calculating the corresponding siRNA coverage. The resulting high upstream/downstream ratios depicted

in Figure 2-14 demonstrate that the cleavage of a moderately expressed gene also leads to a comparable induction of the small RNA response. The resulting small RNA coverage also reflected in the reads-per-basepair values, however, is much lower compared to a more intensely expressed gene. The strength of the break-derived siRNA response in terms of numbers of generated small RNAs thus seems to correlate with the expression level of the target gene, again suggesting that the mRNA transcript could function as sense strand in the dsRNA precursor of the formed endo-siRNAs and/or that nascent transcription and RNA processing serves as the trigger for dsRNA generation.

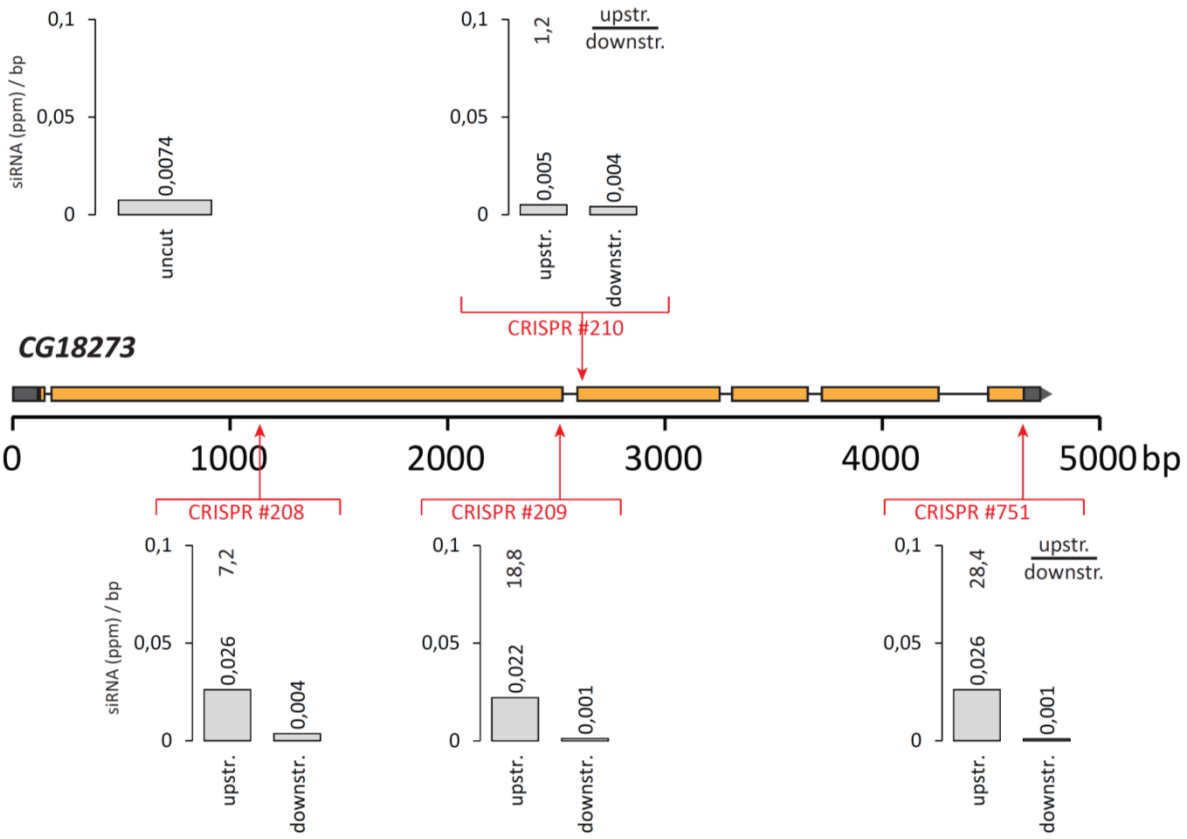

Figure 2-14: The observed siRNA response to cleavage within the weakly expressed gene *CG18273* is clearly detectable but substantially weaker than for *CG15098*. The obtained small RNA reads mapping to the *CG18273* gene were normalized on microRNAs and transposon-mapping reads of the respective library. To account for the intrinsic siRNA formation in the 3' region of the gene, the siRNA background inferred from the uncut sample was subtracted from the siRNA coverage of each sample. The reads per basepair average was determined for the regions upstream and downstream of the break and the ratios were calculated as described for the highly expressed genes.

2.4.3 Cleavage of genes with few introns or Pol III-transcribed snRNA also triggers siRNA response

Having shown that the DSB-induced siRNA response is dependent on transcription and promoted by an intron upstream of the break through investigating genes with either several or no introns and different transcription levels, the study was then extended to include the highly expressed intronless gene *RPII-33* which, however, overlaps with an convergently transcribed intron-containing gene, the moderately expressed gene *R2D2* with only two short introns (one of them in the promotor region) and the RNA polymerase III-transcribed *7SK snRNA*.

As shown in Figure 2-15a cleavage of the intronless *RPII-33* results in a small number of siRNAs despite the high transcription level of the gene. Interestingly the induced small RNAs cover not only the targeted gene but also the moderately expressed, convergently transcribed *mRpS23* gene. Both genes have overlapping 3' UTRs, with the cleavage site just a few nucleotides upstream of the *RPII-33* stop codon and thus only slightly outside the annotated 3' UTR of *mRpS23*. Although there is a weak induction of siRNA production at the locus, further experiments with different cleavage positions would be needed to clarify the contributions of the two genes to the observed small RNA coverage.

The second investigated gene *R2D2* also shows partially overlapping 3' UTRs with the convergently transcribed *cdc14* gene, resulting in cleavage-independent siRNA production downstream of the stop codon as illustrated in Figure 2-15b. In the region upstream of the stop codon, however, the siRNA production is rather weak, but clearly induced upon cleavage at the stop codon, thus confirming the observation that short introns are sufficient to induce break-derived siRNA production in moderately expressed genes.

For the RNA Pol III-transcribed 7SK snRNA, which prior to cleavage only shows a small number of reads that explicitly map in sense direction, the introduction of the DSB-break at the 3' end of the RNA transcript results in an extremely strong small RNA production. The observed small RNA response consists of equal amounts of sense and antisense reads, thus arguing for antisense transcription and dsRNA formation upon cleavage. This result clearly shows that break-derived siRNAs are not limited to RNA Pol II-transcribed genes, but can also occur in response to cleavage of genes transcribed by RNA Pol III.

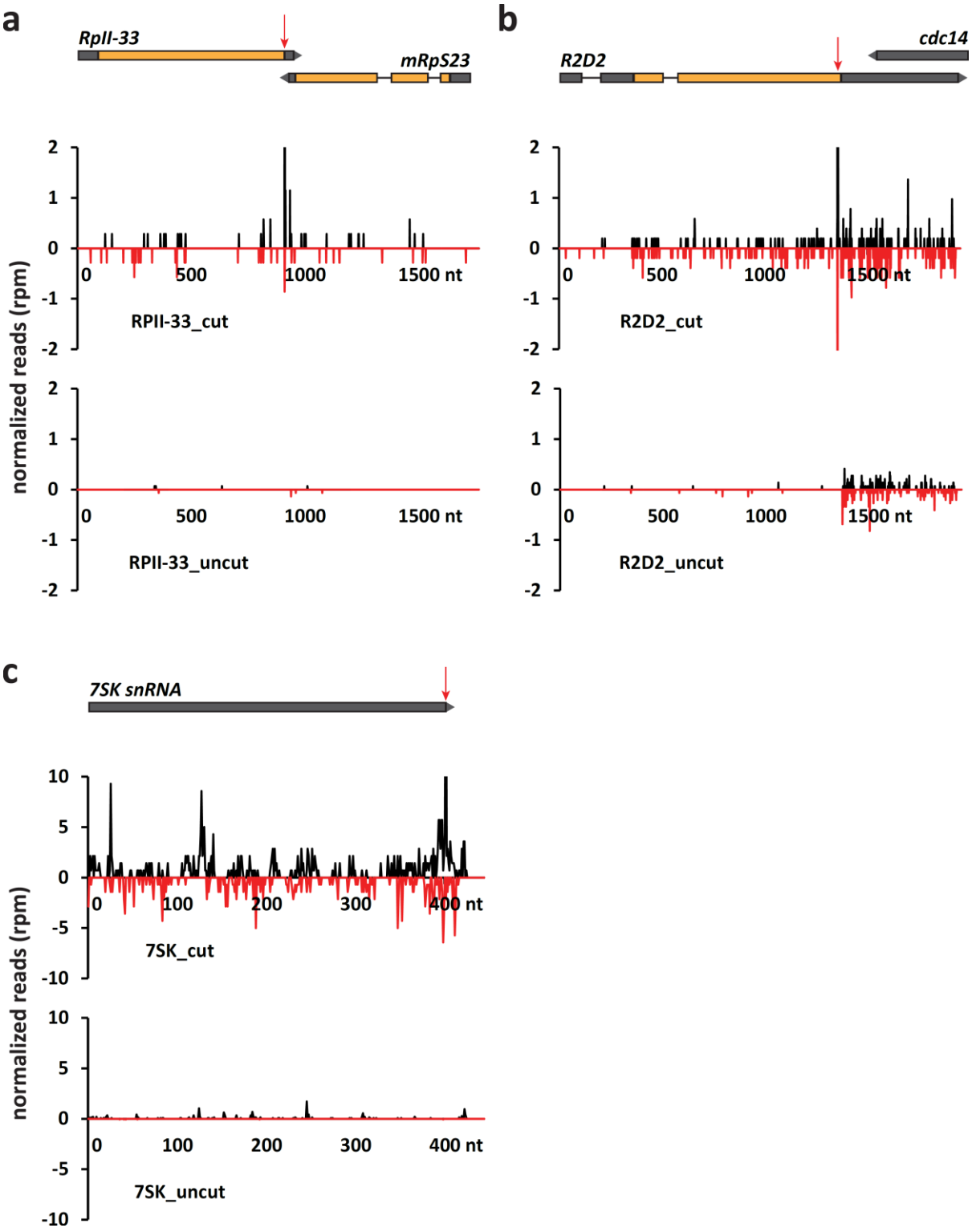

Figure 2-15: The introduction of DNA double-strand breaks in additional genes including the Pol II-transcribed *7SK snRNA*, the intronless *RpII-33* and *R2D2* leads to differently pronounced siRNA responses. The target genes were cleaved at the positions indicated with arrows in the cartoons. Deep-sequencing libraries were prepared and analyzed as described for *CG15098*, *CG18273* and *Tctp*. Due to the low sequencing depth, the results from two independent sequencing runs of the same library were concatenated for *RPII-33* and *R2D2* thus reaching about 25% of the usual sequencing depth. The obtained reads were normalized on the total number of microRNA- and transposon-mapping reads. The traces indicate the position of the 5'-end for sense (black) and of the 3'-end for antisense-oriented small RNAs (red). a) Cleavage close to the stop codon of the intronless *RpII-33* gene induces a weak siRNA response in both the target gene and the intron-containing *mRpS23* gene which has overlapping 3' UTRs with *RPII-33*. b) *R2D2* displays a clear siRNA response to a double-strand break introduced at the stop codon in addition to and clearly distinguishable from the ubiquitous siRNA production in the 3' UTR of the gene. c) Cleavage at the 3'end of the RNA polymerase III-transcribed *7SK snRNA* triggers a strong siRNA response covering the whole transcript.

2.4.4 MicroRNA and transposon-mapping small RNA levels remain unchanged

Comparable to the different cuts in *CG15098* which do not induce any major changes in miRNAs and transposon-mapping siRNAs as well as transcript-mapping small RNAs, the analysis shown in Figure 2-16 demonstrates that this holds true for *Tctp*, *CG18273*, *R2D2*, *RPII-33* and *7SK snRNA* as well.

Concerning miRNAs, *Tctp* and *CG18273* exhibit high correlations with the uncut sample (Figure 2-16a). For both the transposon- and transcript-mapping reads (Figure 2-16c and e), the samples with cuts in different genes show a high correlation with each other, however, the respective correlations with the uncut samples are slightly lower. This can be attributed to the fact, that the uncut sample was not grown and prepared in parallel to the others. The outliers in Figure 2-16e, that are each specific for one of the cleaved genes, again mark the siRNAs mapping to the target gene and the homology region of the respective sgRNA.

When comparing the individual miRNA levels of *R2D2*, *7SK snRNA* and *RPII-33*, one specific miRNA, namely the so far uncharacterized miR-9388, clearly becomes more abundant upon cleavage of RPII-33 (Figure 2-16b), although the two loci do not share any sequence similarity. The transposon- and transcript-mapping reads (Figure 2-16d and f) in contrast do not show any striking differences for any of the target genes. When comparing the transcript-mapping reads between the *R2D2* and *RPII-33* samples, the most obvious datapoint with an increased number of reads after *RPII-33* cleavage can be attributed to the gene itself, whereas the outliers with increased small RNA production upon DSB-induction in *R2D2* are two transcripts of *R2D2* itself as well as 3 transcripts of the cdc14 gene exhibiting a 3' UTR overlap with *R2D2*.

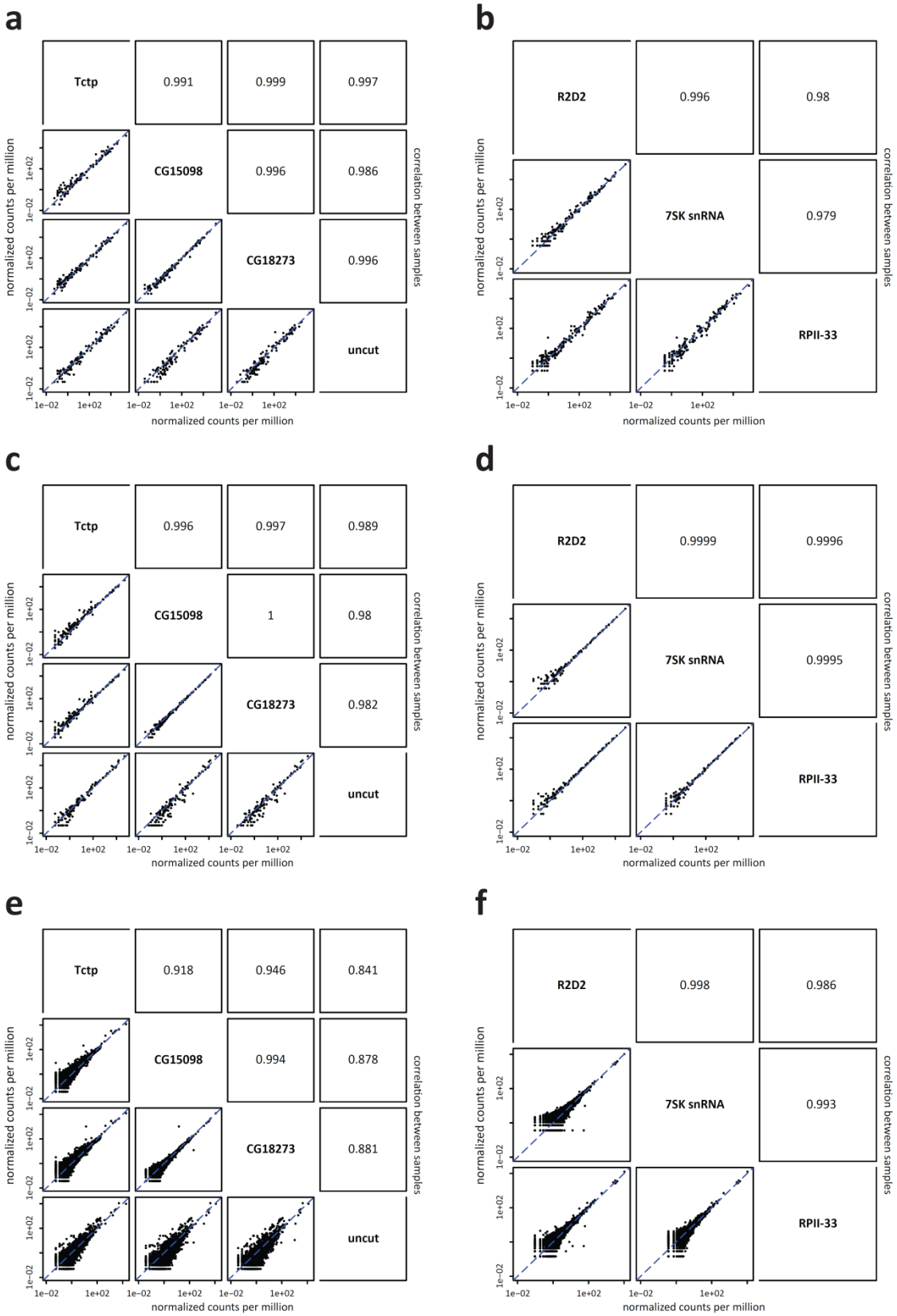

Figure 2-16: MicroRNAs, transposon-mapping small RNAs and most transcript-mapping small RNAs are comparable for all cleaved genes and can therefore not explain the differences in cleavage-induced siRNA production at different target loci. The abundance of microRNAs and transposon- and transcript-mapping small RNAs was normalized to the reads of the specific library mapping to *D. melanogaster* and are displayed in reads per million. The results from different target genes, each cut near the stop codon or 3' end of the transcript, and an uncut library are compared to each other and the correlation is indicated

by the listed R-values. a) Comparison of microRNAs upon cleavage in *Tctp*, *CG15098* and *CG18273*. b) Comparison of microRNAs upon cleavage in *R2D2*, *7SK snRNA* and *RPII-33*. c) Comparison of siRNAs mapping to transposons upon cleavage in *Tctp*, *CG15098* and *CG18273*. d) Comparison of siRNAs mapping to transposons upon cleavage in *R2D2*, *7SK snRNA* and *RPII-33*. e) Comparison of transcript-mapping small RNAs upon cleavage in *Tctp*, *CG15098* and *CG18273*. f) Comparison of transcript-mapping small RNAs upon cleavage in *R2D2*, *7SK snRNA* and *RPII-33*.

2.5 Transfection of U6-sgRNA template triggers siRNAs specifically covering the sgRNA

In addition to break-derived siRNAs covering the area between the transcription start site and the double-strand break, extremely abundant small RNAs mapping to the cleavage position are detected for all cut samples. To further investigate these small RNAs, they were mapped to different CRISPR U6-sgRNA template sequences as shown in Figure 2-17.

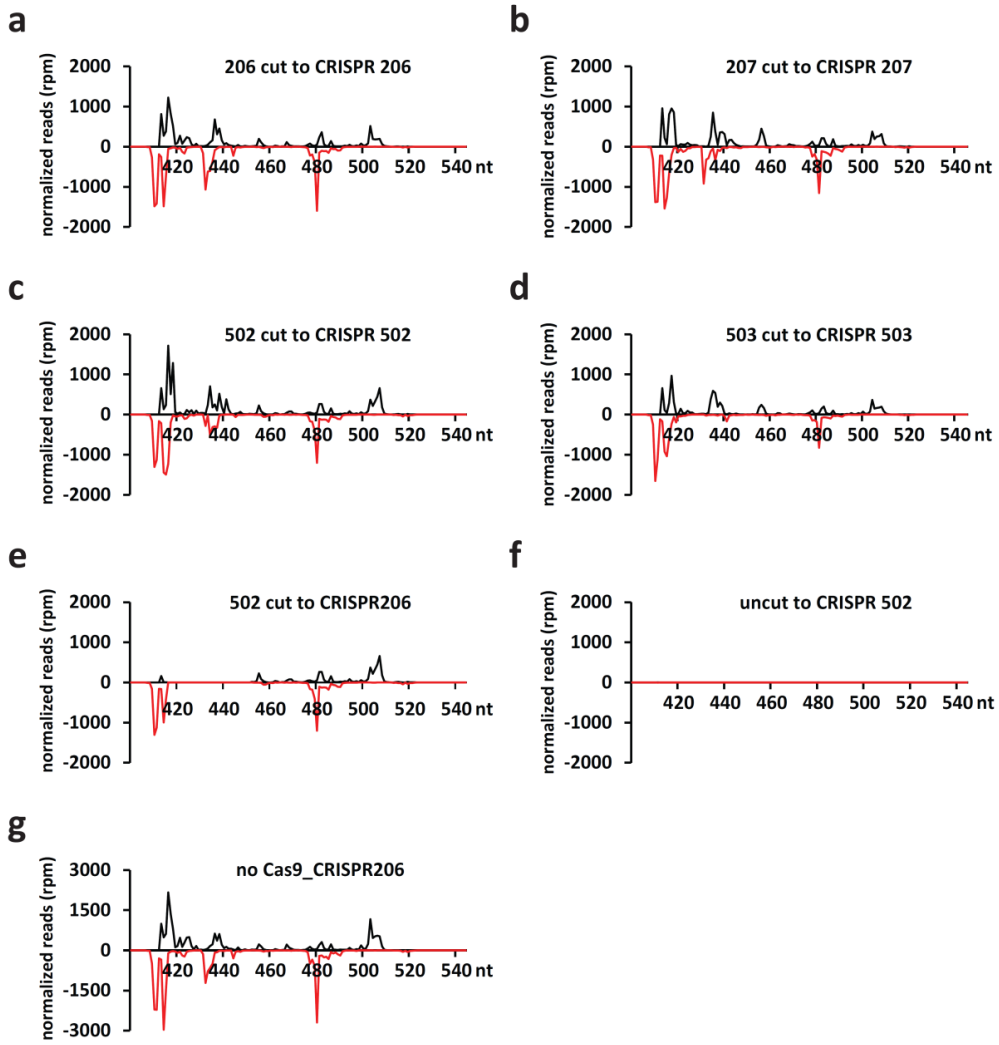

Figure 2-17: Transfection with CRISPR sgRNAs induces the production of small RNAs that cover both, the target-specific homology region and the general region of the structured sgRNA. The size-selected (19-25nt), normalized reads (normalization on microRNAs and transposon-mapping reads) were mapped to the indicated CRISPR U6-sgRNA templates. Sense/antisense reads are depicted in black/ red for the transcribed region of the U6-sgRNA template (position 416 to ~530). The indicated positions (nt) refer to the PCR-product including the U6-promoter and termination sequence flanking the transcribed sgRNA. Reads are mapped to the 5'-end for sense and the 3'-end for antisense-oriented RNAs. a) Deep-sequencing library cut with CRISPR 206 mapped to the corresponding sgRNA. b) Deep-sequencing library cut with CRISPR 207 mapped to the corresponding sgRNA. c) Deep-sequencing library cut with CRISPR 503 mapped to the corresponding sgRNA. d) Deep-sequencing library cut with CRISPR 502 mapped to the corresponding sgRNA. e) Deep-sequencing library cut with CRISPR 502 mapped to the non-corresponding CRISPR 206 sgRNA. f) Deep-sequencing library of untransfected Cas9-expressing cells mapped to the absent CRISPR 502 sgRNA. g) Deep-sequencing library of ordinary S2 cell (no Cas9 expressed) transfected with a CRISPR 206 mapped to CRISPR 206 sgRNA.

Upon mapping to the sgRNA that was used in the specific sample to introduce the cut as depicted for four different samples in Figure 2-17 a-d, the resulting small RNA distributions show profound similarities. For all samples, the maximum of normalized reads mapping to any position within the sgRNA is greater than 1000 and thus more than 200x higher than the strongest break-derived response of the corresponding genes. Furthermore, the reads are not equally distributed but show a reproducible pattern of covered and non-covered regions of the sgRNA. In order to distinguish between reads that are specific for the sgRNA matching the targeted position and reads that map to the sequence common to all sgRNAs, the reads obtained while programming the Cas9 nuclease with CRISPR 502 were also mapped to CRISPR 206, as shown in Figure 2-17e. While the pattern downstream of the homology region remains unchanged, all reads mapping to or overlapping with the gene-specific region downstream of position 434 of the CRISPR sgRNA template are not detected anymore.

To determine whether the utilized S2 cell-derivates endogenously expressing the Cas9 nuclease per se produce any siRNAs that match the sequence of the sgRNA, the small RNAs of untransfected cells were also mapped to CRISPR 502. Figure 2-17f illustrates, that these cells do not produce any small RNAs with sequence homologies to the sgRNA. Furthermore, as demonstrated in Figure 2-17g, the production of sgRNA-matching small RNAs is only dependent on the presence of the sgRNA and does not require the Cas9-protein to be expressed or present in the cells, as customary S2 cells and Cas9-expressing cells show an identical response to CRISPR sgRNAs.

Having clarified the prerequisites and general features of the small RNA response to sgRNAs, Figure 2-18 focuses on the detailed characteristics of these observed small RNAs. As mentioned above, the small RNA reaction to different sgRNAs shows similar patterns of unequally distributed RNAs some of them mapping to the region homologous to the target gene, and can thus potentially influence these targets. Figure 2-18a shows that a CRISPR-independent and reproducible number of roughly 15% of the obtained CRISPR-mapping reads correspond to this homology region. To see whether the strength of the break-induced siRNA response correlates with the number of reads mapping to the sgRNA (or to the homology region), the siRNA responses quantified by the respective upstream/downstream ratios were normalized to the CRISPR-mapping reads for different cleavage positions in *CG15098* as shown in Figure 2-18b. The hence normalized values display the same features as the initial ratios, thus indicating that there is no correlation between the quantities of the break-induced and CRISPR-induced small RNAs.

To answer the question whether the CRISPR-induced small RNAs show the characteristics of typical siRNAs and what regions of the highly structured sgRNA particularly stimulate small RNA production, Figure 2-18c contains a detailed alignment of the small RNAs to the structured sgRNA including a length distribution of the sgRNA-mapping reads. Although the numbers of sense and antisense reads are not comparable for all positions and no exact processing-footprint of Dcr-2 can be detected, the presence and at least partly correlating distribution of sense and antisense reads as well as the clear length preference

for 21 nucleotides argue for the CRISPR-mapping small RNAs to be classical siRNAs. The alignment to the sgRNA shows further that the pattern of siRNA formation indeed correlates with the location of certain structural features. The first major peak is located at the end of the U6 promotor enabling the in-vivo transcription of the sgRNA from the transfected DNA template by RNA polymerase III (Dieci et al. 2007). The sgRNA-induced siRNA response is thus consistent with the finding that genes transcribed by Pol III can also give rise to siRNA production as shown for cleaved 7SK snRNA. The second peak within the siRNA distribution can be attributed to the starting position of the homology region. Although the sgRNA does not display any specific intrinsic structure at this position, it can pair with the target DNA, thus creating DNA-RNA hybrids (Jiang et al. 2016). The next smaller peak is located at the transition from the homology region to the first hairpin structure of the sgRNA followed by extremely strong small RNA production at the position where the hairpin structure is again basepaired after an interrupting bulge. This position, where predominantly antisense reads are created, is also paired with the exact position of the previous smaller peak of mainly sense RNA production. The last distinct peak of sense-orientated small RNAs can be assigned to the region directly preceding the second hairpin. There are no siRNAs detected covering the region between the U6-termination signal and the end of the PCR-template (position 520 to 540 in Figure 2-17).

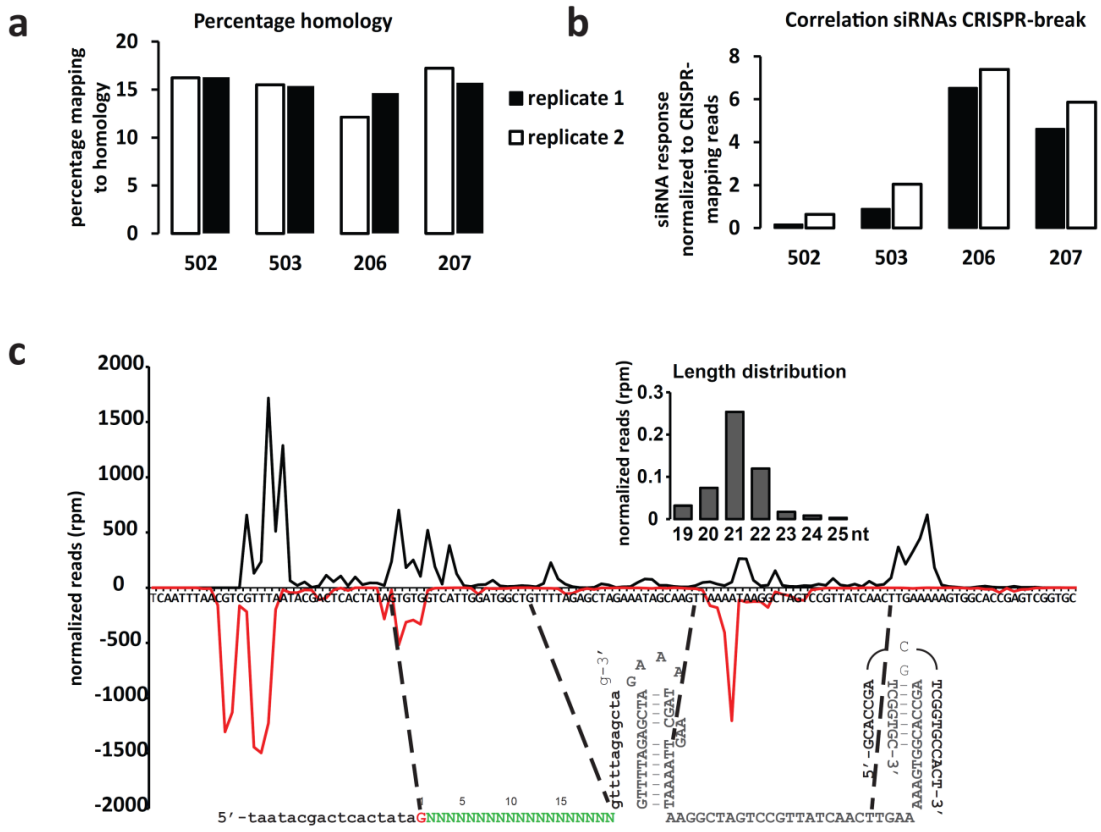

Figure 2-18: CRISPR-mapping small RNAs show typical characteristics of siRNAs and are distributed in a structure-dependent manner along the sgRNA, independent of the specific homology sequence or effect on the target gene. The normalized, size selected reads from the indicated deep-sequencing libraries are mapped to the corresponding sgRNA templates and analyzed. The 5'-end for sense-oriented and the 3'-end for antisense-oriented siRNAs are mapped to the target sequence. a) The percentage of CRISPR-mapping reads mapping to the region homologous to the respective target gene is comparable for all displayed sgRNAs. b) The numbers of CRISPR-mapping small RNAs are compared for different sgRNAs inducing strong or

weak small RNA responses at the target positions. Therefore, the calculated upstream/downstream ratios of the indicated target positions were normalized to the number of CRISPR-mapping small RNAs in the respective sample. c) Detailed mapping of the small RNAs from the deep-sequencing library cleaved with CRISPR 502 with respect to the scaffold structure of the corresponding sgRNA. Small RNAs in sense orientation are shown in black, antisense reads in red. The length distribution of the small RNAs mapping to the sgRNA scaffold is included and shows a clear preference for typical 21nt siRNAs.

Taken together, the sgRNA-induced small RNA response shows characteristic features of typical siRNAs and the distribution pattern appears to be connected to the structural features of the sgRNA.

2.6 Stalling of the pre-catalytic spliceosome promotes siRNA formation

High transcription levels and the existence of an intron upstream of a DNA double-strand break and thus the presence of the splicing machinery at the nascent transcript promote break-induced siRNA formation. Among the various splicing factors identified in the genome-wide screen, those associated with the Prp19/Prp19-related complex enabling the transition of the pre-catalytic spliceosome to the first catalytic step are particularly abundant within the positive hits. This puts forward the question, whether instead of the splicing reaction, the DSB-induced stalling of a spliceosome in a pre-catalytic step could perhaps trigger the formation of dsRNA with the unspliced pre-mRNA as template for antisense transcription.

2.6.1 At which step of the splicing reaction is double-stranded RNA generated?

Double-stranded RNA is a prerequisite for the formation of functional siRNAs. The observed correlation between the transcription levels of the target locus and the strength of the small RNA response to DNA double-strand breaks suggests the mRNA transcript to contribute the sense strand to this dsRNA precursor. To determine whether the transcripts are correctly or partly spliced prior to dsRNA formation and thus further refine the role of the splicing machinery for siRNA formation, the siRNAs formed upon cleavage of *CG15098* were analyzed in detail with respect to the underlying gene structure. As shown in Figure 2-19a, the small RNAs uniformly cover intronic and exonic regions of *CG15098* upstream of the induced break and extend all the way to the transcription start site of the cleaved locus. Furthermore, mapping of the small RNAs to the spliced mRNA sequence of the target locus reveals that no small RNAs spanning the exon-exon junctions can be detected. In order to quantify the small RNA coverage with respect to the underlying gene structure, the number of reads per base pair was calculated for all exons and introns as well as 5' and 3' splice junctions along the gene as illustrated in Figure 2-19b not revealing any correlation between gene structure and siRNA coverage. While intron 2 shows a lower coverage compared to the preceding exon ($p=0.034$, two-sided Student's t-test, $n=4$), the contrary is observed for intron 3 ($p=0.242$). If the sense transcript were fully spliced prior to dsRNA formation, few siRNAs covering the 3' splice junction should be detected. However, Figure 2-19c, summarizing the siRNA coverage along the gene according to gene structure, reveals, that intron-exon-junction spanning reads are even slightly enhanced compared to exonic reads ($p=0.052$) thus suggesting that the second transesterification reaction does clearly not precede dsRNA formation. In case the splicing reaction was stalled after the first catalytic step, the 5' junction should exhibit lower small RNA coverage compared to

exons. While this is true for the 5' junction of the second intron (p=0.012), there is no diminished coverage of the 5' splice site of neither the first nor the third intron. Furthermore, when compared to the adjacent intron, no reduction of the reads can be observed for the second 5' junction (p=0.946).

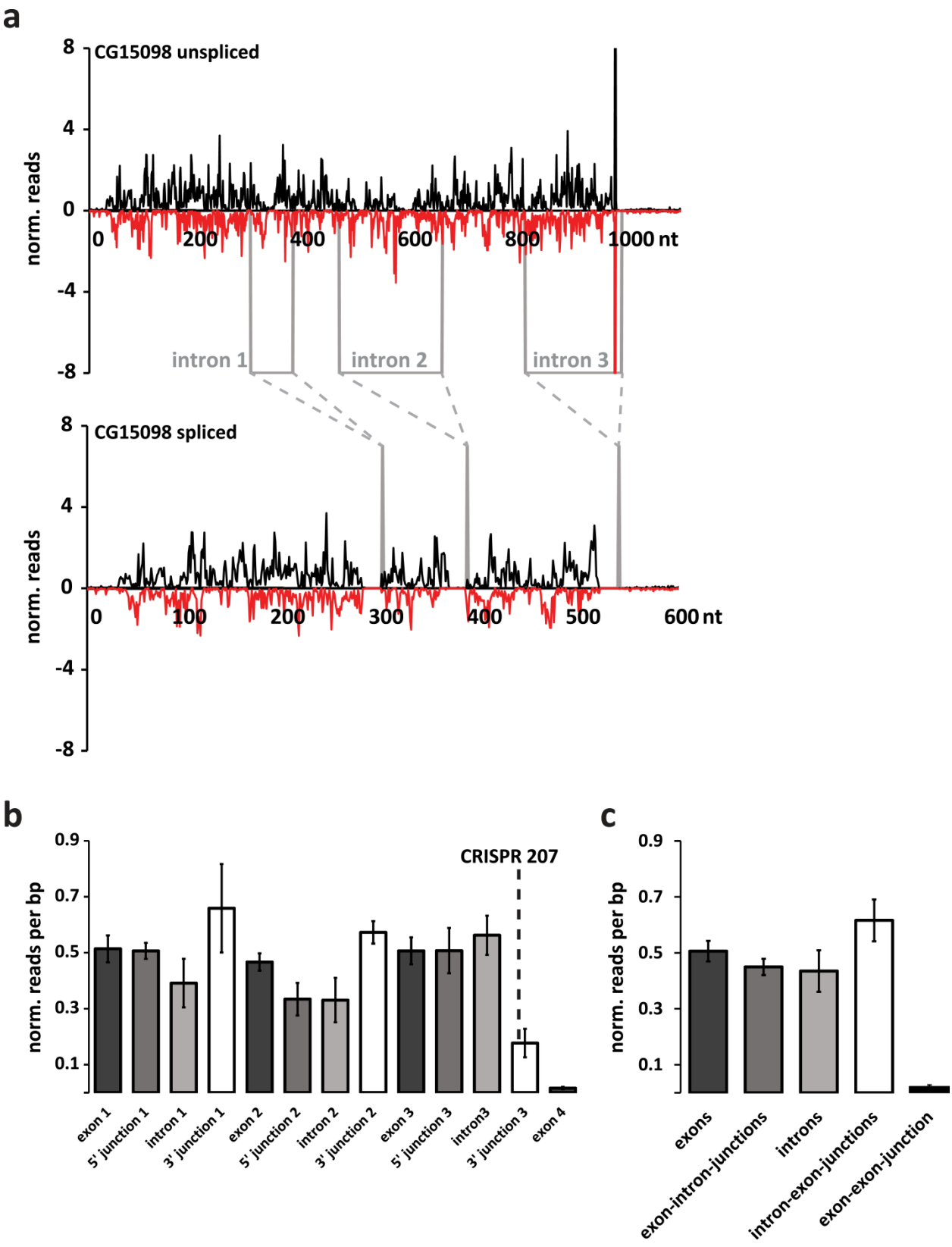

Figure 2-19: Intronic and exonic regions of the unspliced *CG15098* transcript are covered by DSB-induced siRNAs with no conclusive correlation between the coverage pattern and the underlying gene structure. The normalized deep-sequencing

results from four replicates (two biological replicates with two technical replicates each) with cleavage at the end of the third intron (CRISPR 207) were averaged. The 25nt window surrounding the 5' and 3' ends of introns were considered splicing junctions and the cleavage position (5nt frame) was excluded from the analysis. a) Small RNA coverage of spliced and unspliced transcripts. b) The average per bp siRNA coverage was quantified for each exon, intron and splice junction. Error bars indicate the standard deviation of the replicates. c) The per bp siRNA coverage upstream of the induced break was averaged across all introns/exons/splice-junctions of the four replicates. Exon-exon-junction spanning reads were considered in a 19nt window surrounding the splice junction of the spliced transcript. **In parts adapted from (Merk et al. 2017). Differences in the here presented graphs compared to the publication result from re-mapping of the data to an updated/corrected version of the gene model.**

The reduced coverage of intron 2 and its 5' splice junction as well as the mostly unchanged coverage of the other introns is observed for different cleavage positions as depicted in Figure 2-20. This indicates that the reduced siRNA coverage is most likely due to the local characteristics of this gene region such as the specific sequence rather than the underlying gene structure in general.

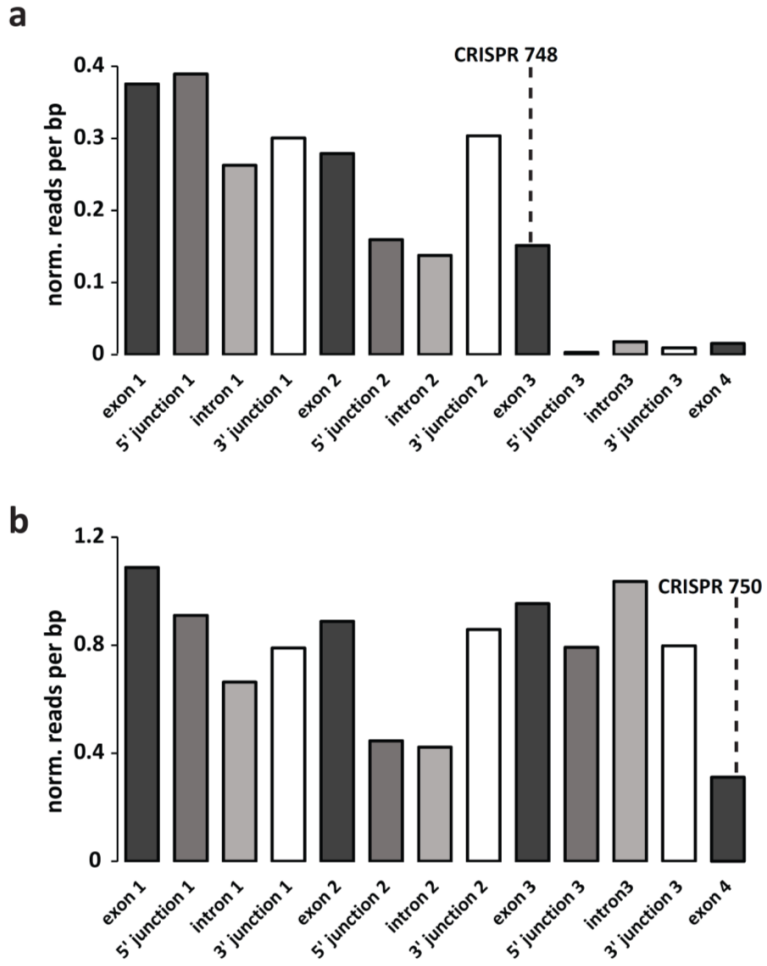

Figure 2-20: The siRNA coverage with respect to the underlying gene structure is identical for different cleavage position within *CG15098*. The normalized per bp siRNA coverage was quantified for each exon, intron and splice junction (25nt window) upon cleavage with the indicated sgRNA (5nt frame around cleavage position excluded from analysis) within the target gene *CG15098*. a) Small RNA coverage upon cleavage with CRISPR 748. b) Small RNA coverage upon cleavage with CRISPR 750. **Adapted from (Merk et al. 2017). Differences in the here presented numbers compared to the publication result from re-mapping of the data to an updated/corrected version of the gene model and slightly changed calculations without impact on the conclusions.**

In contrast to the uniform coverage of intronic and exonic regions as well as exon-intron and intron-exon junctions, exon-exon spanning siRNAs are almost completely absent for all possible exon-exon junctions ($p=0.0001$ compared to exonic reads) thus suggesting that the unspliced transcript indeed contributes the

sense strand to the dsRNA precursor and the siRNAs are not generated by RNA-dependent antisense transcription of the mature mRNA.

Taken together, although upstream splicing stimulates siRNA formation, the resulting small RNAs are uniformly distributed across the locus including exons, introns and both splicing junctions. Exon-exon spanning siRNAs cannot be detected thus implying that the splicing reaction does not occur prior to siRNA formation. This strongly supports the idea that the spliceosome gets indeed stalled in a pre-catalytic state and the unspliced transcript is used for dsRNA formation.

2.6.2 Splicing of different genes is not impaired upon knockdown of selected splicing factors

Having shown that introns and thus the presence of the splicing machinery upstream of the break are important for the formation of DSB-induced siRNAs, the next step was to determine whether the knockdown of certain splicing factors diminishes mRNA splicing at the target locus and thus reduces siRNA formation without a direct role of these splicing factors. Therefore, the splicing efficiencies of the different introns at the previously used target locus *CG15098* were investigated upon knockdown of three different splicing factors. Additionally, a second target gene, *tsr* which also comprises three introns within the coding sequence, however in contrast to *CG15098* with a first small intron and two subsequent larger ones, was examined as well. Furthermore, to consider and possibly distinguish the effect of the knockdowns on the splicing of intron-containing promotors, three target genes with introns of different lengths exclusively within the promotor (*Act5C*: intron 1.7kb, *Act42A*: intron 200bp, and α-tubulin84B: intron-containing promotor used for firefly reporter plasmid) were studied likewise. The selected splicing factors comprise the SR protein kinase Darkener of apricot (Doa) involved in alternative splicing, l(1)10Bb, a component of the Prp19-related complex, and the hnRNP protein and splicing regulator Hrb27C, which had all shown a strong effect in the genome-wide screen for DSB-induced siRNA formation. As shown for Doa and Hrb27C in Figure 2-21a, the knockdown of the respective factors is efficient and the subsequently measured splicing efficiencies are thus detected for strongly reduced levels of these splicing components. Probably due to the low fraction of tagged l(1)10Bb protein within the investigated non-clonal cell-line, this knockdown could not equally be demonstrated via Western blot.

In order to measure final splicing efficiencies via qRT-PCR, total RNA was isolated four days after knockdown (identical timeframe compared to DSB luciferase-reporter and CRISPR deep-sequencing experiments) and the levels of unspliced pre-mRNA and spliced mRNA were quantified. The thus obtained values were normalized to total transcript levels obtained via amplification of a splicing-independent sequence within the respective target locus. As shown in Figure 2-21 b-d for nearly all different target introns and RNAi against selected splicing factors, no increased levels of unspliced transcript can be detected compared to the control knockdown. Furthermore, the amount of spliced transcripts in the steady-state situation matches the total amounts of transcripts at the target locus, indicating that the vast majority of transcripts are still correctly spliced upon knockdown of distinct

splicing factors. The sole exception is the second intron of *tsr* which shows an increase of unspliced transcript upon knockdown of l(1)10Bb. However, even in this case the levels of spliced transcripts are comparable to the total transcript amounts at the locus.

Thus, the influence of the selected splicing factors in the siRNA-formation does not result from any changes in the steady-state levels of correctly spliced transcripts upon knockdown, but can rather be explained by changed splicing kinetics due to the DNA double-strand break. The trigger for siRNA production is hence not the splicing machinery per se, but a signal created when the process of splicing is compromised at a DNA double-strand break.

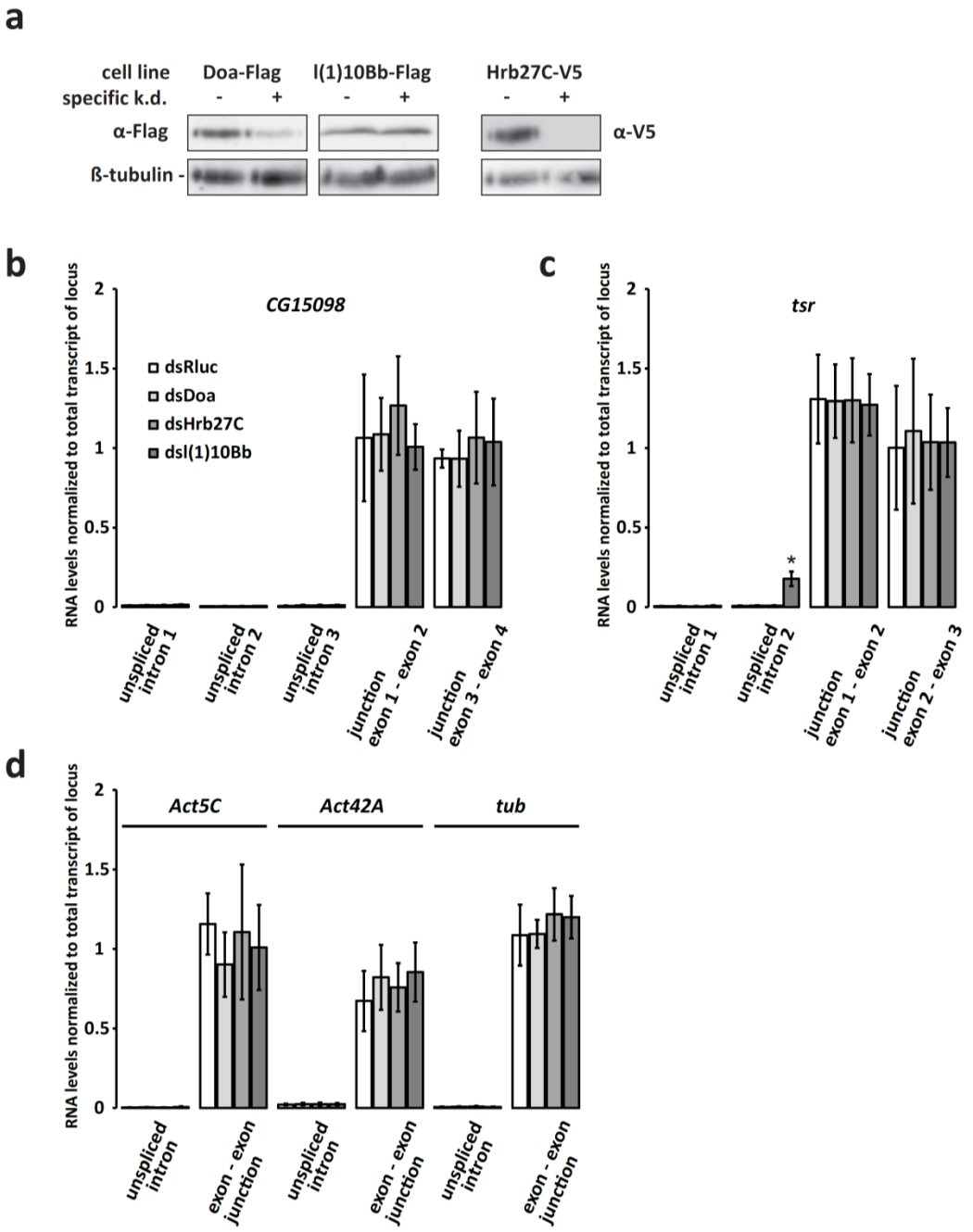

Figure 2-21: Steady state levels of efficiently spliced transcripts are not generally decreased upon knockdown of selected splicing factors. a) Endogenously tagged S2 cells were subjected to RNAi using the indicated dsRNAs against selected splicing

factors. After four days, protein was isolated and analyzed via Western blotting to evaluate the success of the specific knockdown of the Flag- and V5-tagged splicing factors with ß-tubulin as loading control. b) S2 cells were subjected to RNAi against selected splicing factors identified in the genome-wide screen. Four days after knockdown, RNA was isolated and analyzed using qRT-PCR with specific primers to quantify the levels of unspliced pre-mRNA (primer pair with one primer each binding in the intronic/exonic sequence of the specific intron) and spliced mRNA transcript (primer pair with one primer spanning the exon-exon junction). A third amplicon within a fully exonic part of the mRNA thus quantifying the transcript level of the investigated locus was used for normalization. All values are shown as mean ± SD of three biological replicates. For all three introns of the depicted target locus *CG15098*, no increase of unspliced pre-mRNA or decrease of fully spliced mRNA could be detected. b) The second target locus *tsr* showed significantly increased levels (p=0.023, two-sided student's t-test, n=3, marked with *) of unspliced pre-mRNA specifically at the second intron upon knockdown of l1(10)Bb. However, even in this case the majority of transcripts are still fully spliced. All values are shown as mean ± SD of three biological replicates normalized on transcript levels. c) No changes or differences in splicing efficiency could be detected for *Act5C* (large intron in promotor), *Act42A* (short intron in promotor) or *α-tubulin84B* (intron-containing promotor used for firefly reporter plasmid). All values are shown as mean ± SD of three biological replicates normalized on transcript levels of the specific target locus. **In parts published in (Merk et al. 2017).**

2.7 Is the integration of a short intron sufficient for siRNA production?

While the cleavage of a genomic locus downstream of an intron leads to the biogenesis of small RNAs covering the area between the introduced double-strand break and the transcription start site of the gene, such an siRNA response cannot be observed for the intronless gene *Tctp*. In order to show that a single intron upstream of the cleavage site is sufficient to trigger break-induced siRNA formation, I introduced a functional intron into *Tctp* and analyzed the resulting small RNA response.

2.7.1 A functional intron can be integrated into an intronless gene in *Drosophila* cells

In *Drosophila*, few genes are indeed intronless, especially promotors often comprise at least a short intron (Mount et al. 1992). As a first step for intron-integration into *Tctp*, the relatively short 74bp intron from *GAPDH2* was amplified via PCR as depicted in Figure 2-22a. *GAPDH2*, the source of the intron, and the target gene *Tctp* have similar expression levels and are of comparable length. Both genes have no intron within the coding sequence and the template intron located 18nt upstream of the open reading frame in *GAPDH2* was integrated into *Tctp* directly upstream of the start codon to enable a similar genetic environment. The exact position for intron integration was chosen to obtain a functional 3' splice site, and the 5' splice site was added during PCR amplification to enable efficient splicing of the intron in *Tctp*. In addition, a control construct without typical splicing donor sequence was created. The integration was facilitated in analogy to the Cas9-mediated genomic protein-tagging in S2 cells (Bottcher et al. 2014). After Cas9-mediated cleavage and intron insertion via homologous recombination (Figure 2-22a), single cell clones were obtained via marker-free clonal selection as the intron was integrated without any additional linker or marker sequences. As demonstrated in Figure 2-22b for several clones, the endogenous intron is indeed successfully integrated at the new target locus and for clone C5 (wildtype splice sites) and B4' (compromised splice site) no remaining intronless allele of *Tctp* can be detected. The comparison of gDNA and cDNA in Figure 2-22c demonstrates further, that the integrated intron is spliced out in clone C5 in contrast to the splicing-defective construct in clone B4'. The clonal cell-line C5 with the splicing-competent intron in *Tctp* was then subjected to CRISPR-mediated cleavage downstream of the intron and analyzed by deep-sequencing.

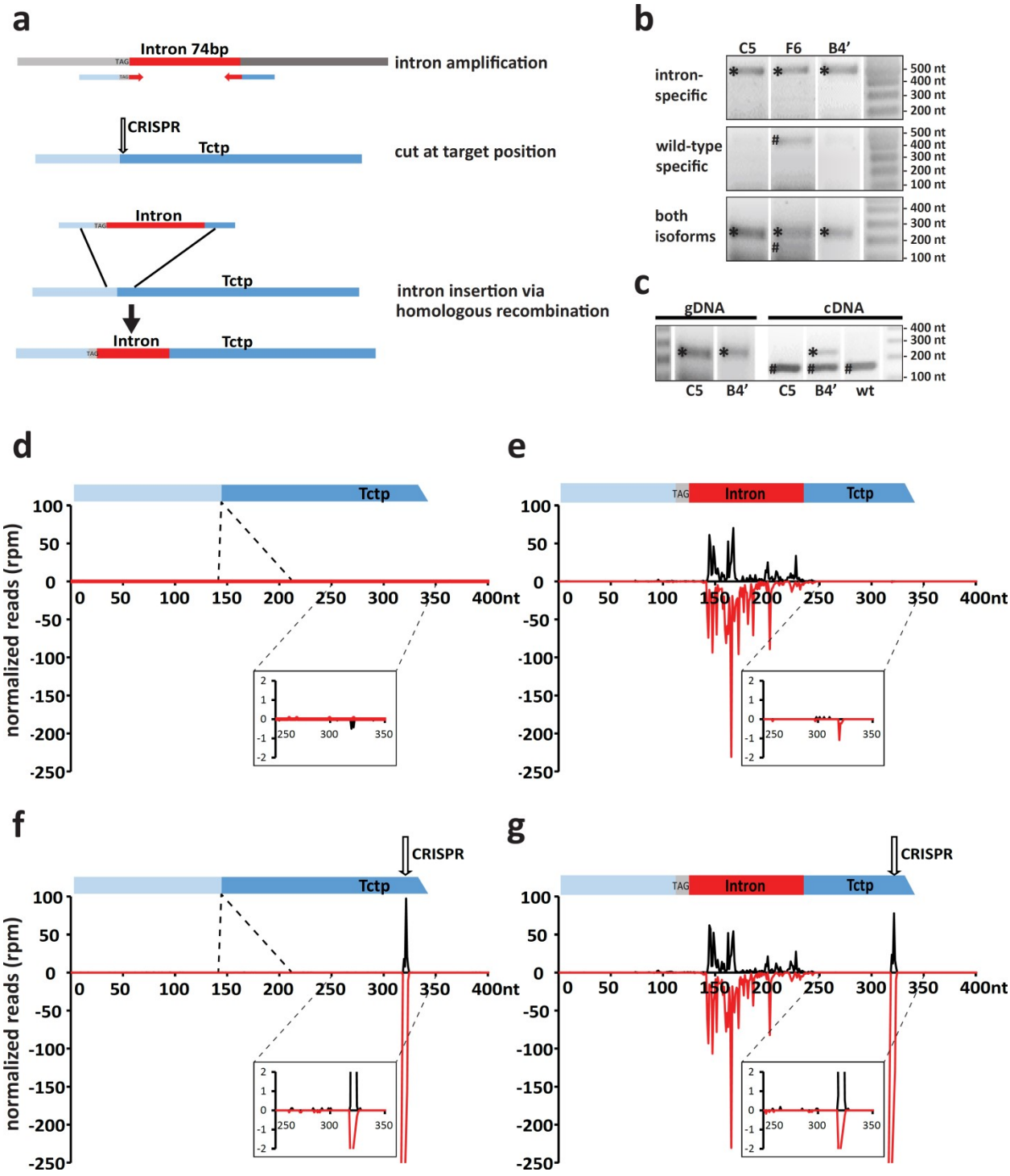

Figure 2-22: Successful integration of an endogenous intron at a different locus within the genome of *Drosophila* cells provokes a strong siRNA response to this newly integrated intron. a) Schematic workflow of intron integration into *Tctp:* The 74bp-intron of *GAPDH2* was amplified with primers carrying extensions homologous to the target position in *Tctp*. A CRISPR sgRNA was then co-transfected with the PCR product into Cas9-expressing *Drosophila* cells to enable integration of the intron via homologous recombination repair of the induced DSB. b) Test-PCRs on cell clones C5, F6 (both clones with correct splice sites) and B4' (unfunctional splice site) reveal that the intron is successfully integrated into the target locus. The three different utilized primer pairs show further that in contrast to clone F6, the clones C5 and B4' do not carry any remaining wildtype alleles of *Tctp*. c) Test-PCRs with primers that amplify both, the intronless and intron-containing *Tctp* locus, were conducted on gDNA and cDNA to evaluate the splicing-efficiency of the integrated intron. d) Deep-sequencing results of the uncut sample without intron integration were normalized on microRNA and transposon-mapping reads and mapped onto the intron-containing *Tctp* locus. The sequence not present in the wildtype locus is visible by comparison to the corresponding cartoon. Sense reads are shown in black (5'-end of siRNA mapped), antisense reads (3'-end mapped) in red. e) Deep-sequencing results of the uncut sample of clone C5 carrying the integrated *GAPDH2*-intron. Normalization and mapping was carried out as in d) and the cartoon indicates the relevant gene structure of the target locus. f) Normalized deep-sequencing results of the wildtype sample cleaved

with CRISPR 202 in *Tctp*. The zoom-in shows the region where the break-induced siRNA-response is expected. g) Normalized deep-sequencing results of the intron-containing clone C5 cleaved downstream of the intron with CRISPR 202 in *Tctp*. The sample was normalized in analogy to the uncut sample and the zoom-in shows the region where the break-induced siRNA-response is expected.

2.7.2 The newly integrated intron provokes an siRNA response

In contrast to wildtype cells (Figure 2-22d), even the uncut intron-containing cell clone revealed tremendous amounts of small RNAs mapping to the inserted intron sequence as shown in Figure 2-22e. This cellular response to (marker-free) intron integration is roughly 50 times stronger than the siRNA induction observed for DNA cleavage at intron-containing genes. The same results are obtained for the cleaved samples as depicted in Figure 2-22f and g. A closer look at the region between the strong intron-induced small RNA coverage and the actual cleavage site, where cleavage-induced siRNAs should be observed, reveals no increased small RNA levels in the cleaved intron-containing (g) compared to the uncleaved (d,e) or cut wildtype (f) samples. The strength of the intron-induced small RNA response prior to cleavage, however, might in this setting prevent the formation or detection of a supposedly much weaker response to the introduced double-strand break. Furthermore, the large amount of siRNA generated upon intron integration suggests that by copying an endogenous intron into a different, naturally intronless locus, a strong cellular response is triggered which most probably affects mRNA biogenesis at the target locus. The obtained results could thus not directly be compared to the wild-type situation as originally intended.

To further investigate the nature of the intron-induced small RNA response, the read distributions at both, the original *GAPDH2* and the target *Tctp* gene were analyzed in more detail. In the wildtype situation depicted in Figure 2-23a/c, no small RNAs mapping to the *GAPDH2* intron can be detected. After introduction of the intron into *Tctp*, reads mapping to this intron could be matched to both the original sequence and the modified *Tctp* locus and the origin remains undetermined. However, when being mapped the *GAPDH2* locus, the reads are sharply restricted to the intron sequence and do not extend to the surrounding sequence specific for *GAPDH2* as shown in Figure 2-23b/d. In contrast, the detailed mapping of the obtained reads to the modified *Tctp* locus in Figure 2-23d reveals that both sense and antisense reads not only cover the exact position of the intron, but also extend further into the region up to 30nt downstream of the intron. This region is specific for the target locus *Tctp* but also part of the HR donor DNA (length of downstream homology region is 59nt). The fact that these small RNAs span the newly created intron-exon junction suggest the small RNAs in the uncut intron-containing sample to indeed originate from the modified *Tctp* locus or possibly an additional integration site of the HR donor PCR product at a different genomic location via end-joining rather than the natural *GAPDH2* locus. In addition to the sense and antisense orientation, the intron-specific small RNA reads demonstrate a narrow size distribution with a preferential length of 21nt, thus displaying the characteristic features of endo-siRNAs. Taken together, the introduction of an endogenous intron at a different locus via genome editing seems to induce a strong siRNA response against the newly introduced intron.

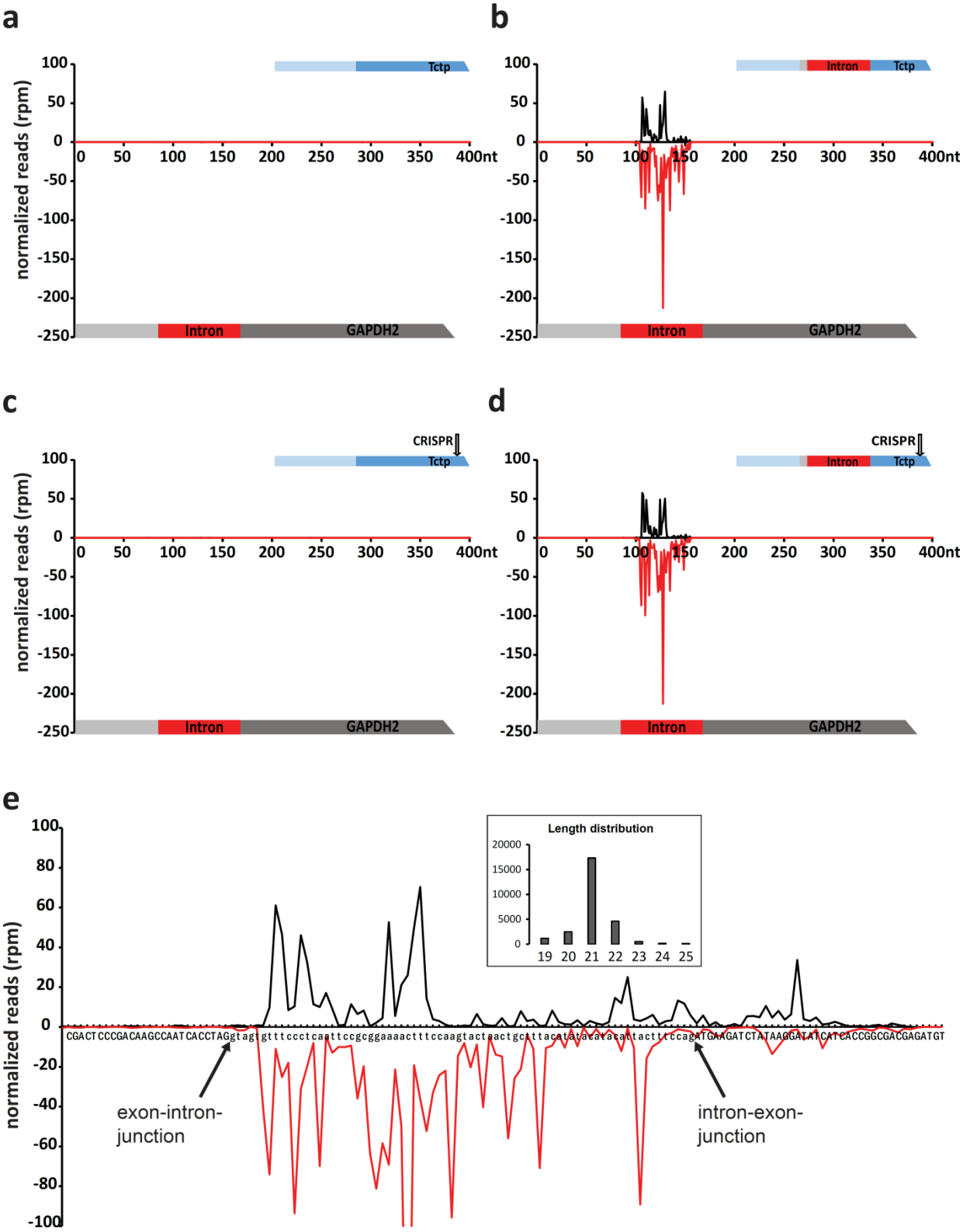

Figure 2-23: Small RNAs mapping to the *GAPDH2* intron are only detected after integration of the intron into *Tctp* with the observed siRNAs covering the intron and extending into the adjacent *Tctp* locus. Deep-sequencing results were normalized on microRNA and transposon-mapping reads and mapped onto the *GAPDH2* locus schematically depicted below each graph. Sense reads are shown in black (5'-end of small RNA mapped), antisense reads (3'-end mapped) in red. The cartoons in the upper right corners (a-d) illustrate the characteristics of the sample concerning the *Tctp* locus. a) No siRNAs are detected at the *GAPDH2* locus in the uncut intronless sample. b) Small RNAs cover the *GAPDH2* intron when the intron was additionally

integrated into *Tctp*. c) No siRNAs are detected at the *GAPDH2* locus in the intronless sample cleaved at the *Tctp* locus. d) Small RNAs cover the *GAPDH2* intron when both intron integration and Cas9-mediated cleavage was performed at the *Tctp* locus. e) Normalized sense and antisense reads (black/red) of the uncleaved intron-containing cell line (clone C5) were mapped to the nucleotide-sequence at the integration site. 5' and 3' splice sites are indicated with arrows. The depicted length distribution is based on the small RNA reads mapping to the integrated intron.

2.8 Is the number of cleavage-induced siRNAs reduced in absence of certain splicing factors?

In the validated results from the genome-wide screen presented in chapter 2.1, the knockdown of several splicing factors leads to a de-repression of the endo-siRNA reporter, thus suggesting a reduction of the double-strand break induced small RNA response in the absence of these factors. In order to directly demonstrate this effect in deep-sequencing experiments, I attempted to create clonal knockout cell-lines of the SR protein kinase Darkener of apricot (Doa) involved in alternative splicing, a very prominent positive hit in the screen, via CRISPR-mediated genome editing. Unfortunately, no clones with a complete or conditional knockout or disruption of all *Doa* alleles could be obtained. Most likely, the desired knockouts are simply not viable as it has been shown for homozygous *Doa* knockout flies (Rabinow and Birchler 1989). Therefore, the cells were subjected to RNAi against *Doa* prior to small RNA sequencing. This procedure, however, results in a large and highly variable fraction (20-60%) of the sequenced small RNAs originating from the knockdown constructs thus partly compromising the reproducibility between several replicates. The data shown in Figure 2-24 was thus obtained using reduced amounts of knockdown siRNA in comparison to the reporter-based screen. Some of the knockdowns for this experiment were kindly performed by R. Böttcher from AG Förstemann.

As depicted in Figure 2-24a, the combined siRNAs from two biological replicates each, clearly show reduced siRNA levels for the cleaved CG15098 locus upon *Doa* compared to control knockdown. Considering their catalytic mode of action, the observed 2-fold reduction of the siRNAs is in the range expected from the corresponding 3- to 5-fold de-repression of the plasmid-based reporter. Unfortunately, a second experiment, identically comprising two independent samples for each knockdown, could not confirm this 2-fold reduction as shown in Figure 2-24b. The quantification of the obtained siRNAs and ratios of reads upstream/downstream of the induced break (Figure 2-24c and d) reveals that in both experiments, as well as a third identical experiment of much lower sequencing depth and sample quality, the siRNA response upon cleavage of the target gene is reduced in the *Doa* knockdown background. However, the strength of the effect considerably varies between a clear 2-fold reduction and a very slight reduction, thus rendering the overall experiment inconclusive in terms of a significant and reproducibly detectable reduction of siRNAs levels upon knockdown of this splicing factor.

In addition to quantitative changes, *Doa* knockdown could possibly also affect the break-induced siRNA response in a qualitative way. The direct overlay in Figure 2-24e though shows that even in replicates with clearly reduced small RNA level for *Doa* compared to control knockdown, the siRNA distribution along the target gene is identical for both samples. Furthermore, the length distribution, which is

important for the loading and thus the functionality of endo-siRNAs, remains the same as illustrated in Figure 2-24f.

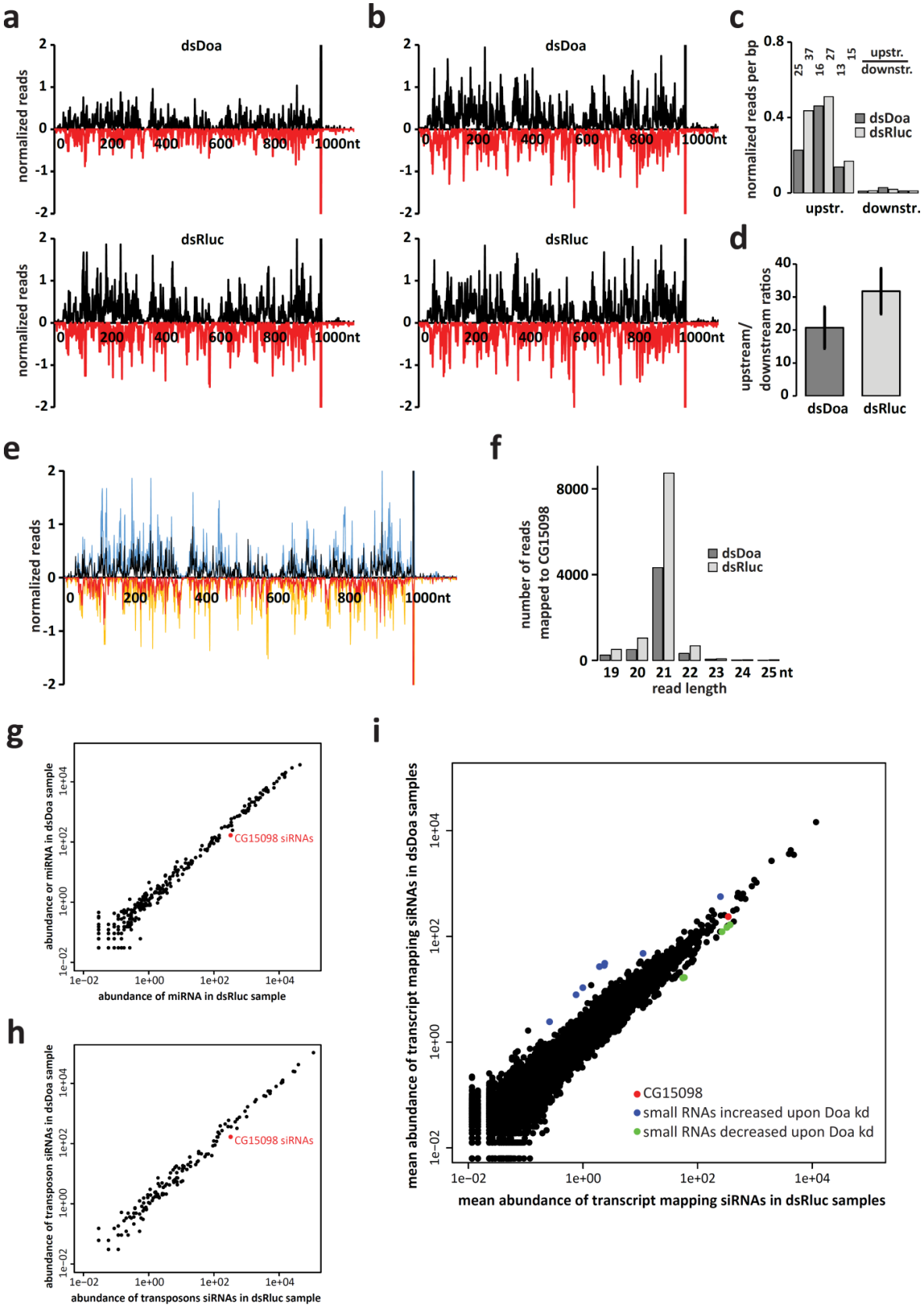

Figure 2-24: The break-induced siRNA response is reduced but not changed with respect to distribution pattern or small RNA characteristics upon knockdown of the splicing factor Doa. Cas9-expressing cells were subjected to either *Doa* or

control knockdown (Rluc) and transfected with CRISPR 207 sgRNA template to facilitate cleavage in *CG15098*. Four days after transfection, small RNAs were analyzed via deep-sequencing and normalized on *Drosophila*-mapping reads without those mapping to *Doa*. Sense-oriented reads (black/blue) are mapped to the 5'-end and antisense-oriented small RNAs (red/yellow) to the 3'-end of the siRNA. a) Two replicates for each, the dsDoa and dsRluc knockdown, were averaged and the sense and antisense reads mapping to the cleaved target gene *CG15098* are depicted showing a clearly reduced siRNA response upon knockdown of *Doa*. b) A second, identical experiment cannot reproduce the strong reduction shown in a) and reveals almost equal small RNA levels for *Doa* and control knockdowns upon cleavage in *CG15098*. c) Small RNA coverage per basepair upstream and downstream of the induced break as well as the corresponding upstream/downstream ratio quantifying the small RNA response to the double-strand break are depicted for three independent experiments consisting of two replicates each for *Doa* and control knockdown. d) The averaged quantification of the small RNA response via the upstream/downstream is depicted for the two experiments shown in a) and b). Error bars indicate the standard deviation of the four replicates of the respective knockdown. e) Despite the quantitative differences depicted in a), the overlay of the small RNA coverage in a dsDoa (black/red) and dsRluc (blue/yellow) background reveals identical distribution patterns. f) The depicted length distribution of small RNAs mapping to *CG15098* upstream of the induced break (experiment shown in a)) reveals no difference between the *Doa* and Rluc knockdown. g) The abundance of microRNAs was normalized to the reads of the specific library mapping to *D. melanogaster* after excluding the reads mapping to the knockdown target *Doa* and are displayed in reads per million. The dsDoa and dsRluc samples displayed in a) are compared and reveal no difference in microRNA levels. The reads mapping to the cleavage target *CG15098* (CRISPR-mapping reads excluded) are indicated in red for comparison. h) Normalized transposon-mapping reads are compared for *Doa* and control knockdowns and displayed in reads per million. The samples reveal only minor differences regarding the numbers of siRNAs mapping to each specific transposon. Reads mapping to the cleavage target *CG15098* (CRISPR-mapping reads excluded) are indicated in red for comparison. i) Normalized transcript-mapping small RNAs were averaged from the two experiments displayed in a) and b) and are compared for *Doa* and control knockdown samples. The cleaved *CG15098* locus is depicted in red and example transcripts that show increased small RNA coverage upon knockdown of *Doa* are colored in blue, transcripts with decreased coverage in green.

As a next step, I checked whether the levels of other (small) RNAs in the cells are influenced by the *Doa* knockdown. Neither miRNAs nor transposon mapping endo-siRNAs exhibit any obvious quantitative changes as depicted in Figure 2-24 g and h. Although this is also true for the number of reads mapping to most transcripts (comprising of both, siRNAs and mRNA degradation products of the selected size), a small number of transcripts shows either increased or decreased levels of small RNA reads exceeding the effect observed for the cleaved target gene *CG15098*. The marked transcript reads increased upon knockdown mostly belong to certain proteins involved in proteolysis, possibly connected to resulting viability effects. In contrast, the transcripts with decreased numbers of small RNAs cannot be attributed to proteins with a common molecular function or biological process. However, the fact that all of the marked transcripts belong to proteins subjected to alternative splicing implies that the observed reduced small RNA levels of certain transcripts upon knockdown of *Doa* might be associated with its known role in the regulation of alternative splicing.

Taken together, the above experiments suggest, but due to technical reasons cannot prove, a quantitative reduction without any obvious qualitative changes of break-induced endo-siRNAs in absence of the splicing factor Doa.

2.9 Do Doa and l(1)10Bb interact thus suggesting a certain molecular mechanism?

As one of the most apparent and effective positive hits in the genome-wide screen, Doa, a dual-specificity kinase of the Lammer kinase family (Yun et al. 1994) involved in the regulation of alternative splicing was a promising candidate to check for potential interaction partners and thus gain further insight into the molecular mechanism underlying the splicing-dependent siRNA response to DNA double-strand breaks and high-copy loci. In addition to specific and essential roles in alternative splicing and *Drosophila* sex-determination (Zhao et al. 2015; Du et al. 1998), Doa also phosphorylates EF1γ and thus inhibits transport

of membrane organelles (Fan et al. 2010; Serpinskaya et al. 2014). Furthermore, mutations in Doa lead to changes in transcription of the *copia* transposon and *copia*-induced mutations within the *white* locus (Rabinow, Chiang, and Birchler 1993), and cause over-proliferation resulting from failed mitosis to meiosis-transition (Zhao et al. 2013).

The several different isoforms of *Doa* are expressed in both, cytoplasm and nucleus (Yun et al. 2000) and can partly be ascribed to specific functions (Kpebe and Rabinow 2008; Rabinow and Samson 2010). Published sequencing data (Brown et al. 2014; Graveley et al. 2011), as well as additional experiments performed during my Master thesis (Merk 2013) with different knockdown constructs and site-specific PCR to verify transcript expression in the utilized S2 cells suggest, that the effect on siRNA production can be attributed to a number of similar Doa isoforms with a molecular weight between 60 and 68 kDa, whereas the larger isoforms of Doa can be excluded.

Phosphorylation by the dual-specificity kinase Doa most frequently takes place on serine residues (Nikolakaki et al. 2002), for example in the unstructured serine-arginine-stretches close to the N-terminus of SR-proteins which are often targeted by Lammer kinases in different species (Lee et al. 1996). Furthermore, structure models suggest that the phosphorylated residues are preferably found in a basic surrounding (Farkaš et al. 2014). Besides Doa, the Bud31 homologue l(1)10Bb, a highly conserved component of the Prp19-related complex (Herold et al. 2009) triggered a remarkably strong de-repression of the siRNA-reporters thus raising the question whether this factor could be the substrate of the Doa kinase within the pre-catalytic spliceosome. The 18 kDa protein indeed shows potential target sites, with the most prominent one being a serine close to the N-terminus surrounded by arginines.

To check this potential protein-protein interaction via co-immunoprecipitation, Doa (C-terminus, Flag-tag) and l(1)10Bb (C- and N-terminus, different tags) were endogenously tagged in S2 cells using the CRISPR/Cas9 nuclease in combination with PCR-based donors for homologous recombination (Bottcher et al. 2014).

As depicted in Figure 2-25a, the suspected interaction could not be shown when C-terminally tagged Doa was immunoprecipitated. Since DNA damage could be a prerequisite for phosphorylation of l(1)10Bb or interactions between Doa and l(1)10Bb in general, the S2 cells were then treated with camptothecin (CPT) prior to protein extraction. Camptothecin blocks DNA topoisomerase I thus causing DNA damage through collisions between advancing replication forks and the trapped enzyme. At higher concentrations, camptothecin also induces replication-independent DNA damage. In Figure 2-25b and c, depicting attempted co-immunoprecipitations of Doa with Strep- and Gfp-tagged l(1)10Bb, the interaction between both proteins is not visible neither in absence, nor at low or high CPT-concentrations. As a next possibility, N-terminally tagged l(1)10Bb was used to investigate the potential interaction between Doa and l(1)10Bb. In addition to avoiding artefacts that might result from the tag at the C-terminus, the N-terminal tagging procedure also facilitates the introduction of specific mutations at the potential Doa

target site near the N-terminus of l(1)10Bb. While a Ser to Asp could unfortunately not be established, the N-terminally tagged l(1)10Bb carrying a single Ser to Ala mutation does not display any detectable interaction with Doa as shown in Figure 2-25d.

Another conceivable interaction partner of the Doa kinase is Rm62. This DEAD-box helicase has several different transcripts and can bind both pri-miRNA and viral RNA to facilitate processing (Moy et al. 2014). It is also a regulator of alternative splicing (Herold et al. 2009; Park et al. 2004) and was found in a complex with the RNA binding protein FMR1, ribosomal proteins and the core RNAi factor Ago2 (Ishizuka, Siomi, and Siomi 2002). Furthermore, it was shown to interact with a complex containing blanks (Gerbasi et al. 2011), a nuclear protein supposedly involved in RNAi which was also among the prominent hits in the screen. As shown in Figure 2-25e, however, no interaction between Doa and Rm62 could be detected.

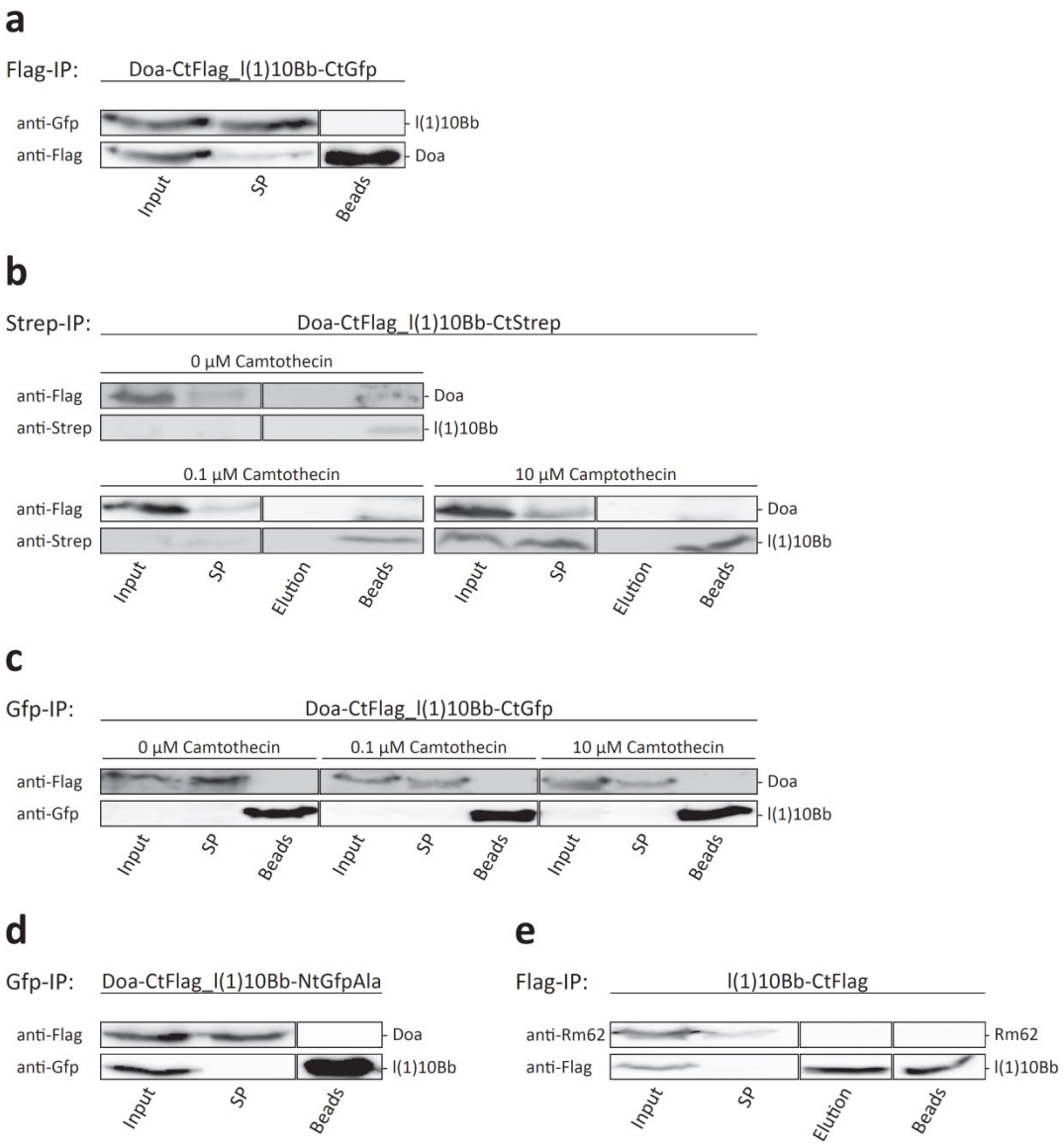

Figure 2-25: Co-immunoprecipitation experiments do not reveal any direct interactions between Doa and l(1)10Bb. Protein extracts from cell lines with the indicated endogenously tagged Doa/l(1)10Bb were subjected to immunoprecipitation

with appropriate beads and the elution/boiled beads were subjected to analysis via Western blotting. a) Western blot of Doa-CtFlag_l(1)10Bb-CtGfp cells upon protein-immunoprecipitation with Flag-beads. Conditions: 2.5 mg protein input; 20 µL uncrosslinked, pre-incubated beads; α-Flag-HRP 1:10000; α-Gfp/α-mouse 1:10000; control: l(1)10-CtFlag. b) Western blot of Doa-CtFlag_l(1)10Bb-CtStrep cells upon protein-immunoprecipitation with Strep-beads. The indicated amounts of camptothecin were added to the cells 24 hours prior to protein extraction. Conditions: 1.5 mg protein input; 20 µL beads blocked with avidin; α-Strep-HRP 1:5000; α-Flag/α-mouse 1:10000; control: Dcr-2-CtStrep_R2D2-CtFlag. c) Western blot of Doa-CtFlag_l(1)10Bb-CtGfp cells upon protein-immunoprecipitation with Gfp-beads. The indicated amounts of camptothecin were added to the cells 24 hours prior to protein extraction. Conditions: 1.5 mg protein input; 25 µL beads; α-Flag/α-mouse 1:10000; α-Gfp/α-mouse 1:10000; control: untagged cells, Doa-CtFlag_l(1)10BbNtGfpAla. d) Western blot of Doa-CtFlag_l(1)10Bb-NtGfpAla cells upon protein-immunoprecipitation with Gfp-beads. The growth medium was supplemented with 200 µM Cu^{2+} to induce l(1)10Bb-NtGfpAla expression. Conditions: 1.5 mg protein input; 25 µL beads; α-Flag/α-mouse 1:10000; α-Gfp/α-mouse 1:10000; control: untagged cells, Doa-CtFlag_l(1)10BbNtGfp. e) Western blot of Doa-CtFlag cells upon protein-immunoprecipitation with Flag-beads. Conditions: 1.5 mg protein input; 25 µL crosslinked Flag-beads; α-Flag/α-mouse 1:10000; α-Rm62/α-rat 1:5000; control: Dcr-2CtStrep_R2D2CtFlag.

3 Discussion

3.1 Validation screens show distinct and common results between reporters

A genome-wide RNAi screen for factors involved in the small RNA response to DNA double-strand breaks reveals a great variety of validated hits and processes all the way from the initial recognition of the double-strand break to the loading of the functional siRNAs into Ago2. The additional validation with the high-copy number reporters and the comparison between the reporters show, that some factors such as the MRN complex and the sliding-clamp loading RfC-complex are specific for the formation of DSB-related siRNAs and could therefore be involved in either the recognition of the break or in a process that happens upstream of dsRNA precursor formation. In contrast, splicing factors are highly enriched in both the positive hits for the DSB-induced siRNA reporter and the high-copy reporter. This suggests that the splicing machinery is not only involved in translating DNA double-strand breaks into an siRNA-signal, instead a more general mechanism for the formation of endo-siRNAs in response to transcription obstacles such as high-copy loci and DNA damage supported by the splicing machinery seems likely.

3.2 Direct introduction of DSBs in *Drosophila* cells triggers siRNA formation

The more direct readout compared to the plasmid-based reporter for siRNA formation at a double-strand break is the cleavage of the DNA double-strand by a targeted nuclease. In *Drosophila* cells stably expressing the Cas9 endonuclease, the endo-siRNA formation from the reporter setting was reproduced upon cleavage in an actively transcribed gene downstream of an intron and the siRNAs were directly detected and characterized by deep-sequencing. Furthermore, the dependency on transcription levels and intron structure was extensively studied upon selection of defined target sites. In a similar setting in human and mouse cells (Bonath et al. 2018) as well as in *A. thaliana* (Miki et al. 2017) the small RNA production observed for reporters could not be reproduced upon endonucleolytic cleavage in unique genes. The cleavage in repetitive loci or highly expressed transgenes, however, did trigger detectable siRNA formation in these organisms thus suggesting that either the siRNA response to DSB is much stronger in *Drosophila* cells compared to for example human cells or that a DSB introduced by a nuclease is indeed sufficient for siRNA formation in *Drosophila* whereas additional conditions need to add up to trigger small RNA formation in other organisms.

3.3 A model for the endo-siRNA formation at DNA double-strand breaks

The validated results from the genome-wide screen and the deep-sequencing as well as additional (e.g. qPCR and Western blot) experiments merge into the model for spliceosome-stimulated endo-siRNA formation at double-strand breaks depicted in Figure 3-1.

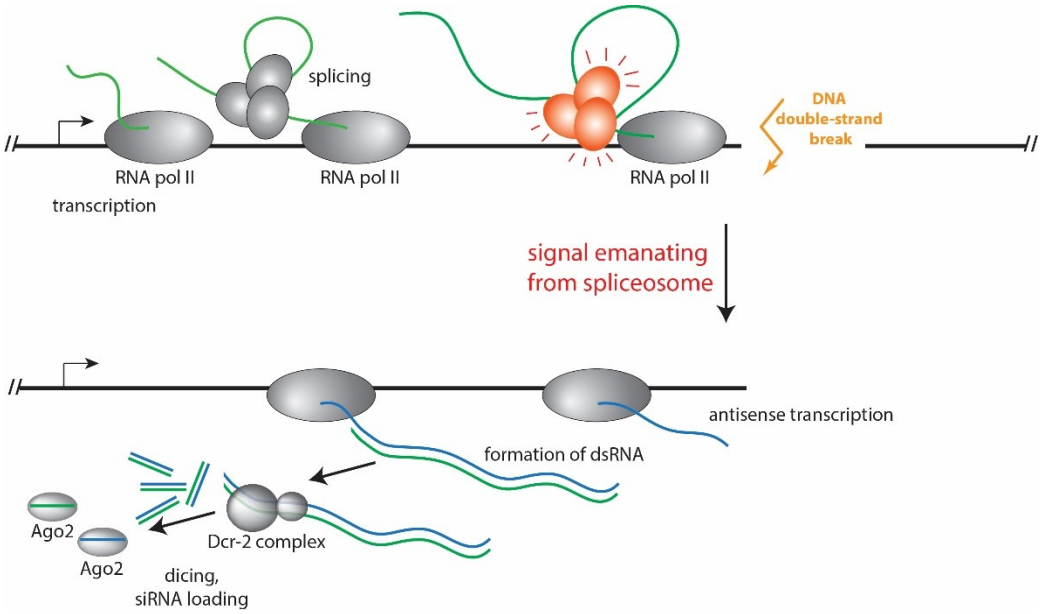

Figure 3-1: A model for the events leading to DSB-induced endo-siRNA formation in *Drosophila*: When the RNA polymerase reaches the DNA double-strand break, transcriptional stalling, possibly caused or accompanied by R-loop formation, occurs and the co-transcriptional splicing process is impacted. The pre-catalytic stalling of the spliceosome induces a signal leading to antisense transcription by RNA polymerase II from the break to the transcription start site. The antisense transcript pairs with the unspliced sense transcript and the resulting dsRNA is cleaved by Dcr-2 into 21nt endo-siRNAs which are then loaded onto Ago2. **Model published in (Merk et al. 2017)**

The different events leading to siRNA formation are illustrated in the model and will be discussed in the following paragraphs.

3.4 The MRN complex acts upstream of the siRNA response

The initial steps at the DNA double-strand break, namely recognition, initial signaling and limited end-resection by the MRN-complex, are independent and happen prior to siRNA formation. The same is observed for the equivalent process in human cells (Francia et al. 2016). In the validated genome-wide screen the MRN-complex as well as the CtIP nuclease but not any actual repair factors involved in later steps of DNA repair appear to stimulate siRNA formation. This is consistent with the observation that the siRNA response covers the whole area between the transcription start site and the break thus suggesting that antisense transcription must initiate at the break ruling out or possibly even preventing extensive end-resection.

3.5 Transcription and R-loop formation at the break

At actively transcribed genes, stalling of the transcription machinery at obstacles such as DNA damage including double-strand breaks can lead to extensive R-loop formation. The nascent RNA pairs with the template DNA thus leaving the non-template strand behind as single-stranded DNA. The co-transcriptional splicing process is also stalled as a consequence of transcriptional stalling. This can either lead to a defined signal through structural changes, or the R-loop structure can persist and function as a

signal. In addition, the single stranded DNA can either be subject to initiation of antisense transcription or can be bound by the ss-DNA binding protein RPA (Sikorski et al. 2011).

A similar process has been observed for transcription-blocking lesions (not double-strand breaks) where the removal of the spliceosome from the transcript leads to R-loop formation between the freed RNA and the unwound DNA thus triggering ATM signaling and changes in alternative splicing (Tresini, Marteijn, and Vermeulen 2016).

3.6 Stalling of the spliceosome in a pre-catalytic state

The results from the Cas9-cleavage-induced small RNA deep-sequencing experiments emphasize that in addition to being transcribed, the targeted gene needs an intron to trigger break-derived small RNA formation. Furthermore, the inserted lesion needs to be at least a minimum distance downstream of the first intron for proper siRNA biogenesis induction. This is consistent with the fact, that even though splicing occurs co-transcriptionally, the transcribing polymerase needs to move on away from the intron for the spliceosome to be able to bind to the nascent RNA.

It has been shown for *S. cerevisiae* that the absence of an intron leads to increased R-loop formation and genomic instability (Bonnet et al. 2017). The insertion of an intron into the intronless gene, however, rescues the phenotype. I also tried to directly compare the small RNA production at the same gene with and without an intron by introducing an endogenous intron from a gene of similar length and transcription level into the chromosomal location of an intronless gene. However, this insertion - even in absence of a DNA double strand break - gave rise to an extensive siRNA response thus rendering it impossible to check for the much less abundant break-derived siRNAs. While this is nonetheless a very interesting observation at first sight, it was impossible to unambiguously assign the precise genomic origin of the corresponding reads. Further validations are thus required to substantiate the conclusion.

Similar to the here proposed model, stalled spliceosome-dependent pathways have been described for other organisms. In budding yeast, intronless genes are subject to degradation via spliceosome-mediated decay triggered by non-productive association with the spliceosome (Volanakis et al. 2013). Another process connecting impaired splicing to RNAi is transposon suppression by the SCANR complex in *C. neoformans* (Dumesic et al. 2013). Hereby, transposon transcripts are distinguished from other transcripts via their suboptimal introns causing stalling of the spliceosome after the first catalytic step followed by processing of the splicing intermediate into siRNAs. In contrast, correctly spliced transcripts are for example prevented from entering the piRNA biogenesis pathway (Zhang, Wang, et al. 2014).

In fission yeast, splicing has been shown to be essential for small RNA-mediated heterochromatin silencing (Bayne et al. 2008). However, the RNAi effector proteins themselves are often subject to splicing thus raising the question whether splicing factors are indeed involved in a certain process or whether they are instead required for the correct splicing of the actual effector proteins. For the siRNA

biogenesis factors in *Drosophila* cells, the knockdown of selected splicing factors did not impact the protein levels nor were the transcript levels of certain genomic loci changed.

Several results indicate that the spliceosome is being stalled in a pre-catalytic state. First of all, although splicing factors of various different complexes are among the factors promoting endo-siRNA formation, the components of the Prp19/Prp19-related complex which functions in activating the B complex prior to the first catalytic step are particularly enriched among the validated positive hits from the genome-wide screen. Second, there are no exon-exon junction spanning small RNAs and the observed small RNAs equally map to intronic and exonic regions of the target gene. This indicates that the transcript was not spliced prior to dsRNA formation.

3.7 Antisense transcription initiates at the double-strand break

In response to the DNA double-strand break and the subsequent stalling of the spliceosome, antisense transcription must be initiated at the double-strand break. As both, sense and antisense small RNA reads, match to the complete area between the transcription start site and the double-strand break, antisense transcription must initiate directly at the break. Both the intronic and exonic regions are equally covered by small RNA reads thus suggesting that the unspliced transcript pairs with the antisense transcript to form the dsRNA precursor. Furthermore, the fact that the antisense reads cover exonic and intronic regions and there are no exon-exon-spanning reads excludes the possibility that the spliced transcript could serve as a template for a potential RNA-dependent RNA polymerase. Moreover, in contrast to *N. crassa* (Lee et al. 2009), *A. thaliana* (Wei et al. 2012) and rice (Chen et al. 2013) there is no RNA-dependent RNA polymerase in *Drosophila*. Recently published NET-seq data show that in *Drosophila* the antisense transcription initiating at the DNA double-strand break is performed by RNA polymerase II (Böttcher et al. 2021). The NET-seq data also confirm the information derived from the distribution of small dsRNAs that antisense transcription starts at the break and extends all the way to the transcription start site. For human cells, it has recently been observed that antisense transcription by RNA polymerase III is initialized at resected DNA double-strand breaks (Liu et al. 2021). The obtained RNA molecule hybridizes with the single-stranded DNA overhang thus protecting it from further resection. In contrast, antisense transcription of long RNA precursors at DNA damage was reported to be performed by Pol II in human cells (Michelini et al. 2017). While the NET-seq data (Böttcher et al. 2021) does not indicate any Pol III antisense transcription at the selected loci, the here presented small RNA deep-sequencing data of the cut in the Pol III transcribed 7SK snRNA locus show sense and antisense small RNA reads. The dsRNA precursor for these small dsRNAs is thus either a hybrid of a Pol II and a Pol III transcript, or the antisense transcription might in this case be performed by Pol III as described for human cells.

3.8 Processing of dsRNA via the endo-siRNA pathway

The long dsRNA consisting of the unspliced transcript and the complementary Pol II derived antisense transcript is then processed by the classical endo-siRNA pathway. Besides Dcr-2 and Ago2 as main effector proteins and the dsRBP loquacious as co-factor of Dcr-2 (non-isoform specific knockdown construct), the dsRBP blanks which was described to shuttle double-stranded non-coding RNAs between the nucleus and the cytoplasm (Nitschko et al. 2020) was identified among the factors promoting DSB-induced small RNA biogenesis. It can thus be assumed that the dsRNA is exported to the cytoplasm by blanks and processed into perfectly complementary 21nt dsRNAs by Dcr-2 with loquacious as co-factor. The dsRNAs are then loaded onto Ago2.

3.9 Possible functions of DSB-derives siRNAs

A potential role of DSB-derived small dsRNAs in DSB repair has been described for various organisms (Francia et al. 2016; Francia et al. 2012; Wei et al. 2012; Gao et al. 2014; Lee et al. 2009). However, many observations depend on the role of Dicer and Argonaute proteins which are either involved in different pathway such as miRNA and siRNA pathways or have several redundant homologues making it difficult to dissect the individual functions. In *Drosophila*, mostly separate pathways exist for microRNAs and siRNAs. Knockout experiments with DSB repair reporters show, that homologous recombination repair is still fully functional in the absence of Dcr-2 (Schmidts et al. 2016). Instead of the functional siRNAs, the processing of the transcript into small RNAs or the induced antisense transcription could function in DSB repair for example by regulating the accessibility of ssDNA for processing factors (Ohle et al. 2016). RNA synthesis, binding and degradation can influence the formation and persistence of R-loops which in turn can play a role in both DSB formation and repair. In *S. pombe,* collisions between the RNA polymerase and the replication machinery induce siRNA formation via an RdRP. This leads to the release of Pol II thus preventing extensive R-loop formation (Zaratiegui et al. 2011). Furthermore, the DSB-induced dsRNAs can also function in transcript surveillance and degradation instead of DNA repair as double-strand breaks are only one of several triggers of a common biogenesis pathway. Especially the endo-siRNAs generated at high-copy loci show many similarities to DSB-derived small RNAs.

3.10 More siRNAs independent of introduced DNA double-strand breaks

In addition to the investigated double-strand break derived siRNAs, other in most cases much more abundant siRNAs were present in the deep-sequencing libraries. The transfection of either S2 or Cas9-expressing S2 cells with the U6-sgRNA template PCR product induces a very strong siRNA response around 500 to 1000-fold higher at the peak position compared to DSB-induced siRNAs. These 21nt dsRNAs are not equally distributed along the sgRNA but display a distinct pattern correlating with the secondary structure of the sgRNA and thus possibly with the stability of the sgRNA or the accessibility

for deep-sequencing primer ligation. Although the small RNA reads do not cover the U6-sgRNA template all the way to the end of the PCR product, a possible effect of the double-stranded DNA end could further be investigated by preparing deep-sequencing libraries after CRISPR sgRNA transfection.

More strikingly, the marker-free integration of an endogenous intron into a gene of similar length and expression strength led to the production of 21nt dsRNAs spanning the whole length of the intron, even extending into the surrounding region. The sense-antisense distribution is not symmetric and there is one strong peak. Apart from this peak the RNA numbers are roughly 10 to 20-fold higher compared to the DSB-induced siRNAs. A similar response, however more symmetric and weaker, has been observed during protein tagging with selection markers (Kunzelmann and Forstemann 2017). However, these siRNAs disappear upon removal of the selection marker and can be avoided when no selection marker is used. As the intron was introduced in a marker-free procedure, and the single clone cell lines had been shown to be splicing competent and were used for deep-sequencing experiments several weeks after intron induction, the actual trigger and the unambiguous assignment of the genomic origin for this small RNA response remains unclear.

3.11 Conclusion and outlook

Taken together, the validation project following the genome-wide screen for factors involved in the siRNA response to DNA double-strand breaks has yielded many new insights into the processes and proteins between DNA repair, co-transcriptional splicing and RNAi. Especially the question of how a stalled spliceosome can function as a sensor for double-strand breaks and how this stalling can then be translated into a small RNA response was elucidated a bit further through the here presented data. However, the validated results from the genome-wide screen contain many additional factors beyond splicing for which the function within the process still remains unknown and can be further explored.

4 Materials and Methods

4.1 Materials

4.1.1 Laboratory equipment

Agarose gel running chamber	Carl Roth; Karlsruhe, Germany
Amersham Imager 600	GE Healthcare; Chalfont St Giles, UK
BioPhotometer	Eppendorf; Hamburg, Germany
Bioruptor Next Gen	Diagenode; Liege, Belgium
Centrifuge 5417R , 5417C, 5415R	Eppendorf; Hamburg, Germany
Centrifuge Universal 16R	Andreas Hettich; Tuttlingen, Germany
Centrifuge, Rotanta 460 R	Andreas Hettich; Tuttlingen, Germany
Desk centrifuge Sprout	Biozym; Hessisch Oldendorf, Germany
FACS Calibur flow cytometer	Becton, Dickinson; Franklin Lakes, USA
Heater	HLC BioTech; Pforzheim, Germany
Incubator	Binder; Tuttlingen, Germany
Incubator ECF-F Control	Fiocchetti; Luzzarra, Italy
Incubator Innova 43	Eppendorf; Hamburg, Germany
INTAS UV Imaging System	INTAS; Göttingen, Germany
LAS 3000 mini Western Imager	Fujifilm; Tokyo, Japan
Magnetic Stand	Life technologies; Carlsbad, USA
Magnetic Stirrer, MR 1000	Heidolph; Schwabach, Germany
Nanodrop	Thermo Fisher Scientific; Waltham, USA
Overhead incubator RS-RD5	Phoenix Instrument; Garbsen, Germany
Overhead Shaker, REAX 2	Heidolph; Schwabach, Germany
PAGE-electrophoresis chamber	BioRad; Hercules, USA
Power supply	BioRad; Hercules, USA
Roller Mixer SRT9	Stuart; Stone, UK
Shaker, Polymax 1040	Heidolph; Schwabach, Germany
SpectraMax Paradigm Plate Reader	Molecular Devices; Sunnyvale, USA
SterilGARD cell culture workbench	The Baker Company; Sanford, USA

TELEVAL 31	Zeiss; Oberkochen, Germany
Thermo cycler BIoER Life touch	BioRad; Hercules, USA
Thermocycler Sensoquest	Sensoquest; Göttingen, Germany
Thermomixer compact	Eppendorf; Hamburg, Germany
TOptical Thermocycler	Biometra; Jena, Germany
Vortex Genie 2	Scientific Industries; Bohemia, USA
Water Bath	GFL; Burgwedel, Germany

4.1.2 Laboratory chemicals

Acetic acid	Carl Roth; Karlsruhe, Germany
Agarose	Biozym; Hessisch Oldendorf, Germany
Aluminium sulfate	Carl Roth; Karlsruhe, Germany
Ammoniumperoxodisulfat (APS)	Carl Roth; Karlsruhe, Germany
Ampicillin	Carl Roth; Karlsruhe, Germany
Avidin	IBA; Göttingen, Germany
Bacto Yeast Extract	Becton, Dickinson; Franklin Lakes, USA
Blasticidin HCL	Thermo Fisher Scientific; Waltham, USA
Boric acid	Carl Roth; Karlsruhe, Germany
Bromophenol blue	Sigma-Aldrich; St. Louis, USA
Bovine Serum Albumin (BSA)	Jackson ImmunoResearch; West Grove, USA
Camptothecin	Sigma-Aldrich; St. Louis, USA
Chloroform	Merck; Darmstadt, Germany
Coomassie Brilliant Blue G250	Carl Roth; Karlsruhe, Germany
Coomassie Brilliant Blue R	Sigma-Aldrich; St. Louis, USA
$CuSO_4$	Sigma-Aldrich; St. Louis, USA
Dimethyl pimelimidate (DMP)	Sigma-Aldrich; St. Louis, USA
Dimethylsulfoxid (DMSO)	Carl Roth; Karlsruhe, Germany
Dithiothreitol (DTT)	Carl Roth; Karlsruhe, Germany
EDTA	AppliChem; Darmstadt, Germany
Ethanol p.a.	VWR; Radnor, USA
Formamide	Sigma-Aldrich; St. Louis, USA

Glucose	Carl Roth; Karlsruhe, Germany
Glycerol	Carl Roth; Karlsruhe, Germany
Glycine	Carl Roth; Karlsruhe, Germany
HEPES	Carl Roth; Karlsruhe, Germany
Hydrochloric acid	Carl Roth; Karlsruhe, Germany
Isopropanol	VWR; Radnor, USA
Kaliumacetate (Kac)	Merck; Darmstadt, Germany
Kanamycine	Sigma-Aldrich; St. Louis, USA
KCl	Carl Roth; Karlsruhe, Germany
KH_2PO_4	Carl Roth; Karlsruhe, Germany
$LiCl_2$	Carl Roth; Karlsruhe, Germany
Magnesium acetate tetra-hydrate (MgAc)	Sigma-Aldrich; St. Louis, USA
Methanol p.a.	Sigma-Aldrich; St. Louis, USA
$MgCl_2$	Merck; Darmstadt, Germany
$MgSO_4$	Merck; Darmstadt, Germany
Na_2HPO_4	Merck; Darmstadt, Germany
NaCl	Merck; Darmstadt, Germany
NaOH solution	Merck; Darmstadt, Germany
Phosphoric acid	Carl Roth; Karlsruhe, Germany
Puromycin-Dihydrochlorid	Carl Roth; Karlsruhe, Germany
Sodium dodecylsulfate (SDS)	Merck; Darmstadt, Germany
Sodium tetraborate ($Na_2B_4O_7$)	Honeywell; Seelze, Germany
Spermidine	Sigma-Aldrich; St. Louis, USA
TEMED	Carl Roth; Karlsruhe, Germany
Tergitol Type NP-40	Sigma-Aldrich; St. Louis, USA
Thimerosal	Sigma-Aldrich; St. Louis, USA
Tris	Carl Roth; Karlsruhe, Germany
Tris-Hydrochlorid	Carl Roth; Karlsruhe, Germany
Triton X-100	Sigma-Aldrich; St. Louis, USA
Tryptone	Carl Roth; Karlsruhe, Germany
Tween-20	Carl Roth; Karlsruhe, Germany

Urea	Carl Roth; Karlsruhe, Germany
Water, nuclease-free	Carl Roth; Karlsruhe, Germany
Xylene cyanol	Carl Roth; Karlsruhe, Germany

4.1.3 Kits

CloneJET PCR Cloning Kit	Thermo Fisher Scientific; Waltham, USA
Dual-Glo Luciferase Assay System	Promega; Madison, USA
DyNAmo Flash SYBR Green qPCR Kit	Thermo Fisher Scientific; Waltham, USA
QIAGEN Plasmid Midi/Maxi Kit	Qiagen; Hilden, Germany
QIAquick Gel Extraction Kit	Qiagen; Hilden, Germany
QIAquick PCR Purification Kit	Qiagen; Hilden, Germany
Reliaprep gDNA Tissue Miniprep System	Promega; Madison, USA
RNA Clean and Concentrator-5	Zymo Research; Irvine, USA
ZR small-RNA PAGE Recovery Kit	Zymo Research; Irvine, USA

4.1.4 Markers

PageRuler Unstained Protein Ladder	Thermo Fisher Scientific; Waltham, USA
PageRuler Prestained Protein Ladder	Thermo Fisher Scientific; Waltham, USA
GeneRuler DNA Ladder Mix	Thermo Fisher Scientific; Waltham, USA
50bp DNA Ladder	New England Biolabs; Ipswich, USA
microRNA Marker	New England Biolabs; Ipswich, USA

4.1.5 Other materials

10 mM ATP for Solexa 5' ligation	New England Biolabs; Ipswich, USA
10x *Taq* buffer NH_4SO_4	Thermo Fisher Scientific; Waltham, USA
AMPure XP beads	Beckman Coulter; Brea, USA
ATP, UTP, GTP, CTP	Crystal Chem.; Elk Grove Village, USA
Blotting paper	Macherey-Nagel; Duren, Germany
Bradford reagent	BioRad; Hercules, USA
Cell culture dishes and plates	Sarstedt; Nümbrecht, Germany
cOmplete Mini EDTA-free	Roche; Basel, Switzerland
dNTPs 10 mM each	Biozym; Hessisch Oldendorf, Germany
Dynabeads Protein G	Thermo Fisher Scientific; Waltham, USA

FACS Flow/Clean/Rinse	Becton, Dickinson; Franklin Lakes, USA
FuGENE HD	Promega; Madison, USA
GFP Trap A	ChromoTek; Martinsried, Germany
MagStrep “type2HC” Beads	IBA; Göttingen, Germany
NEB buffer 2	New England Biolabs; Ipswich, USA
PCR 8er-stripes; PCR plates	Biozym; Hessisch Oldendorf, Germany
PEG-800 for Solexa 3' ligation	New England Biolabs; Ipswich, USA
PhosSTOP	Roche; Basel, Switzerland
Polyvinylidene fluoride (PVDF) membrane	Millipore; Billerica, USA
qPCR plates	Eppendorf; Hamburg, Germany
Random hexamers	Eurofins Genomics; Ebersberg, Germany
RestoreTM Western Blot Stripping Buffer	Thermo Fisher Scientific; Waltham, USA
RiboLock RNAse inhibitor	Thermo Fisher Scientific; Waltham, USA
SuperSignal West Dura Extended Duration	Thermo Fisher Scientific; Waltham, USA
SybrGold nucleic acid stain	Thermo Fisher Scientific; Waltham, USA
SybrSafe nucleic acid stain	Thermo Fisher Scientific; Waltham, USA
TRIzol Reagent	Thermo Fisher Scientific; Waltham, USA
UltraPure ProtoGel	National Diagnostics; Atlanta, USA
UltraPure SEQUAGEL UreaGel	National Diagnostics; Atlanta, USA

4.1.6 Bacterial cells and media

4.1.6.1 Media

SOC-medium:
0.5% (w/v) yeast extract
2% (w/v) Tryptone
10 mM NaCl
2.5 mM KCl
10 mM $MgCl_2$
10 mM $MgSO_4$
20 mM Glucose
pH 7.0

LB medium:
1% (w/v) Tryptone
0.5% (w/v) yeast extract
1% (w/v) NaCl
pH 7.2

Antibiotics (ampicillin, kanamycin) were added at a final concentration of 100 µg/mL.

LB-agar plates containing antibiotics were obtained from in-house supplies.

4.1.6.2 Bacterial cells

XL2-blue ($CaCl_2$-competent)	Plasmid amplification, *Taq* polymerase expression
BL21 (DE3 pLys S)	T7 RNA polymerase expression

4.1.7 *Drosophila melanogaster* cells and media

4.1.7.1 Media and additives

Schneider's Drosophila Medium	Bio & Sell; Feucht, Germany
Fetal Calf Serum (FCS)	Biochrom; Berlin, Germany
Pen/Strep	Thermo Fisher Scientific; Waltham, USA

4.1.7.2 Cell lines

Table 1: *Drosophila* cell lines used for screening and further experiments

Name	Description	Origin
S2B2	parental Schneider cell line clone B2	laboratory stock
Klon4	endo-siRNA reporter cell line derived from S2B2; Fluc (~ 2 copies) and Rluc (~ 9 copies) reporters	R. Böttcher
63-33	Low-copy GFP reporter cell line derived from S2B2	K. Ellmer
63N1	endo-siRNA GFP reporter cell line derived from S2B2	Laboratory stock (Hartig et al. 2009)
67-1D	siRNA reporter cell line with two perfect binding sites for miR-277 in GFP 3'-UTR	Laboratory stock (Forstemann et al. 2007)
S2RB14 5-3 Hygro	S2B2 cells stably expressing myc-Cas9, hygromycin resistant clone 5-3	Laboratory stock (Bottcher et al. 2014)

4.1.8 Plasmids

Table 2: Plasmids used as reporters and for endogenous tagging of proteins

Plasmid	Description	Reference
pRB1	constitutive myc-tagged *Renilla* luciferase expression; ubiquitin-promotor; SV40 3' untranslated region (UTR); ampicillin-resistance	(Michalik, Bottcher, and Forstemann 2012)
pRB2	FLAG-tagged Firefly luciferase expression; tubulin promotor and 3' UTR; ampicillin-resistance	(Michalik, Bottcher, and Forstemann 2012)

Plasmid	Description	Reference
pRB3	inverse *Renilla* luciferase; ubiquitin-promotor; SV40 3' UTR; ampicillin-resistance	R. Böttcher
pRB4	truncated *Renilla* luciferase; ubiquitin-promotor; ampicillin-resistance	R. Böttcher
pRB17	template plasmid to add U6-promotor to sgRNA via overlap-PCR; T7-promotor; ampicillin-resistance	(Bottcher et al. 2014)
pMH4	template plasmid for HR donor PCR; C-terminal 2x FLAG tag with Blasticidin resistance cassette; ampicillin-resistance	(Bottcher et al. 2014)
pMH3	template plasmid for HR donor PCR; C-terminal eGFP tag with Blasticidin resistance cassette; ampicillin-resistance	(Bottcher et al. 2014)
pIW1	template plasmid for HR donor PCR; C-terminal Twin-Strep tag with Blasticidin resistance cassette; ampicillin-resistance	(Bottcher et al. 2014)
pSK25	template plasmid for HR donor PCR; C-terminal 2x FLAG tag with Puromycin resistance cassette; ampicillin-resistance	(Kunzelmann et al. 2016)
pSK23	template plasmid for HR donor PCR; C-terminal eGFP tag with Puromycin resistance cassette; ampicillin-resistance	(Kunzelmann et al. 2016)
pSK24	template plasmid for HR donor PCR; C-terminal Twin-Strep tag with Puromycin resistance cassette; ampicillin-resistance	(Kunzelmann et al. 2016)
pRB34	template plasmid for HR donor PCR; N-terminal 3x FLAG tag with inducible mtnDE promotor, constitutive *copia*-Blasticidin cassette; ampicillin-resistance	(Kunzelmann et al. 2016)
pRB40	template plasmid for HR donor PCR; N-terminal eGFP tag with inducible mtnDE promotor, constitutive *copia*-Blasticidin cassette; ampicillin-resistance	(Kunzelmann et al. 2016)
pRB32	template plasmid for HR donor PCR; N-terminal Strep tag with inducible mtnDE promotor, constitutive *copia*-Blasticidin cassette; ampicillin-resistance	(Kunzelmann et al. 2016)

Plasmid	Description	Reference
pRB26	template plasmid for HR donor PCR; N-terminal eGFP tag with inducible mtnDE promotor, constitutive *copia*-Puromycin cassette; ampicillin-resistance	R. Böttcher
pRB27	template plasmid for HR donor PCR; N-terminal Strep tag with inducible mtnDE promotor, constitutive *copia*-Puromycin cassette; ampicillin-resistance	R. Böttcher

4.1.9 Enzymes

DNase I RNAse free	Thermo Fisher Scientific; Waltham, USA
EcoRI-HF	New England Biolabs; Ipswich, USA
Pfu polymerase	Laboratory stock
Phusion DNA polymerase	New England Biolabs; Ipswich, USA
Phusion Hot Start DNA polymerase	Thermo Fisher Scientific; Waltham, USA
Proteinase K	New England Biolabs; Ipswich, USA
SuperScript III Reverse Transcriptase	Thermo Fisher Scientific; Waltham, USA
T4 RNA ligase 1	New England Biolabs; Ipswich, USA
T7 endonuclease I	New England Biolabs; Ipswich, USA
T7 polymerase	Laboratory stock
Taq polymerase	Laboratory stock
Truncated T4 RNA ligase 2	Laboratory stock

4.1.10 Oligonucleotides

4.1.10.1 Primers for dsRNA constructs

DsRNAs for controls, tagging and specific knockdowns:

dsAgo2_s	taatacgactcactataggGCTGCAATACTTCCAGCACA
dsAgo2_as	taatacgactcactataggCTCGGCCTTCTGCTTAATTG
dsDcr2_s	taatacgactcactataggATTGTTGACCAAAGCGGAAC
dsDcr2_as	taatacgactcactataggATTCCCAAAACGCTCAACAC
dsGFP_s	taatacgactcactataggACCCTCGTGACCACCCTGACCTAC
dsGFP_as	taatacgactcactataggGGACCATGTGATCGCGCTTCTCGT
dsRed_s	cgtaatacgactcactataggAGGACGGCTGCTTCATCTAC
dsRed_as	cgtaatacgactcactataggTGGTGTAGTCCTCGTTG

dsRluc_s	taatacgactcactataggATGGCTTCCAAGGTGTACGACC
dsRluc_as	taatacgactcactataggCATTTTCTCGCCCTCTTCGCTC
dsThread_s	taatacgactcactataggCGATGTGATGCGCGTATATT
dsThread_as	taatacgactcactataggGCTTACGATAACTGGCAGGC
dsLoqs_s	cgtaatacgactcactataggGCAACCACAAATATCAGT
dsLoqs_as	cgtaatacgactcactataggTTGCACGGTTTTCGGGAG
dsDoa_s	taatacgactcactataggAAGCTGCCAAGCTGGAAATA
dsDoa_as	taatacgactcactataggCTTACCCAATGGACCAAACG
dsLig4_s	taatacgactcactatagggCCCAATGATCCAAAGTGTTTTTGCA
dsLig4_as	taatacgactcactatagGGAAGTAGGATGCCTTCGCGA
mus308_s	taatacgactcactataggGCTGGGACTCCACCGGAAAG
mus308_as	taatacgactcactatagggTACCGTCGCCGTCCAGTAATG

DsRNAs for validation screens:

The locus-specific primer sequences following the "taatacgactcactataggg"-T7 sequence are listed in Appendix 1: DsRNA sequences of validation screens for all dsRNA constructs for validation screening.

4.1.10.2 Primers for qPCR

tsr-1_intron_s	gcacacacacttggttgaca
tsr-1_as	TGCCACAGTCTCCACATCAA
tsr-1_transcript_s	ACTGTGTCTGATGTCTGCAAG
tsr-1_exon-exon_s	AAAATGGCTTCTGGTGTAACTGT
tsr-2_intron_s	caagtaatggcgcagaactca
tsr-2_as	GGTGGCCCGGAGTTTCTC
tsr-2_exon-exon_s	CCCGATACTGCCAAGGTCAA
CG15098-1_s	TTCTTGGCCACGATCCTCTT
CG15098-1_intron_as	atacgattggttgtgggtgc
CG15098-1_exon-exon_as	TGAAAACGATGATTAGCACACTG
CG15098-2_s	accggcaagtttctttgcaa
CG15098-2_as	GATGATCTGGGCCAGCATTG
CG15098-3_s	GAACGCCATTTACTCCTGCT
CG15098-3_transcript_as	CAGGCCAACATCCACGAAAA
CG15098-3_intron_as	ccacaatttcgcaacaacgg
CG15098-3_exon-exon_as	AGAGATACCAGGTCAGCAAGAG
α-tub84B_intron_s	gcggctctgcaattatcgat
α-tub84B_as	AGTCTCGCTGAAGAAGGTGT
α-tub84B_transcript_s	ATGTTGGTCAGGCTGGTGT
α-tub84B_exon_s	ACTCAATATGCGTGAATGTATCTCT
Act5C_intron_s	gtccaaggaaaccacgcaaa
Act5C_as	CTCATCACCCACGTACGAGT
Act5C_both_s	TGTGACGAAGAAGTTGCTGC
Act5C_exon_s	AACACACCAAATCTTACAAAATGTG

Act42A_exon_s	TCCTACATATTTCCATAAAAGATCCA
Act42A_as	ACGGCCGACAATAGAAGGAA
Act42A_intron_s	cgctgtcctgtctccttgtt
Act42A_both_s	GTGTGACGAAGAGGTTGCAG
rp49_s	ATCGGTTACGGATCGAACA
rp49_as	ACAATCTCCTTGCGCTTCTT

4.1.10.3 CRISPR and T7 endonuclease assay oligonucleotides for DSB-introduction

Table 3: Target-specific primers for U6-sgRNA template synthesis

CRISPR	Sequence	Target
202	taatacgactcactataGCCAGGGTGACGATATCAAGCgttttagagctag	Tctp
206	taatacgactcactataGGCCGTTGTTGCGAAATTGgttttagagctag	CG15098
207	taatacgactcactataGCTCGTTTTCAGTGCTGACCgttttagagctag	CG15098
208	taatacgactcactataGGTATAATCCTCTCCTGCgttttagagctag	CG18273
209	taatacgactcactataGATTCTGGAGCTACTAATgttttagagctag	CG18273
210	taatacgactcactataGCTTGCAGATCTTATGCATTTgttttagagctag	CG18273
502	taatacgactcactataGTGTGGTCATTGGATGGCTgttttagagctag	CG15098
503	taatacgactcactataGTTGGTTGCAGGTACCGTTAgttttagagctag	CG15098
602	cctattttcaatttaacgtcgTATACGCATTAAATCAACAgtttaagagctatgctg	R2D2
653	cctattttcaatttaacgtcgGTTCGCTGCAGCAAAAGAAgtttaagagctatgctg	7SK snRNA
748	cctattttcaatttaacgtcgTTCAAGCACGGTCTGGAGGgtttaagagctatgctg	Tctp
749	cctattttcaatttaacgtcgATCACCAATATTGCAATGCgtttaagagctatgctg	CG15098
750	cctattttcaatttaacgtcgTCCAGTGTAGCTTCCCGTTgtttaagagctatgctg	CG15098
751	cctattttcaatttaacgtcgTACCACAAATTTAGGGGCAgtttaagagctatgctg	CG18273
1077	cctattttcaatttaacgtcgTTGTTATATGTAAATCAAAgtttaagagctatgctg	RPII-33
1173	cctattttcaatttaacgtcgATATCTAATTTCTTTTTACgtttaagagctatgctg	Tctp

Table 4: Primers for T7 endonuclease assay

Primer	Sequence	T7 assay
212	ACGCCATCTTCCGACTCCCGACA	CRISPR 202
213	TTTCGCGGTATTCCACCAGGGCC	CRISPR 202
216	ATCTTCGCATTAGCACGAAATC	CRISPR 503

Primer	Sequence	T7 assay
217	GCCGGTTTTTTTTACAGAATAG	CRISPR 502
218	TTTACTCCTGCTGCCCTGGC	CRISPR 206,207
219	ATGAAATAAGAGCTGGCCAA	CRISPR 206,207
220	ATGTCAGCCTGATCCTTTTG	CRISPR 208
221	CAAAGTGTACTTTCCCAACC	CRISPR 208
222	TCCTAAGGGATAACGCGTCG	CRISPR 209,210
223	GACGACGCTCTTCCGCTGGC	CRISPR 209,210
224	GAACGCCATTGGCCAAGTCGA	CRISPR 751
225	AATTTTGGACAACTCAATCA	CRISPR 751
504	GTCTGGCAACGCCGCTGTACC	CRISPR 502
505	GCCAGGGCAGCAGGAGTAAA	CRISPR 503
752	CTGGCGTGGATGTTGTGCTT	CRISPR 748
753	GACAGGCATTTTTGTCAGCG	CRISPR 748
754	TTTACCTGGCGCTCAAGGTG	CRISPR 749
755	GTCACTTACAGAGCAGGCCA	CRISPR 749
756	GGTATCTCTACTACGGCATC	CRISPR 750
757	GCAGCCAATCGAAAAATCTTG	CRISPR 750
1172	CATATGGCGGCGTCTACCTACG	CRISPR 1173

4.1.10.4 Oligonucleotides for tagging and intron integration

Table 5: Oligonucleotides used for endogenous tagging of proteins

Primer	Sequence	Description
C-terminal tagging of Doa		
909	cctattttcaatttaacgtcgCTGTAATTCTCATC TGGAGgtttaagagctatgctg	CRISPR cuts at C-terminus of Doa
153	ATAGGCTGCCACCACACCATCGAGT AGGTGAGGTCAGCAACAAGCAGCCC CTTTCGTCGGGCAGCAGCAGCCGCG AACGATCGCATAGCCTCTCCAGAGga tcttccggatggctcgag	HR-primer Doa Ct_s

Primer	Sequence	Description
154	CGACTGCCTTGAGTGTATGTTTGCATGAGTGAGTACTTAATTAGCGCTGTCTGTCAACAATAACCAAATCGCTGTAATTCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCCATATG	HR-primer Doa Ct_as
170	TGGCCTATCAATTATGCTACTCTGTG	check C-term tagging of Doa_s
171	GGTTGATCTTTCGATATTCGGTTGTAAAC	check C-term tagging of Doa_as
229	Ctatgagccatacccgctggac	check C-term tagging of Doa_s2
1008	Catctatacaagtcaagttc	check C-term tagging of Doa_as2
N-terminal tagging of Doa		
872	Tcgctgaaggacaagctcat	check which isoform of Doa is expressed_s_1
873	Cactgtctcccagacgtggt	check which isoform of Doa is expressed_s_2
874	Gaatgatcggcagtatcgtg	check which isoform of Doa is expressed_as
890	cctattttcaatttaacgtcgTTCAGCGAGGGTAGTTGCAgtttaagagctatgctg	CRISPR cuts at N-terminus of Doa
888	aaaaggtgaacattcccatgcaaacactcgcacctaagcagaaacaacagtacaaaacccgaagttcctatactttctagagaataggaacttccatatg	HR-primer Doa Nt_s
889	tttcgcttgctgcacgctgccgctgggcatgagcttgtccttcagcgagggtagttgcataccgccgcttggagcagcTGGAGA	HR-primer Doa Nt_as
1010	Gacgagctgctagaggcgaa	check N-term tagging of Doa_s
1009	Acaaatcccatagacgggtagca	check N-term tagging of Doa_as
Inactivation of Doa		
721	cctattttcaatttaacgtcgCACGACAATCGCGAACACTgtttaagagctatgctg	internal CRISPR to mutate/inactivate Doa
722	cctattttcaatttaacgtcgattacCGTGCCATGCGATAgtttaagagctatgctg	internal CRISPR to mutate/inactivate Doa
734	cgaaatttcgcctacagGTGCATCTTGTTCGAACTGTATCTGGGAATCACGCTCTTCCAAggatcttccggatggctcgag	internal HR-primer to mutate/inactivate Doa
735	CCGTGTGACGTTTGGTCCATTG	check internal mutation of Doa_s
736	CTGGACTTCTCATCCCAATCTAACTTACC	check internal mutation of Doa_as

Primer	Sequence	Description
C-terminal tagging of l(1)10Bb		
763	cctattttcaatttaacgtcgCGTCCACATGGAATCTAACgtttaagagctatgctg	CRISPR cuts at C-terminus of l(1)10Bb
764	AAACTGGAAGAGGGTCGGATCGTTGAGTGTGTCCACTGCGGTTGTCGCGGCTGCTCCGGTggatcttccggatggctcgag	HR-primer l(1)10Bb Ct_s
765	TATGTATTATGAGGATTATTTGTGTCGAATAAAGATGCCAATAATGCGTCCACATGGAATGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCCATATG	HR-primer l(1)10Bb Ct_as
760	TTTCGGCACGAACTGCATATGC	check C-term tagging of l(1)10Bb_s
761	GTATTTACATACTATGAAGAAATCAGGCG	check C-term tagging of l(1)10Bb_as
N-terminal tagging of l(1)10Bb		
762	cctattttcaatttaacgtcgTGAACATTTAAGATGCCCAgtttaagagctatgctg	CRISPR cuts at N-terminus for tagging of l(1)10Bb
766	AATTAAGCATTTTTAAGAGAAATTAAACAGTTTTACAGTTAAATAAATTGAACATTTAAGgaagttcctatactttctagagaataggaacttccatatg	HR-primer l(1)10Bb Nt_s
767	AGTTGGTTCGATCAGCTCCCAACCGTCTGGCGGTGGTTTGCGACTGCGGCGAACCTTGGGaccgccgcttggagcagcTGGAGA	HR-primer l(1)10Bb Nt_as_wt
768	TGGTTCGATCAGCTCCCAACCGTCTGGCGGTGGTTTGCGAgcGCGGCGAACCTTGGGaccgccgcttggagcagcTGGAGA	HR-primer l(1)10Bb Nt_as_mutation Ala
769	TGGTTCGATCAGCTCCCAACCGTCTGGCGGTGGTTTGCGAtcGCGGCGAACCTTGGGaccgccgcttggagcagcTGGAGA	HR-primer l(1)10Bb Nt_as_mutation Asp
758	CCCATGTGGAAAGTGTCATCGC	check N-term tagging of l(1)10Bb_s
759	CCTCGTGCGGTTCCGTTTCGG	check N-term tagging of l(1)10Bb_as
General primers for tagging		
12	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	Scaffold primer for generation of sgRNA
252	GCTCACCTGTGATTGCTCCTAC	Sense primer for *Drosophila* U6 promotor
254	gcttattctcAAAAAAGCACCGACTCGGTGCCACT	Antisense pimer for sgRNA template PCR

Primer	Sequence	Description
336	GTTTaAGAGCTAtgctgGAAAcagcaTAGCAAGTTtAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	Optimized scaffold primer for generation of sgRNA
451	aataggaacttcTAATCCAAAATGgga	Universal primer to check N-term tagging_s
646	GGTATTCTCTTACAATATGTTTTATGGCATAAAAGG	Universal *copia*_as primer to check C-term tagging (Blast+Puro)
231	GTAGGTTGAATAGTATATTCCAACAGCATATG	Universal *copia*_as primer to check C-term tagging (Blast only)
340	AACATATGCTGTTGGAATATACTATTCAACC	*Copia*_s primer to check knockout
351	ACAATCAACAGCATCCCCATCTC	Sense primer for Split-Blast system
352	TTCTCATTTCCGATCGCGACGATAC	Antisense primer for Split-Blast system
781	GGACGTTGGCTGCCGC	Sense primer for Split-Puro system
782	CCCCTGCTTCCACGCT	Antisense primer for Split-Puro system
Intron integration		
558	TAGTTGCATTTTCGCCAGCGATTTAGACGCCATCTTCCGACTCCCGACAAGCCAATCACCTAGgtagtgtttccctcaatt	homology Primer Tctp with GAPDH-intron-overlap sense
559	ATCTTGTAGGTGTCGGCAAACATCTCGTCGCCGGTGATGATATCCTTaTAGATCTTCATctggaaagtaatgtatgtat	homology Primer Tctp with GAPDH-intron-overlap antisense
560	taatacgactcactatagATCACCATGAAGATCTACAgttttagagctag	CRISPR for Tctp-intron-integration, cuts near start codon
561	TAGTTGCATTTTCGCCAGCGATTTAGACGCCATCTTCCGACTCCCGACAAGCCAATCACCgtagtgtttccctcaattccg	homology Primer Tctp with GAPDH-intron-overlap without splice site_sense
585	GGTCCGGTGCAGTTGTAGTT	to check GAPDH-Intron integration in Tctp -> before intron sense
586	TCCTCCTCCAGACCGTGCTT	to check GAPDH-Intron integration in Tctp -> coding sequence antisense
587	GCCAATCACCATGAAGATCTA	to check GAPDH-Intron integration in Tctp -> wt only sense
588	cattactttccagATGAAGATCT	to check GAPDH-Intron integration in Tctp -> with intron only sense
634	ACCTCGTAGATCACATCATCC	alternative antisense primer to 586 to check Tctp-intron (shorter product)

4.1.10.5 SOLEXA oligonukleotides

Solexa 3' linker (IDT 25 µM)	AMP-5'p=5'pCTGTAGGCACCATCAATdideoxyC-3'
Solexa 5' RNA linker (MWG Eurofins 50 µM)	5'-rArCrArCrUrCrUrUrUrCrCrCrUrArCrArCrGrArCrGrCrUrCrUrUrCrCrGrArUrCrU-3'
RT primer index	5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCGATTGATGGTGCCTACAG-3'
Illumina 5' ext	5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'
3'PCR primer index1	5'-CAAGCAGAAGACGGCATACGAGATatcacgGTGACTGGAGTTCAGACGTG-3'
3'PCR primer index2	5'-CAAGCAGAAGACGGCATACGAGATcgatgtGTGACTGGAGTTCAGACGTG-3'
3'PCR primer index3	5'-CAAGCAGAAGACGGCATACGAGATttaggcGTGACTGGAGTTCAGACGTG-3'
3'PCR primer index4	5'-CAAGCAGAAGACGGCATACGAGATtgaccaGTGACTGGAGTTCAGACGTG-3'
3'PCR primer index5	5'-CAAGCAGAAGACGGCATACGAGATcagaatGTGACTGGAGTTCAGACGTG-3'
3'PCR primer index6	5'-CAAGCAGAAGACGGCATACGAGATgctgtaGTGACTGGAGTTCAGACGTG-3'
3'PCR primer index7	5'-CAAGCAGAAGACGGCATACGAGATtcgcacGTGACTGGAGTTCAGACGTG-3'
3'PCR primer index8	5'-CAAGCAGAAGACGGCATACGAGATcagtggGTGACTGGAGTTCAGACGTG-3'
3'PCR primer index9	5'-CAAGCAGAAGACGGCATACGAGATggtatcGTGACTGGAGTTCAGACGTG-3'

4.1.11 Antibodies

Table 6: Antibodies used for immunoprecipitation and Western blotting

antibody	2nd antibody	Dilution	source
α-GFP	Mouse	1:10000	Santa Cruz, F0908
α-FLAG M2	Mouse	1:10000	Sigma, F1804
α-FLAG-HRP M2	not required	1:5000	Sigma
α-Strep-tag II-HRP	not required	1:10000	IBA
α-anti-Rm62 (1E7)	Rat	1:5000	Imhof Lab
α-Ago2 9D6	Mouse	1:5	Siomi Lab
α-Dcr-2 8-59	Mouse	1:2000	Siomi Lab
α-Loqs 1E4	Mouse	1:1000	Siomi Lab
α-beta-tubulin	Mouse	1:10000	DSHB

4.1.12 Buffers

ATP-free T4 RNA ligase buffer:
100 mM $MgCl_2$
100 mM DTT
600 µg/mL BSA
500 mM Tris-HCl, pH 7.5

Buffer A (genomic DNA extraction):
100 mM Tris-HCl, pH 7.5
100 mM EDTA
100 mM NaCl
0.5 % SDS

Colloidal Coomassie staining solution:
50 g/L aluminum sulfate
2 % (v/v) H_3PO_4 (conc.)
10 % (v/v) ethanol
0.5 % (v/v) Coomassie G250 stock

6x DNA loading buffer:
0.25 % (w/v) bromophenol blue
0.25 % (w/v) xylene cyanol
30 % (w/v) glycerol

2x Formamide loading dye:
80 % (w/v) formamide
10 mM EDTA, pH 8

	1 mg/mL xylene cyanol 1 mg/mL bromophenol blue
4x Laemmli SDS loading buffer:	200 mM Tris-HCl, pH 6.8 8 % (w/v) SDS 40 % (v/v) glycerol 0.4 % (w/v) bromophenol blue 200 mM DTT (freshly added)
5x SDS Running buffer:	125 mM Tris/HCl, pH 7.5 1.25 M glycine 5 % SDS
10x TBE:	0.9 M Tris 0.9 M boric acid 0.5 M EDTA (pH 8)
10x TBS:	50 mM Tris, pH 7.4 150 mM NaCl
10x Western solution:	250 mM Tris/HCl, pH 7.5 1.92 M glycine
1x Western transfer buffer:	10 % 10x Western solution 10 % ethanol
Coomassie G250 stock solution:	0.5 g/L Coomassie G250 in 100% methanol
Coomassie staining solution:	45 % (v/v) methanol 10 % acetic acid 0.25 % (w/v) Coomassie Brilliant Blue R
Coomassie destain:	45 % (v/v) methanol 10 % acetic acid
10x PBS:	137 mM NaCl 2.7 mM KCl 10 mM Na_2HPO_4 2 mM KH_2PO_4, pH 7.4
TBS-T:	TBS supplemented with 0.05 % Tween-20
10x *Taq* buffer:	500 mM KCl 100 mM Tris-HCl, pH 8.3 optional: 1 % Triton X-100 For cell culture PCRs, the pH was changed to 9.3 (Bu, Huang, and Zhou 2008)

10x T7 transcription buffer:	400 mM Tris pH 7.9 10 mM Spermidine 260 mM $MgCl_2$ 0.1 % Triton X-100
LiCl/KAc Solution:	1.43 mM KAc 4.3 mM LiCl
DMP/borate:	0.1M sodium tetraborate pH 9.0 20 mM DMP
Lysis Buffer (LB):	150 mM KAc pH 7.4 30 mM HEPES pH 7.4 100 mM glycine pH 7.4 5 mM MgAc 1 mM DTT 15 % Glycerol 1 % Tergitol 1 tablet/10 mL buffer cOmplete Mini (-EDTA) 1 tablet/10 mL buffer PhosSTOP
IP-Washing Buffer 1:	150 mM KAc pH 7.4 30 mM HEPES pH 7.4 5 mM MgAc 1 mM DTT 15 % Glycerol 0.1 % Tergitol 1 tablet/10 mL buffer cOmplete Mini (-EDTA) 1 tablet/10 mL buffer PhosSTOP
IP-Washing Buffer 2:	150 mM KAc pH 7.4 30 mM HEPES pH 7.4 5 mM MgAc
1x TAE:	40 mM Tris 20 mM acetic acid 2 mM EDTA pH 8.0

4.1.13 Software

Microsoft Excel 2010/2013	Microsoft Corporation; Redmond, USA
R (3.3.2), Bioconductor (2.11)	Fred Hutchinson Cancer Research Center; Seattle, USA
cellHTS2 (2.22.0) package	Boutros et al. (Boutros, Bras, and Huber 2006)
Adobe Illustrator CS6	Adobe Systems Software; Dublin, Ireland
Inkscape (1.2.0.0)	Inkscape Project/Software Freedom Conservancy; NY, USA

ApE plasmid Editor	M. Wayne Davis, University of Utah; Salt Lake City, USA
BD Cell Quest	Becton, Dickinson; Franklin Lakes, USA
BOWTIE	Langmead et al. (Langmead et al. 2009)
PERL scripts	Hartig et al. (Hartig et al. 2009)
GALAXY platform	Giardine et al. (Giardine et al. 2005)
Flowing Software 2.5.1	Perttu Terho; University of Turku; Turku, Finland
qPCRsoft 3.2	Analytik Jena; Jena, Germany
GOrilla	Eden et al. (Eden et al. 2009)

4.2 Methods

4.2.1 *Drosophila* cell culture

4.2.1.1 Cell maintenance and storage

Drosophila S2B2 cells and derivates were cultured in Schneider's medium supplemented with 10% fetal calf serum (FCS) and Penicillin/Streptomycin (Pen/Strep). They were grown at 25 °C in appropriate cell culture flasks and split when a confluency of approximately 90% was reached for up to 20 passages.

For long-term storage, 450 µL of dense cells were supplemented with 100µL dimethyl sulfoxide (DMSO) and 450 µL of fresh medium in cryovials, slowly cooled to -80 °C in an isopropanol freezing container and then transferred to liquid nitrogen.

For culturing N-terminally tagged cells, the growth medium was supplemented with $CuSO_2$ at a final concentration of 40 µM (maintenance) to 400 µM (protein overexpression).

4.2.1.2 RNA-Interference (RNAi)

4.2.1.2.1 Preparation of dsRNA

In order to generate locus-specific dsRNAs, corresponding primers containing T7-promotor sites adjacent to the gene-specific sequences were designed and used in the following PCR reaction on cDNA from S2 cells.

PCR mix:

10x *Taq* buffer NH_4SO_4	5 µL
dNTPs (10µM each)	1 µL
$MgCl_2$ (25mM)	5 µL
cDNA (0.5mg/mL)	1 µL
Taq polymerase	1 µL
Primer Mix (10µM)	1 µL
H_2O (nuclease-free)	ad 50 µL

PCR program:

95 °C	3 min	
95 °C	1 min	┐
55 °C	30 s	├ 35 cycles
72 °C	1 min	┘
72 °C	5 min	

To improve PCR outcome, buffer conditions (KCl instead of NH_4SO_4), $MgCl_2$ concentration, polymerase (*Phusion* polymerase instead of *Taq* polymerase) and template (gDNA 90 ng/µL instead of cDNA) were adjusted for certain constructs.

The PCR products were then used as templates for *in vitro* transcription (IVT) with T7 polymerase at 37 °C overnight.

IVT-mix:

10x T7 transcription buffer	10 µL
DTT (1M)	0.5 µL
ATP	4.25 mM end conc.
CTP	4.25 mM end conc.
UTP	4.25 mM end conc.
GTP	6.5 mM end conc.
T7 polymerase	2 µL
H_2O (nuclease-free)	ad 75 µL
Template PCR	25 µL

After *in vitro* transcription, 1 µL of DNAse I was added per 100 µL IVT reaction and incubated at 37 °C for another 30 min. The magnesium pyrophosphate precipitate was then removed by centrifugation. To enable proper strand annealing, the samples were heated to 95 °C for 5 minutes and slowly cooled down to room temperature. The concentration of dsRNA was determined via agarose gel electrophoresis.

For IVTs conducted in 96-well plates, no DNAse I incubation was performed and the success of the *in vitro* transcription was confirmed via agarose gel electrophoresis after annealing. Subsequently, the IVT reactions were diluted (1:3 or 1:5) with H_2O to avoid toxicity of poorly removed magnesium pyrophosphate precipitate.

4.2.1.2.2 Specific gene knockdown in *Drosophila* cells

To induce gene-specific knockdowns via RNAi, *Drosophila* cells were seeded at $0.5x10^6$ cells/mL (63-33, 67-1D and 63N1 cells) or $1x10^6$ cells/mL (S2 and Klon4 cells). DsRNA was added to a final concentration of 2.5 ng/µL or 5 µL of diluted IVT per 96-well. After 4 days of soaking (on day 5), the cells were harvested and analyzed.

4.2.1.3 Transfection of plasmids

4.2.1.3.1 Preparation of plasmids

In order to be used for validation screens, the plasmids pRB1, pRB2 and pRB4 (see Table 2) were re-transformed in $CaCl_2$-competent XL2-blue *E.coli* via heat-shock and subsequent growth on LB plates supplemented with appropriate antibiotics. Midi/Maxi-preps were prepared from 100 mL/400 mL bacterial culture using the respective Qiagen kit and verified via test restrictions. The plasmid pRB4 was

then linearized using the EcoRI-HF restriction enzyme in digestion reactions of 100 µg each at 37 °C overnight followed by heat inactivation. After confirming full linearization, the digested plasmids were purified using a Qiagen gel extraction kit according to manufacturer's instructions without preceding agarose gel electrophoresis.

4.2.1.3.2 Plasmid transfection

For transfection, plasmid DNA (for exact amounts see chapters 4.2.1.4 and 4.2.1.5) was pre-diluted with serum-free medium and complemented with the transfection reagent FuGENE HD. The mixture was then incubated for 30 min and added to the cell suspension.

4.2.1.4 Validation screens

For validation experiments, S2 cells/reporter cells were seeded in 96-well plates (65 µL of S2 and Klon4 cells, 100 µL of GFP reporter cells) and treated with dsRNA as described in 4.2.1.2.2.

The first round of validation was performed for positive and negative hits from the genome-wide screen with dsRNAs identical in sequence to those present in the HD2 library. For the second round of validation, two additional, independent dsRNA constructs (DesignA and DesignB, Appendix 1: DsRNA sequences of validation screens) were designed for the selected hits and used for knockdown. The same dsRNAs as in the genome-wide screen were used as controls: positive controls (known siRNA pathway factors Ago2, Dcr-2, Ago1), negative controls (GFP and DsRed), a dsRNA construct against the *Renilla* luciferase reporter as positive control for potential repressors and knockdown efficiency and a dsRNA against the apoptosis inhibitor Thread to identify and exclude dead cells.

All validation experiments were performed in two independent biological replicates and with S2, Klon4, 63-33, 63N1 and 63-33 cells.

In order to induce siRNA formation in S2 cells, linearized pRB4 was used as trigger. Therefore, a 10 µL transfection mix containing 10 ng pRB1, 25 ng pRB2, 40 ng linearized pRB4 and 0.3 µL FuGENE in serum free medium was added 48 h after knockdown.

96 hours after knockdown, the cells were analyzed by dual-luciferase readout (transfected S2 cells and Klon4 cells) or flow cytometry (GFP-reporter cells).

4.2.1.5 CRISPR-Cas9-mediated genome editing

4.2.1.5.1 Introduction of double-strand breaks

In order to induce DNA double-strand breaks in S2 cells that stably express the Cas9 nuclease, U6-sgRNA templates for intracellular sgRNA expression were generated via the following overlap extension PCR reaction.

PCR mix:		PCR program:		
10x*Taq* buffer KCl	5 µL	95 °C	5 min	
dNTPs (10 µM each)	1 µL	95 °C	30 s	┐
$MgSO_4$ (25 mM)	8 µL	68 °C	30 s	├ 5 cycles
primer U6 promotor sense (10 µM)	1 µL	72 °C	30 s	┘
primer sgRNA antisense (10 µM)	1 µL	95 °C	30 s	┐
pRB17 (10 ng/µL)	1 µL	62 °C	30 s	├ 5 cycles
specific CRISPR oligo (1 µM)	1 µL	72 °C	30 s	┘
oligo sgRNA scaffold (1 µM)	1 µL	95 °C	30 s	┐
Taq/Pfu polymerases (1:1)	1 µL	55 °C	30 s	├ 30 cycles
H_2O	ad 50 µL	72 °C	30 s	┘
		72 °C	5 min	

The thus obtained CRISPR PCR product was then purified and concentrated using a Qiagen PCR purification kit according to manufacturer's instructions, and quantified via agarose gel electrophoresis.
For transfection, the Cas9-expressing S2RB14 cells were plated at a density of $1x10^6$ cells/mL in 6-well plates (3 mL). 1500 ng of CRISPR DNA and 24 µL FuGENE were diluted to a total transfection volume of 600 µL in serum-free Schneider's medium, incubated for 30 min and added to the cells. After 4 days at 25 °C, the cells were harvested for RNA and gDNA isolation and analysis.

4.2.1.5.2 Endogenous protein tagging

For endogenous protein tagging, suitable U6-sgRNA templates to induce a Cas9-mediated break at the desired position and corresponding homology donor PCR products that carry the desired tag flanked by the target sequence were simultaneously transfected into S2RB14 cells.

The U6-sgRNA templates were synthesized as described in chapter 4.2.1.5.1 and the HR donor PCR was carried out according to the following scheme:

PCR mix:		PCR program:		
template plasmid (100 pg/µL)	6 µL	94 °C	2 min	
homology primer sense (10 µM)	2 µL	94 °C	30 s	┐
homology primer antisense (10 µM)	2 µL	50 °C	30 s	├ 30 cycles
10x *Taq* buffer without detergence	10 µL	72 °C	2 min	┘
$MgCl_2$ (25 mM)	6 µL	72 °C	5 min	
dNTPs (10 mM each)	2 µL			
H_2O	70 µL			
Taq/Pfu polymerase mix (1:1)	2 µL			

The template plasmids used for C-terminal and N-terminal tagging are listed in Table 2, the homology primers are listed in Table 5. Homology primers which otherwise show perfect homology to the corresponding U6-sgRNA region were designed with a silent point mutation close to the impending cutting position to prevent re-cutting of integrated HR-donors. Furthermore, the introduction of non-silent mutations in the antisense homology primer during N-terminal tagging of l(1)10Bb was used to deliberately change a specific amino acid of the protein.

The obtained PCR products were quantified via agarose gel electrophoresis. When detergence-free *Taq* buffer was used, U6-sgRNA templates and HR donors were not purified prior to transfection.

S2RB14 cells were treated with *mus308* and *lig4* dsRNA (1 µg/mL end conc.) 72 hours prior to transfection. For transfection, the cells were seeded at a density of $1x10^6$ cells/mL in Schneider's medium supplemented with 10 % FCS and a transfection mix of 300 ng U6-sgRNA, 200 ng HR donor and 4 µL FuGENE in 100 µL serum-free medium was added to 500 µL cell suspension.

Four days after transfection, the cells were split 1:5 into Schneider's medium with 10% FCS containing 10 µg/mL Blasticidin or 0.5 µg/mL Puromycin.

For N-terminal tagging, the medium was supplemented with $CuSO_4$ (end concentration 40 µM) after transfection and during selection.

Double-tagging was performed sequentially by using a cell line after clonal selection to repeat the tagging procedure for a second protein with a different selection marker. The selection was then carried out in medium containing both Blasticidin and Puromycin.

The successful integration of the tag was verified by Western Blot (see 4.2.4.4) and integration PCR (see 4.2.2.4) and, where beneficial, single cell clones were selected as described in chapter 4.2.1.6.

4.2.1.5.3 Intron integration

The tagging procedure described in the previous chapter was analogously used to integrate the intron of *GAPDH2* after the start codon of the intronless gene *Tctp* with either a correct or an incorrect splicing junction. To achieve this, primers that amplify the intron sequence flanked by extensions homologous to the *Tctp* target region were used in PCR. In one of the constructs, an additional TAG-codon was inserted between the *Tctp* and the intron sequence as potential 5' splice site. Instead of a template plasmid, 2 µL of gDNA were used as PCR template and the annealing temperature was adjusted to 42 °C. The obtained products were then checked for point mutations by inserting them into pJET1.2 with a CloneJET PCR Cloning Kit and performing colony PCR according to manufacturer's instructions. For each construct, the colony PCR product of two clones was also analyzed by sequencing.

In analogy to chapter 4.2.1.5.2, S2RB14 cells were pretreated with *lig4* dsRNA for 4 days and then transfected with CRISPR #560 U6-sgRNA and homology PCR products from primers #558/#559 or #558/#561. After 4 days, the transfected cells were split 1:5 into Schneider's medium with 10% FCS (no selection medium). This population was then used for clonal selection.

4.2.1.6 Clonal selection of cell lines

Clonal selection was carried out in Schneider's medium with 10 % FCS and 20 % of conditioned medium from S2 cells. For Blasticidin/Puromycin-resistant populations, the corresponding selection antibiotic was

added to the medium. Furthermore, for cell lines with copper-inducible N-terminally tagged proteins, the medium was supplemented with $CuSO_4$ (40 µM end concentration).

For single cell dilutions, 150 µL of suitable cell culture medium containing 1000 cells (alternatively 3000 or 5000 cells) each were added to 24 wells of a 96-well plate. 30 µL from each well were then transferred to 24 additional wells containing 120 µL medium thus creating a 1:5 dilution. This dilution was then used to subsequently establish 1:25 and 1:125 dilutions of the original cell suspension. The plates were sealed, incubated for at least two weeks at 25 °C and regularly checked for developing colonies. Clearly visible colonies were marked at the outside of the plate and transferred to a new 96-well plate containing 100 µL of suitable medium with a P20 pipette in about 10 µL volume. Densely grown single cell clones were then analyzed and scaled up for further experiments.

4.2.2 DNA analysis

4.2.2.1 Agarose gel electrophoresis

Depending on the expected individual length and length distribution of the analyzed nucleotides, 1-2 % agarose gels were prepared with 1x TAE buffer and stained with 1x SybrSafe during preparation. Samples were diluted with 6x DNA loading buffer and loaded on the gel. The gels were run at 55 V for 20-45 min and documented with an INTAS UV Imaging System.

4.2.2.2 Isolation of genomic DNA (gDNA)

For gDNA isolation, different approaches were used dependent on the later application of the DNA.

Genomic DNA used as PCR template for dsRNA production or as control in other PCRs was isolated in a workflow based on the Berkeley Drosophila Genome Project fly protocol. Therefore, 500 µL of S2 cell culture were harvested by centrifugation washed with 200 µL PBS, resuspended in 100 µL Buffer A, and incubated at 65 °C for 10 minutes. After addition of 400 µL LiCl/KAc solution, the lysate was incubated on ice for 10 minutes and centrifuged at 13000 rpm for 15 minutes. The supernatant was then precipitated with 500 µL isopropanol. The pellet was washed with 70 % EtOH and re-dissolved in 40 µL H_2O with 5 % buffer P1 (Qiagen Mini Kit to introduce RNAse) through incubation at 42 °C for 10 min. The thus obtained gDNA was then quantified via agarose gel electrophoresis.

For the isolation of gDNA to verify successful HR-based tagging upon CRISPR-Cas9 restriction, a small-scale isolation protocol using spin columns and buffers from a QIAquick Gel Extraction Kit was applied. 100 µL densely grown cells were lysed with 300 mL buffer QG. The DNA was precipitated with 100 µL isopropanol and the lysate was transferred to a spin column. The flow-through was discarded and the column washed with 700 µL buffer PE. After an additional 1 min centrifugation step to completely dry the matrix, the gDNA was eluted in 50 µL EB.

For more sensitive applications such as the T7 endonuclease assay, gDNA was isolated using the Reliaprep gDNA Tissue Miniprep System. 500 µL cells were harvested and re-suspended in 160 µL PBS. Following manufacturer's instructions for gDNA isolation from animal tissue, 20 µL Proteinase K Solution and 200 µL Cell Lysis Buffer were added to the cell suspension and incubated at 56 °C for 60 minutes. Afterwards, 20 µL of RNase A Solution were added and the mix incubated at 56 °C for another 10 minutes. After addition of 250 µL Binding Buffer and centrifugation, the supernatant was transferred to the binding column and washed three times with 500 µL Column Wash Solution. The genomic DNA was then eluted in 50 µL nuclease-free water.

4.2.2.3 T7 endonuclease assay

In order to evaluate the cutting efficiency of the CRISPRs used for the introduction of double-strand breaks, a roughly 500bp region around the target position was amplified via PCR on isolated gDNA.

PCR mix:

5x *Phusion* HF buffer (suppl. with enzyme)	4 µL
dNTPs (10 µM each)	0.4 µL
primer forward (10 µM)	1 µL
primer reverse (10 µM)	1 µL
gDNA (~150 ng)	2 µL
Phusion DNA polymerase	0.2 µL
H_2O	11.4 µL

PCR program:

98 °C	5 min	
98 °C	10 s	⎫
T_{anneal} °C	30 s	⎬ 35 cycles
72 °C	30 s	⎭
72 °C	10 min	

The T7 endonuclease assay primers and corresponding annealing temperatures for the different cutting positions targeted by the indicated CRISPRs are shown in Table 7.

Table 7: Primers and annealing temperatures for T7 endonuclease assay

CRISPR	Primers	T_{anneal} [°C]
502	504, 217	51
503	216, 505	59
749	754, 755	59
206	218, 219	63
207	218, 219	61
750	756, 757	55
202	212, 213	52
748	752, 753	57
1173	752, 1172	61
208	220, 221	52
209	222, 223	65
210	222, 223	65
751	224, 225	61

The thus obtained PCR products were then denatured and re-annealed in the following steps:

95 °C	5 min
95 °C to 85 °C	-2 K/s
85 °C to 25 °C	-0.1 K/s
4 °C	∞

Next, 10 µL of the PCR reaction were added to 10 µL T7 endonuclease I-mastermix (2 µL buffer NEB2, 0.5 µL T7 endonuclease I, 7.5 µL H_2O) on ice and then incubated at 37 °C for 16 min.

Immediately after digestion, the digest (10 µL + 6x DNA loading dye) and the remaining PCR reaction (5 µL + 6x DNA loading dye) were analyzed and compared on a 1.5 % agarose gel.

4.2.2.4 Evaluation of tagging success

To evaluate the tagging success or identify correctly tagged clones, different PCRs were attempted. Whenever possible, especially for screening larger numbers of clones, the PCR was performed directly with cells. If this was not possible due to inhibitory effects of the cell medium or background, gDNA was isolated as described above.

In order to check the correct integration of a C-terminal tag, a control PCR with a common antisense primer binding within the *copia* promotor of the C-terminal tagging construct (#321 for Blasticidin only or #646 for Blast or Puro) and a gene-specific sense primer was performed. For N-terminal tags, a sense primer binding between the resistance marker and the N-terminal tag was combined with a gene-specific antisense primer.

PCR mix for C-terminal tagging:

10x PCR buffer *(high pH for cells, normal pH for gDNA)*	2.5 µL
primer gene specific sense (10 µM)	1µL
primer *copia* as (10 µM)	1 µL
$MgCl_2$ (25 mM)	3 µL
dNTPs (10 mM each)	0.5 µL
H2O	16 µL
Taq polymerase	0.5 µL
gDNA or re-suspended cells	2 µL

PCR mix for N-terminal tagging:

10x PCR buffer *(high pH for cells, normal pH for gDNA)*	2.5 µL
primer gene specific antisense (10 µM)	1µL
primer us_common sense (10 µM)	1 µL
$MgCl_2$ (25 mM)	3 µL
dNTPs (10 mM each)	0.5 µL
H2O	16 µL
Taq polymerase	0.5 µL
gDNA or re-suspended cells	2 µL

PCR program:

94 °C	2 min	
94 °C	30 s	┐
55 °C*	30 s	├ 35 cycles
72 °C	2 min	┘
72 °C	5 min	

*annealing temperatures were optimized if necessary

Desired mutations that were inserted via specifically altered homology primers were verified by sequencing of the obtained PCR product.

To further distinguish between clonal cells lines that carry only tagged alleles and cell lines or populations that carry both wild-type (wt) and modified alleles, the PCRs were also carried out with gene-specific primer pairs which bind upstream and downstream of the CRISPR target site and can thus lead to shorter products for unmodified alleles and longer products where the tag was successfully integrated. The absence of the shorter wt band was considered a clear indicator that all alleles were tagged.

4.2.2.5 Intron integration analysis via PCR

The successful integration of the *GAPDH2* intron into *Tctp* was verified in analogy to the control PCR to verify C-terminal tagging success. Positive clones where identified using primers #588/#586 directly on cultured cells. To counter-screen for clones carrying both, wildtype and modified alleles, the positive clones were then analyzed with the wt-specific primers #587/#586 as well as a primer pair positioned upstream and downstream of the integration site (#585/#634).

Identical primers were further used on cDNA instead of gDNA/cells to test whether the integrated intron can be successfully spliced out or whether it is still present in the resulting mRNA of the selected cell clones.

4.2.3 RNA analysis

4.2.3.1 RNA isolation

For RNA extraction, 3 mL of densely grown cells were harvested and treated with 1 mL of TRIzol reagent according to manufacturer's instructions. The isolated RNA was re-suspended in 30 µL RNAse-free EB buffer (10 mM Tris-Cl, pH 8.5, QIAGEN) and quantified via photometry.

4.2.3.2 qRT-PCR

After isolation, RNA intended for qRT-PCR was treated with DNAseI.

DNAseI-mix:

RNA	5 µg
DNAse buffer	5 µL
Ribolock	1 µL
DNAse	1 µL
H_2O	ad 50 µL

The digestion mix was incubated at 37 °C for 30 min. After the subsequent addition of 1 µL Proteinase K, the reaction was incubated at 65 °C for 15 min and then purified using an RNA clean and concentrator kit according to manufacturer's instructions. The purified RNA was eluted in 20 µL nuclease-free water and the resulting concentration was determined photometrically.

For reverse transcription, 100 ng of purified RNA were diluted in a total volume of 18.8 µL H_2O and supplemented with 2 µL dNTP mix (10 mM each) and 3.2 µL random hexamers (100 pM). The mix was incubated at 65 °C for 5 min and then chilled on ice.

After addition of 8 µL 5x First strand buffer (supplied with enzyme), 4 µL DTT (0.1 M), and 2 µL nuclease inhibitor (Ribolock), the mix was split into two tubes and supplemented with 1 µL of Superscript III reverse transcriptase (+RT sample) or nuclease-free water (-RT sample) respectively.

The reaction was incubated at 50 °C for 50 min, transferred to 70 °C for 15 min and then diluted with nuclease-free water to obtain cDNA-concentrations suitable for qPCR.

Quantitative PCR was performed on a Biometra TOptical Thermocycler using the DyNAmo Flash SYBR Green qPCR Kit:

qPCR mix:

Template cDNA	1 µL
SYBR Green mix	5 µL
sense primer (5 µM)	0.5 µL
antisense primer (5 µM)	0.5 µL
bromphenolblue	0.1 µL
H_2O	2.9 µL

qPCR program:

95 °C	3 min	
95 °C	1 min	40 cycles
61 °C	30 s	
72 °C	30 s	

The primer sequences are listed in 4.1.10.2 and data analysis was carried out according to the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001).

4.2.3.3 SOLEXA sample preparation

The deep-sequencing libraries were created in the following multi-step procedure based on a published protocol for the generation of small RNA libraries (Elmer et al. 2014):

Small RNA isolation

20 µg of total RNA from TRIzol extraction were mixed with formamide loading dye and separated on a pre-run 20 % polyacrylamide/urea gel (UltraPure SEQUAGEL UreaGel, National Diagnostics; 8 mL concentrate, 1 mL diluent, 1 mL buffer) at 250 V for 75 min.

The gel was then stained with SyBr Gold and the gel section containing the 15 to 28nt-length RNA was cut out using a small RNA ladder at 17-25nt and the 2S rRNA (visible at 30nt) for orientation.

A ZymoResearch PAGE-Recovery Kit was used according to manufacturer's instructions to recover the size-selected small RNAs in 6 µL of nuclease-free water.

3'linker ligation

The following ligation mix was incubated at room temperature for 1 hour and then heat-inactivated for 5 min at 95 °C.

size-selected RNA	6 µL
ATP-free T4 RNA ligase buffer	1 µL
Solexa 3' linker (25 µM)	1 µL
PEG-8000 (50 %)	1 µL
truncated T4 RNA ligase 2	1 µL

To remove the unreacted linker 1, formamide loading dye was added to the ligation reaction and the mix was purified on a 15 % acrylamide/urea gel in analogy to the first gel purification (250 V, 75 min).

In this second gel electrophoresis, a 50bp DNA ladder and the small RNA ladder were utilized to select the 36-41nt fraction of ligated small RNAs.

The RNAs were again extracted with a ZymoResearch PAGE-Recovery Kit and eluted in 5 µL nuclease-free water.

5'linker ligation

ligated and purified small RNA	5 µL
ATP-free T4 RNA ligase buffer	1 µL
DMSO	1 µL
ATP	1 µL
Solexa 5' RNA linker (50 µM)	1 µL
T4 RNA ligase 1 (full-length)	1 µL

The reaction mix was incubated at room temperature for 1 hour, followed by heat-inactivation of the T4 ligase (5 min, 95 °C).

Reverse transcription

For denaturation, 2 µL of oligo RT primer index (5 µM) were added to 9 µL of 5' adapter ligation mix, incubated at 95 °C for 2 min and then cooled on ice for 2 min.

Next, the following components were added and the mix was subsequently split into two tubes (9 µL each) and incubated at 50 °C for 2 min.

5x First strand buffer (supplied with enzyme)	4 µL
100 mM DTT (supplied with enzyme)	2 µL
Ribolock RNase inhibitor	1 µL
dNTP mix (10 mM each)	1 µL

After incubation, 1 µL of Superscript III RT enzyme (+RT sample) and 1 µL of nuclease-free water (-RT sample) were added to the respective tubes and reverse transcription was carried out at 50 °C for 30 min.

Afterwards, the enzyme was heat-inactivated at 95 °C for 5 min.

PCR amplification and purification of DNA library

In order to amplify the obtained cDNA, the following PCR reaction was assembled:

first strand cDNA (+RT) or negative control (-RT)	5 µL
5x *Phusion* HF PCR buffer (supplied with enzyme; including Mg^{2+}, final conc. 2.5 mM)	20 µL
dNTP mix (10 mM each)	2 µL
oligo Illumina 5' ext. (10 µM)	1 µL
oligo 3' PCR primer index (10 µM)	1 µL
nuclease-free water	70 µL
Hot Start *Phusion* polymerase	1 µL

The index primers were chosen according to the anticipated combination of libraries during deep-sequencing. Every run contained libraries with at least one A or C and one G or T at each position of the barcode.

Amplification was performed with the following thermocycler protocol:

98 °C	1 min	
98 °C	15 s	24 cycles
58 °C	30 s	
72 °C	30 s	
72 °C	2 min	

The PCR products were then analyzed on a 2 % agarose gel, the desired band at 160 nt was excised and purified using a Quiagen Gel extraction kit. The amplified small-RNA libraries were eluted in 30 µL nuclease-free water.

To improve the quality of the amplified libraries, the purified products were subjected to the following 1-cycle PCR:

PCR reaction:

purified PCR product	18 µL
5x *Phusion* HF PCR buffer	5 µL
dNTP mix (10 mM each)	0.5 µL
oligo Illumina 5' ext. (10 µM)	0.5 µL
oligo 3' PCR primer index (specific barcode)	0.5 µL
Hot Start *Phusion* polymerase	0.5 µL

Thermocycler protocol:

98 °C	1 min	
98 °C	15 s	1 cycle
58 °C	30 s	
72 °C	30 s	
72 °C	2 min	

After amplification, the reaction mix was purified on magnetic Ampure XP beads. Therefore, 27.5 µL of dissolved Ampure XP beads were added to the 25 µL PCR reaction, mixed and incubated for 5 min. The beads with the bound DNA were then washed twice with 180 µL 70% EtOH for 30 s and after complete removal of the EtOH eluted in 11 µL buffer EB.

The thus obtained purified libraries were quantified on a 2% agarose gel and mixed to obtain the deep-sequencing sample with comparable final concentrations of each single library. The mixed sample was then quantified using a 2100 Bioanalyzer (Agilent Technologies) and analyzed via high-throughput sequencing on a HiSeq1500 instrument (Illumina). In most cases, the library was spiked-in with another sample (1 % of total reads) to check for quality and concentration prior to performing a full-lane run.

4.2.4 Protein analysis

4.2.4.1 Luciferase assay

To perform luciferase-based validation screens, the Dual-Glo Luciferase Assay System was used according to manufacturer's instructions. Therefore, the cultured reporter cells (75 µL/well in clear 96-well cell culture plates) were lysed by adding 75 µL of Dual Glow Reagent and incubating them for 10 min at room temperature and 150 rpm. 100 µL of the lysate were then transferred to black 96-well plates and luminescence of the firefly luciferase was measured for 1s/well on a SpectraMax Paradigm Plate Reader.

In order to subsequently measure the luminescence intensity of the *Renilla* luciferase, 50 µL Dual-Glo Stop & Glo Reagent were added to each well of the plate and the mix was incubated for 10 min, followed by the second luminescence detection interval of 1s/well.

4.2.4.2 Flow cytometry

The fluorescence of the GFP-based reporter cell lines was measured on a FACSCalibur flow cytometer with High Throughput Sampler (HTS). The mixing volume was set to 60 µL, fluorescence was detected in the FL1 channel (530/30) and 10000 cells were measured per 96-well. The measurement was carried out and monitored with the manufacturer's CellQuest software and secondary analysis was done using Flowing Software 2.5.1.

4.2.4.3 Protein extraction

Different methods were used for protein extraction depending on the intended analysis. To qualitatively confirm the successful tagging of endogenous proteins, 450 µL cells were harvested, washed in 1x PBS and re-suspended in 1x SDS loading buffer. After lysis at 95 °C for 10 min, the cell debris was separated by centrifugation (30 min, >12000 rpm) and the supernatant containing the extracted proteins was directly analyzed via SDS-gel electrophoresis and Western Blotting.

For more quantitative Western Blotting, the cells were harvested, twice washed in 1x PBS and then re-suspended in 150 µL 1x PBS containing 8 M urea. The lysis was carried out at 95 °C for 10 min, followed by a 10 min maximum-speed centrifugation step. To determine the protein concentration, 1 µL of the protein extract was mixed with 1 mL of Bradford reagent, incubated at room temperature for 5 min, and the absorption at 595 nm was measured on a BioPhotometer and internally compared to a BSA-standard.

Protein extracts for immunoprecipitation were generated by harvesting the cells via centrifugation, washing the pellet twice in 1x PBS and re-suspending the cells in 500 µL Lysis Buffer (LB). If necessary, the cells were diluted to obtain an estimate concentration of $1x10^8$ cells/mL. Next, the lysate was subjected to snap-freezing in liquid nitrogen, thawed and incubated for 25 min on ice. Afterwards, the

samples were sonicated using a Bioruptor (12x (20s power, 60s break); high intensity) followed by the final centrifugation step (4 min at maximum speed). The obtained protein concentration was then measured photometrically as described above.

4.2.4.4 Western Blotting

Prior to Western Blotting, appropriate amounts of protein extracts (10 to 40 µg total protein per lane) were mixed with SDS loading buffer, incubated at 95 °C for 5 min and then separated using discontinuous polyacrylamide gel electrophoresis (Laemmli 1970). Therefore, 8–12 % polyacrylamide gels were prepared using the UltraPure ProtoGel system, depending on the molecular weight of the analyzed proteins. After loading of the samples including a prestained protein marker, the gels were run in 1x SDS Running buffer at 150 V for 60-75 min in a PAGE-electrophoresis chamber.

Following electrophoresis, the gels were assembled in a blotting chamber and the proteins were transferred onto an activated polyvinylidene fluoride (PVDF) membrane for 1 h at 100 V in 1x Western transfer buffer. After immobilizing the proteins using EtOH, the membranes were washed in water and TBS-T and then blocked for 30 min in 5 % milk in TBS-T (Tween concentration optimized for the respective antibody) at room temperature. Incubation of the blocked membranes with the primary antibody was then carried out in 5 % milk in TBS-T at 4 °C overnight.

Subsequent to three 10-minute washing steps with TBS-T, the membranes were next incubated with the secondary antibody (if needed) for 2 hours at room temperature. Afterwards, the membranes were again washed three times in TBS-T and then evenly coated with Enhanced chemiluminescence (ECL) substrate (SuperSignal West Dura Extended Duration).

The resulting signal was detected using a LAS 3000 mini Western Imager or Amersham Imager 600. If further antibodies were subsequently applied to the same membrane, the membrane was stripped with RestoreTM Western Blot Stripping Buffer for 45 min and then washed several times in H_2O and TBS-T prior to re-blocking.

4.2.4.5 Co-immunoprecipitation

In order to investigate protein-protein interactions, several proteins were endogenously tagged with either FLAG-, GFP- or Strep-tags, thus requiring different protocols for (co-)immunoprecipitation.

For some of the FLAG-IP experiments, the antibodies were crosslinked to magnetic beads prior to incubation. In this case, 200 µL Dynabeads Protein G (Thermo Fisher Scientific) were incubated with 20 µL α-FLAG antibody in 400 µL PBS for 4 h at 4 °C on an overhead incubator. Then, the beads were washed twice with PBS and re-suspended in 1 mL PBS. For crosslinking, the beads were washed twice with 0.1 M borate pH 9.0 and then twice incubated with ice-cold, freshly prepared DMP/borate for 30 min at room-temperature. Afterwards, the crosslinked beads were washed in 50 mM glycine (pH 2.5),

then incubated for 30 min at RT in glycine, washed three times in PBS, checked on a 10 % SDS gel for successful crosslinking and stored in 20 % EtOH.

Instead of crosslinking, for some experiments the Dynabeads were pre-incubated with the α-FLAG-antibody (1 µL antibody per 10 µL beads in 500 µL Lysis Buffer) for 45 min prior to addition of the lysate.

Prior to immunoprecipitation, the beads (20-25 µL per sample) were washed once in 1 mL TBS-T and twice in 500 µL lysis buffer and re-suspended in the original volume of Lysis Buffer. The prepared cell lysates were then diluted with Lysis Buffer to obtain the desired amount of protein in 700 µL and added to the washed beads. To facilitate binding, the mix was incubated for 1 h at 4 °C on an overhead incubator. Afterwards, the magnetic beads were twice washed with IP-Wash Buffer 1 and then twice with IP-Wash Buffer 2. After removing all liquid, the beads were re-suspended in 30 µL 1x SDS loading buffer (4x stock diluted with lysis buffer) and heated to 95 °C for 10 min. The beads were then removed and the supernatant was analyzed by Western Blotting. Alternatively, the bound protein was eluted by re-suspending the beads in 30 µL 0.75 M arginine pH 3.5. The thus obtained elution was neutralized with 4 µL 1 M Tris-HCl pH 8.0 and supplemented with 4x SDS loading buffer.

Co-immunoprecipitations of GFP-tagged proteins were carried out using GFP-Trap_A agarose beads (ChromoTek). Therefore, the same buffers as for FLAG-IPs were applied and all washing and separation steps were performed via centrifugation at 2500 rcf and 4 °C for 2 min. The beads were not natively eluted but boiled in SDS loading buffer as described above.

The protein extracts from cells intended for immunoprecipitation of Strep-tagged proteins were blocked by adding 100 µg avidin to the lysate. The magnetic MagStrep Beads (IBA) were treated and incubated as described above for FLAG-IPs, however, the washing steps were carried out with the Washing buffer supplied with the beads (100 mM Tris/HCl, pH8; 150 mM NaCl, 1 mM EDTA). Elution was attempted with the likewise supplied Biotin Elution Buffer (Washing buffer including 2 mM D-biotin) and the beads with the still bound proteins were afterwards subjected to boiling in SDS loading buffer.

4.2.5 Bioinformatic analysis

4.2.5.1 Genome-wide screen

The primary analysis of the data obtained from the genome-wide screen was performed during the Master thesis in R/Bioconductor using the cellHTS2 package (Boutros, Bras, and Huber 2006). The raw data was log-transformed and normalized on the plate median for each channel separately with no variance adjustment applied. The mean Rluc values of both replicates were plotted against the corresponding mean Fluc values and locally weighted polynomial regression (LOESS) was utilized to fit a smoothened curve (smoothing parameter 0.9). The thus obtained LOESS values were then subtracted from the

corresponding Rluc values to acquire the LOESS-residuals (resi), based on which a z-score was calculated.

Factors with z-scores >4 and <-4 were considered preliminary positive/negative hits and were selected for validation. The z-scores of these factors are listed in Appendix 2: Detailed results from validation screens.

4.2.5.2 Validation screening

For both luciferase-based reporter validations (DSB-reporter with linearized plasmid and Klon4 cells with high-copy reporter) the Rluc/Fluc ratios were calculated and then normalized to the mean ratios of all negative controls (GFP and DsRed) separately for each replicate and dsRNA design thus obtaining the respective fold-change of the reporter signal upon knockdown. The mean fold-change of both biological replicates for each dsRNA design was then calculated and evaluated based on the standard deviation of all negative controls. Factors with fold-changes greater than the standard deviation of the negative controls for at least two out of three designs were considered confirmed hits (see Appendix 2: Detailed results from validation screens values, standard deviation at the end of the table and validation outcome). Constructs which caused a strong viability effect upon knockdown of the target were identified based on the obtained Fluc values and marked accordingly.

Regarding the GFP-reporters, the mean fluorescence (FL1) of the green cells upon knockdown was normalized to the average mean fluorescence of the negative controls. Both replicates were averaged for each design separately. The viability of the cells was evaluated by monitoring the time to reach the desired cell count.

Candidates that could be validated with the DSB-reporter but also showed strongly increased GFP-values in both the low-copy and miR-277 reporters ($\geq$ 1.6 each) were marked as false positives and not counted as validated candidates even if formally validated (see Appendix 2: Detailed results from validation screens).

Gene ontology analysis of the validated hits was performed with Gorilla (Eden et al. 2009) using the genome-wide HD2 library as background reference.

4.2.5.3 Deep-sequencing data

The datasets obtained for the deep-sequencing libraries were de-multiplexed according to the barcode of each included sample using the Galaxy platform (Giardine et al. 2005). The adaptor sequence was clipped off (6nt seed matching), thereby also removing reads from linker-to-linker ligations. Next, all sequences containing any unidentified nucleotides were excluded and a third filter was applied to select for reads with the desired read length of 19 to 25nt.

The small RNA reads were then mapped to the *D. melanogaster* genome, as well as to specific regions of interest (miRNAs, transposons, targeted genes) using bowtie (Langmead et al. 2009). Further analysis

was carried out with in-house Perl scripts (Hartig et al. 2009) counting the number of reads matching each sequence in the reference genome, the distribution of the reads along a reference sequence and the length distribution of the matched hits. The obtained results were normalized on either the miRNA/transposon-mapping reads or the number of reads mapping to the *D. melanogaster* genome to account for sample-specific differences in sequencing depth.

The original deep sequencing datasets were published at the European Nucleotide archive (ENA) (accession number PRJEB20896).

5 Literature

Abraham, R. T. 2001. 'Cell cycle checkpoint signaling through the ATM and ATR kinases', *Genes Dev*, 15: 2177-96.

Adamson, B., A. Smogorzewska, F. D. Sigoillot, R. W. King, and S. J. Elledge. 2012. 'A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response', *Nat Cell Biol*, 14: 318-28.

Aguilera, Andrés, and Hélène Gaillard. 2014. 'Transcription and recombination: when RNA meets DNA', *Cold Spring Harbor perspectives in biology*, 6: a016543.

Ameres, Stefan L., Jui-Hung Hung, Jia Xu, Zhiping Weng, and Phillip D. Zamore. 2011. 'Target RNA-directed tailing and trimming purifies the sorting of endo-siRNAs between the two Drosophila Argonaute proteins', *RNA*, 17: 54-63.

Anand, R., L. Ranjha, E. Cannavo, and P. Cejka. 2016. 'Phosphorylated CtIP Functions as a Co-factor of the MRE11-RAD50-NBS1 Endonuclease in DNA End Resection', *Mol Cell*, 64: 940-50.

Aymard, F., B. Bugler, C. K. Schmidt, E. Guillou, P. Caron, S. Briois, J. S. Iacovoni, V. Daburon, K. M. Miller, S. P. Jackson, and G. Legube. 2014. 'Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks', *Nat Struct Mol Biol*, 21: 366-74.

Bao, P., C. L. Will, H. Urlaub, K. L. Boon, and R. Lührmann. 2017. 'The RES complex is required for efficient transformation of the precatalytic B spliceosome into an activated B(act) complex', *Genes Dev*, 31: 2416-29.

Batenburg, N. L., J. R. Walker, Y. Coulombe, A. Sherker, J. Y. Masson, and X. D. Zhu. 2019. 'CSB interacts with BRCA1 in late S/G2 to promote MRN- and CtIP-mediated DNA end resection', *Nucleic Acids Res*.

Bayer, Fabienne E., Mirjam Zimmermann, Anette Preiss, and Anja C. Nagel. 2018. 'Overexpression of the Drosophila ATR homologous checkpoint kinase Mei-41 induces a G2/M checkpoint in Drosophila imaginal tissue', *Hereditas*, 155: 27-27.

Bayne, E. H., M. Portoso, A. Kagansky, I. C. Kos-Braun, T. Urano, K. Ekwall, F. Alves, J. Rappsilber, and R. C. Allshire. 2008. 'Splicing factors facilitate RNAi-directed silencing in fission yeast', *Science*, 322: 602-6.

Bekker-Jensen, S., C. Lukas, F. Melander, J. Bartek, and J. Lukas. 2005. 'Dynamic assembly and sustained retention of 53BP1 at the sites of DNA damage are controlled by Mdc1/NFBD1', *J Cell Biol*, 170: 201-11.

Birmingham, A., L. M. Selfors, T. Forster, D. Wrobel, C. J. Kennedy, E. Shanks, J. Santoyo-Lopez, D. J. Dunican, A. Long, D. Kelleher, Q. Smith, R. L. Beijersbergen, P. Ghazal, and C. E. Shamu. 2009. 'Statistical methods for analysis of high-throughput RNA interference screens', *Nat Methods*, 6: 569-75.

Blanchette, Marco, Richard E. Green, Stewart MacArthur, Angela N. Brooks, Steven E. Brenner, Michael B. Eisen, and Donald C. Rio. 2009. 'Genome-wide analysis of alternative pre-mRNA splicing and RNA binding specificities of the Drosophila hnRNP A/B family members', *Mol Cell*, 33: 438-49.

Bonath, Franziska, Judit Domingo-Prim, Marcel Tarbier, Marc R. Friedländer, and Neus Visa. 2018. 'Next-generation sequencing reveals two populations of damage-induced small RNAs at endogenous DNA double-strand breaks', *Nucleic acids research*, 46: 11869-82.

Bonnet, A., A. R. Grosso, A. Elkaoutari, E. Coleno, A. Presle, S. C. Sridhara, G. Janbon, V. Geli, S. F. de Almeida, and B. Palancade. 2017. 'Introns Protect Eukaryotic Genomes from Transcription-Associated Genetic Instability', *Mol Cell*, 67: 608-21.e6.

Bottcher, R., M. Hollmann, K. Merk, V. Nitschko, C. Obermaier, J. Philippou-Massier, I. Wieland, U. Gaul, and K. Forstemann. 2014. 'Efficient chromosomal gene modification with CRISPR/cas9 and PCR-based homologous recombination donors in cultured Drosophila cells', *Nucleic Acids Res*, 42: e89.

Böttcher, R., I. Schmidts, V. Nitschko, P. Duric, and K. Förstemann. 2021. 'RNA polymerase II is recruited to DNA double-strand breaks for dilncRNA transcription in Drosophila', *RNA Biol*: 1-10.

Boutros, M., L. P. Bras, and W. Huber. 2006. 'Analysis of cell-based RNAi screens', *Genome Biol*, 7: R66.

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

Brooks, A. N., M. O. Duff, G. May, L. Yang, M. Bolisetty, J. Landolin, K. Wan, J. Sandler, B. W. Booth, S. E. Celniker, B. R. Graveley, and S. E. Brenner. 2015. 'Regulation of alternative splicing in Drosophila by 56 RNA binding proteins', *Genome Res*, 25: 1771-80.

Brown, James B., Nathan Boley, Robert Eisman, Gemma E. May, Marcus H. Stoiber, Michael O. Duff, Ben W. Booth, Jiayu Wen, Soo Park, Ana Maria Suzuki, Kenneth H. Wan, Charles Yu, Dayu Zhang, Joseph W. Carlson, Lucy Cherbas, Brian D. Eads, David Miller, Keithanne Mockaitis, Johnny Roberts, Carrie A. Davis, Erwin Frise, Ann S. Hammonds, Sara Olson, Sol Shenker, David Sturgill, Anastasia A. Samsonova, Richard Weiszmann, Garret Robinson, Juan Hernandez, Justen Andrews, Peter J. Bickel, Piero Carninci, Peter Cherbas, Thomas R. Gingeras, Roger A. Hoskins, Thomas C. Kaufman, Eric C. Lai, Brian Oliver, Norbert Perrimon, Brenton R. Graveley, and Susan E. Celniker. 2014. 'Diversity and dynamics of the Drosophila transcriptome', *Nature*, 512: 393.

Bu, Ying, Huan Huang, and Guohua Zhou. 2008. 'Direct polymerase chain reaction (PCR) from human whole blood and filter-paper-dried blood by using a PCR buffer with a higher pH', *Analytical Biochemistry*, 375: 370-72.

Carissimi, C., I. Laudadio, E. Cipolletta, S. Gioiosa, M. Mihailovich, T. Bonaldi, G. Macino, and V. Fulci. 2015. 'ARGONAUTE2 cooperates with SWI/SNF complex to determine nucleosome occupancy at human Transcription Start Sites', *Nucleic Acids Res*, 43: 1498-512.

Carthew, R. W., and E. J. Sontheimer. 2009. 'Origins and Mechanisms of miRNAs and siRNAs', *Cell*, 136: 642-55.

Castel, S. E., J. Ren, S. Bhattacharjee, A. Y. Chang, M. Sanchez, A. Valbuena, F. Antequera, and R. A. Martienssen. 2014. 'Dicer promotes transcription termination at sites of replication stress to maintain genome stability', *Cell*, 159: 572-83.

Castellano-Pozo, M., J. M. Santos-Pereira, A. G. Rondon, S. Barroso, E. Andujar, M. Perez-Alegre, T. Garcia-Muse, and A. Aguilera. 2013. 'R loops are linked to histone H3 S10 phosphorylation and chromatin condensation', *Mol Cell*, 52: 583-90.

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

Cejka, P. 2015. 'DNA End Resection: Nucleases Team Up with the Right Partners to Initiate Homologous Recombination', *J Biol Chem*, 290: 22931-8.

Celeste, A., O. Fernandez-Capetillo, M. J. Kruhlak, D. R. Pilch, D. W. Staudt, A. Lee, R. F. Bonner, W. M. Bonner, and A. Nussenzweig. 2003. 'Histone H2AX phosphorylation is dispensable for the initial recognition of DNA breaks', *Nat Cell Biol*, 5: 675-9.

Cerosaletti, Karen, Jocyndra Wright, and Patrick Concannon. 2006. 'Active role for nibrin in the kinetics of atm activation', *Molecular and Cellular Biology*, 26: 1691-99.

Chakraborty, A., N. Tapryal, T. Venkova, N. Horikoshi, R. K. Pandita, A. H. Sarker, P. S. Sarkar, T. K. Pandita, and T. K. Hazra. 2016. 'Classical non-homologous end-joining pathway utilizes nascent RNA for error-free double-strand break repair of transcribed genes', *Nat Commun*, 7: 13049.

Chan, S. H., A. M. Yu, and M. McVey. 2010. 'Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in Drosophila', *PLoS Genet*, 6: e1001005.

Chanarat, Sittinan, and Katja Sträßer. 2013. 'Splicing and beyond: The many faces of the Prp19 complex', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1833: 2126-34.

Chang, H. H. Y., N. R. Pannunzio, N. Adachi, and M. R. Lieber. 2017. 'Non-homologous DNA end joining and alternative pathways to double-strand break repair', *Nat Rev Mol Cell Biol*, 18: 495-506.

Chapman, J. R., and S. P. Jackson. 2008. 'Phospho-dependent interactions between NBS1 and MDC1 mediate chromatin retention of the MRN complex at sites of DNA damage', *EMBO Rep*, 9: 795-801.

Chen, H., K. Kobayashi, A. Miyao, H. Hirochika, N. Yamaoka, and M. Nishiguchi. 2013. 'Both OsRecQ1 and OsRDR1 are required for the production of small RNA in response to DNA-damage in rice', *PLoS One*, 8: e55252.

Chiu, Yu-Hsin, John B. Macmillan, and Zhijian J. Chen. 2009. 'RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway', *Cell*, 138: 576-91.

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

Cogoni, C., and G. Macino. 1999. 'Gene silencing in Neurospora crassa requires a protein homologous to RNA-dependent RNA polymerase', *Nature*, 399: 166-9.

Cogoni, Carlo, and Giuseppe Macino. 1997. 'Isolation of quelling-defective (qde) mutants impaired in posttranscriptional transgene-induced gene silencing in Neurospora crassa', *Proc Natl Acad Sci U S A*, 94: 10233-8.

Cruz, C., and J. Houseley. 2014. 'Endogenous RNA interference is driven by copy number', *Elife*, 3: e01581.

Czech, B., and G. J. Hannon. 2011. 'Small RNA sorting: matchmaking for Argonautes', *Nat Rev Genet*, 12: 19-31.

Czech, B., C. D. Malone, R. Zhou, A. Stark, C. Schlingeheyde, M. Dus, N. Perrimon, M. Kellis, J. A. Wohlschlegel, R. Sachidanandam, G. J. Hannon, and J. Brennecke. 2008. 'An endogenous small interfering RNA pathway in Drosophila', *Nature*, 453: 798-802.

de Vries, H. I., L. Uyetake, W. Lemstra, J. F. Brunsting, T. T. Su, H. H. Kampinga, and O. C. Sibon. 2005. 'Grp/DChk1 is required for G2-M checkpoint activation in Drosophila S2 cells, whereas Dmnk/DChk2 is dispensable', *J Cell Sci*, 118: 1833-42.

Denli, Ahmet M., Bastiaan B. J. Tops, Ronald H. A. Plasterk, René F. Ketting, and Gregory J. Hannon. 2004. 'Processing of primary microRNAs by the Microprocessor complex', *Nature*, 432: 231-35.

Dieci, G., G. Fiorino, M. Castelnuovo, M. Teichmann, and A. Pagano. 2007. 'The expanding RNA polymerase III transcriptome', *Trends Genet*, 23: 614-22.

Doré, Andrew S., Adam C. B. Drake, Suzanne C. Brewerton, and Tom L. Blundell. 2004. 'Identification of DNA-PK in the arthropods: Evidence for the ancient ancestry of vertebrate non-homologous end-joining', *DNA Repair (Amst)*, 3: 33-41.

Du, C., M. E. McGuffin, B. Dauwalder, L. Rabinow, and W. Mattox. 1998. 'Protein phosphorylation plays an essential role in the regulation of alternative splicing and sex determination in Drosophila', *Mol Cell*, 2: 741-50.

Dumesic, P. A., P. Natarajan, C. Chen, I. A. Drinnenberg, B. J. Schiller, J. Thompson, J. J. Moresco, J. R. Yates, 3rd, D. P. Bartel, and H. D. Madhani. 2013. 'Stalled spliceosomes are a signal for RNAi-mediated genome defense', *Cell*, 152: 957-68.

Eden, E., R. Navon, I. Steinfeld, D. Lipson, and Z. Yakhini. 2009. 'GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists', *BMC Bioinformatics*, 10: 48.

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

Fan, Y., M. Schlierf, A. C. Gaspar, C. Dreux, A. Kpebe, L. Chaney, A. Mathieu, C. Hitte, O. Gremy, E. Sarot, M. Horn, Y. Zhao, T. G. Kinzy, and L. Rabinow. 2010. 'Drosophila translational elongation factor-1gamma is modified in response to DOA kinase activity and is essential for cellular viability', *Genetics*, 184: 141-54.

Farkaš, Robert, Michaela Kovacikova, Denisa Liszekova, Milan Beno, Peter Danis, Leonard J. Rabinow, Bruce A. Chase, and Ivan Raška. 2014. 'Exploring some of the physico-chemical properties of the LAMMER protein kinase DOA of Drosophila', *Fly (Austin)*, 3: 130-42.

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

Forstemann, K., M. D. Horwich, L. Wee, Y. Tomari, and P. D. Zamore. 2007. 'Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1', *Cell*, 130: 287-97.

Forstemann, K., Y. Tomari, T. Du, V. V. Vagin, A. M. Denli, D. P. Bratu, C. Klattenhoff, W. E. Theurkauf, and P. D. Zamore. 2005. 'Normal microRNA maturation and germ-line stem cell

maintenance requires Loquacious, a double-stranded RNA-binding domain protein', *PLoS Biol*, 3: e236.

Francia, S., M. Cabrini, V. Matti, A. Oldani, and F. d'Adda di Fagagna. 2016. 'DICER, DROSHA and DNA damage response RNAs are necessary for the secondary recruitment of DNA damage response factors', *J Cell Sci*, 129: 1468-76.

Francia, S., F. Michelini, A. Saxena, D. Tang, M. de Hoon, V. Anelli, M. Mione, P. Carninci, and F. d'Adda di Fagagna. 2012. 'Site-specific DICER and DROSHA RNA products control the DNA-damage response', *Nature*, 488: 231-5.

Gaillard, H., and A. Aguilera. 2013. 'Transcription coupled repair at the interface between transcription elongation and mRNP biogenesis', *Biochim Biophys Acta*, 1829: 141-50.

Gao, M., W. Wei, M. M. Li, Y. S. Wu, Z. Ba, K. X. Jin, M. M. Li, Y. Q. Liao, S. Adhikari, Z. Chong, T. Zhang, C. X. Guo, T. S. Tang, B. T. Zhu, X. Z. Xu, N. Mailand, Y. G. Yang, Y. Qi, and J. M. Rendtlew Danielsen. 2014. 'Ago2 facilitates Rad51 recruitment and DNA double-strand break repair by homologous recombination', *Cell Res*, 24: 532-41.

Garcia, Valerie, Sarah E. L. Phelps, Stephen Gray, and Matthew J. Neale. 2011. 'Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1', *Nature*, 479: 241-44.

Gerbasi, V. R., J. B. Preall, D. E. Golden, D. W. Powell, T. D. Cummins, and E. J. Sontheimer. 2011. 'Blanks, a nuclear siRNA/dsRNA-binding complex component, is required for Drosophila spermiogenesis', *Proc Natl Acad Sci U S A*, 108: 3204-9.

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

Ghodke, Indrajeet, and K. Muniyappa. 2013. 'Processing of DNA Double-stranded Breaks and Intermediates of Recombination and Repair by Saccharomyces cerevisiae Mre11 and Its Stimulation by Rad50, Xrs2, and Sae2 Proteins', *Journal of Biological Chemistry*, 288: 11273-86.

Giardine, B., C. Riemer, R. C. Hardison, R. Burhans, L. Elnitski, P. Shah, Y. Zhang, D. Blankenberg, I. Albert, J. Taylor, W. Miller, W. J. Kent, and A. Nekrutenko. 2005. 'Galaxy: a platform for interactive large-scale genome analysis', *Genome Res*, 15: 1451-5.

Graveley, B. R., G. May, A. N. Brooks, J. W. Carlson, L. Cherbas, C. A. Davis, M. Duff, B. Eads, J. Landolin, J. Sandler, K. H. Wan, J. Andrews, S. E. Brenner, P. Cherbas, T. R. Gingeras, R. Hoskins, T. Kaufman, and S. E. Celniker. 2011. "The D. melanogaster transcriptome: modENCODE RNA-Seq data for cell lines."

Hammond, L. E., D. Z. Rudner, R. Kanaar, and D. C. Rio. 1997. 'Mutations in the hrp48 gene, which encodes a Drosophila heterogeneous nuclear ribonucleoprotein particle protein, cause lethality and developmental defects and affect P-element third-intron splicing in vivo', *Molecular and Cellular Biology*, 17: 7260-67.

Hamperl, Stephan, and Karlene A. Cimprich. 2014. 'The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability', *DNA Repair (Amst)*, 19: 84-94.

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

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

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

Herold, N., C. L. Will, E. Wolf, B. Kastner, H. Urlaub, and R. Luhrmann. 2009. 'Conservation of the protein composition and electron microscopy structure of Drosophila melanogaster and human spliceosomal complexes', *Mol Cell Biol*, 29: 281-301.

Hoeijmakers, J. H. 2009. 'DNA damage, aging, and cancer', *N Engl J Med*, 361: 1475-85.

Hoskins, A. A., and M. J. Moore. 2012. 'The spliceosome: a flexible, reversible macromolecular machine', *Trends Biochem Sci*, 37: 179-88.

Hsin, Jing-Ping, and James L. Manley. 2012. 'The RNA polymerase II CTD coordinates transcription and RNA processing', *Genes Dev*, 26: 2119-37.

Ishizuka, A., M. C. Siomi, and H. Siomi. 2002. 'A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins', *Genes Dev*, 16: 2497-508.

Jiang, Fuguo, David W. Taylor, Janice S. Chen, Jack E. Kornfeld, Kaihong Zhou, Aubri J. Thompson, Eva Nogales, and Jennifer A. Doudna. 2016. 'Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage', *Science*, 351: 867-71.

Kallgren, S. P., S. Andrews, X. Tadeo, H. Hou, J. J. Moresco, P. G. Tu, J. R. Yates, 3rd, P. L. Nagy, and S. Jia. 2014. 'The proper splicing of RNAi factors is critical for pericentric heterochromatin assembly in fission yeast', *PLoS Genet*, 10: e1004334.

Kashammer, L., J. H. Saathoff, K. Lammens, F. Gut, J. Bartho, A. Alt, B. Kessler, and K. P. Hopfner. 2019. 'Mechanism of DNA End Sensing and Processing by the Mre11-Rad50 Complex', *Mol Cell*, 76: 382-94.e6.

Katzenberger, R. J., M. S. Marengo, and D. A. Wassarman. 2009. 'Control of alternative splicing by signal-dependent degradation of splicing-regulatory proteins', *J Biol Chem*, 284: 10737-46.

Kawamata, T., and Y. Tomari. 2010. 'Making RISC', *Trends Biochem Sci*, 35: 368-76.

Kent, T., G. Chandramouly, S. M. McDevitt, A. Y. Ozdemir, and R. T. Pomerantz. 2015. 'Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase theta', *Nat Struct Mol Biol*, 22: 230-7.

Keskin, H., Y. Shen, F. Huang, M. Patel, T. Yang, K. Ashley, A. V. Mazin, and F. Storici. 2014. 'Transcript-RNA-templated DNA recombination and repair', *Nature*, 515: 436-9.

Khodor, Y. L., J. Rodriguez, K. C. Abruzzi, C. H. Tang, M. T. Marr, 2nd, and M. Rosbash. 2011. 'Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila', *Genes Dev*, 25: 2502-12.

Kpebe, A., and L. Rabinow. 2008. 'Dissection of darkener of apricot kinase isoform functions in Drosophila', *Genetics*, 179: 1973-87.

Kunzelmann, S., R. Bottcher, I. Schmidts, and K. Forstemann. 2016. 'A Comprehensive Toolbox for Genome Editing in Cultured Drosophila melanogaster Cells', *G3 (Bethesda)*, 6: 1777-85.

Kunzelmann, S., and K. Forstemann. 2017. 'Reversible perturbations of gene regulation after genome editing in Drosophila cells', *PLoS One*, 12: e0180135.

Kurzynska-Kokorniak, A., N. Koralewska, M. Pokornowska, A. Urbanowicz, A. Tworak, A. Mickiewicz, and M. Figlerowicz. 2015. 'The many faces of Dicer: the complexity of the mechanisms regulating Dicer gene expression and enzyme activities', *Nucleic Acids Res*, 43: 4365-80.

Laemmli, U. K. 1970. 'Cleavage of structural proteins during the assembly of the head of bacteriophage T4', *Nature*, 227: 680-5.

Lai, W., H. Li, S. Liu, and Y. Tao. 2013. 'Connecting chromatin modifying factors to DNA damage response', *Int J Mol Sci*, 14: 2355-69.

Langerak, Petra, Eva Mejia-Ramirez, Oliver Limbo, and Paul Russell. 2011. 'Release of Ku and MRN from DNA ends by Mre11 nuclease activity and Ctp1 is required for homologous recombination repair of double-strand breaks', *PLoS Genet*, 7: e1002271-e71.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg. 2009. 'Ultrafast and memory-efficient alignment of short DNA sequences to the human genome', *Genome Biol*, 10: R25.

LaRocque, J. R., B. Jaklevic, T. T. Su, and J. Sekelsky. 2007. 'Drosophila ATR in double-strand break repair', *Genetics*, 175: 1023-33.

Lee, H. C., S. S. Chang, S. Choudhary, A. P. Aalto, M. Maiti, D. H. Bamford, and Y. Liu. 2009. 'qiRNA is a new type of small interfering RNA induced by DNA damage', *Nature*, 459: 274-7.

Lee, H. Y., K. Zhou, A. M. Smith, C. L. Noland, and J. A. Doudna. 2013. 'Differential roles of human Dicer-binding proteins TRBP and PACT in small RNA processing', *Nucleic Acids Res*, 41: 6568-76.

Lee, J. H., and T. T. Paull. 2004. 'Direct activation of the ATM protein kinase by the Mre11/Rad50/Nbs1 complex', *Science*, 304: 93-6.

Lee, K., C. Du, M. Horn, and L. Rabinow. 1996. 'Activity and autophosphorylation of LAMMER protein kinases', *J Biol Chem*, 271: 27299-303.

Lee, R. C., R. L. Feinbaum, and V. Ambros. 1993. 'The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14', *Cell*, 75: 843-54.

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

Lenzken, S. C., A. Loffreda, and S. M. Barabino. 2013. 'RNA splicing: a new player in the DNA damage response', *Int J Cell Biol*, 2013: 153634.

Li, Y., J. Lu, Y. Han, X. Fan, and S. W. Ding. 2013. 'RNA interference functions as an antiviral immunity mechanism in mammals', *Science*, 342: 231-4.

Liao, S., M. Tammaro, and H. Yan. 2016. 'The structure of ends determines the pathway choice and Mre11 nuclease dependency of DNA double-strand break repair', *Nucleic Acids Res*, 44: 5689-701.

Liu, S., Y. Hua, J. Wang, L. Li, J. Yuan, B. Zhang, Z. Wang, J. Ji, and D. Kong. 2021. 'RNA polymerase III is required for the repair of DNA double-strand breaks by homologous recombination', *Cell*, 184: 1314-29.e10.

Livak, K. J., and T. D. Schmittgen. 2001. 'Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method', *Methods*, 25: 402-8.

Lou, Z., K. Minter-Dykhouse, S. Franco, M. Gostissa, M. A. Rivera, A. Celeste, J. P. Manis, J. van Deursen, A. Nussenzweig, T. T. Paull, F. W. Alt, and J. Chen. 2006. 'MDC1 maintains genomic stability by participating in the amplification of ATM-dependent DNA damage signals', *Mol Cell*, 21: 187-200.

Lu, Wei-Ting, Ben R. Hawley, George L. Skalka, Robert A. Baldock, Ewan M. Smith, Aldo S. Bader, Michal Malewicz, Felicity Z. Watts, Ania Wilczynska, and Martin Bushell. 2018. 'Drosha drives the formation of DNA:RNA hybrids around DNA break sites to facilitate DNA repair', *Nat Commun*, 9: 532-32.

Lukas, C., F. Melander, M. Stucki, J. Falck, S. Bekker-Jensen, M. Goldberg, Y. Lerenthal, S. P. Jackson, J. Bartek, and J. Lukas. 2004. 'Mdc1 couples DNA double-strand break recognition by Nbs1 with its H2AX-dependent chromatin retention', *EMBO J*, 23: 2674-83.

Lukas, J., C. Lukas, and J. Bartek. 2011. 'More than just a focus: The chromatin response to DNA damage and its role in genome integrity maintenance', *Nat Cell Biol*, 13: 1161-9.

Madigan, J. P., H. L. Chotkowski, and R. L. Glaser. 2002. 'DNA double-strand break-induced phosphorylation of Drosophila histone variant H2Av helps prevent radiation-induced apoptosis', *Nucleic Acids Res*, 30: 3698-705.

Mahaney, B. L., K. Meek, and S. P. Lees-Miller. 2009. 'Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining', *Biochem J*, 417: 639-50.

Malone, C. D., J. Brennecke, M. Dus, A. Stark, W. R. McCombie, R. Sachidanandam, and G. J. Hannon. 2009. 'Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary', *Cell*, 137: 522-35.

Malone, C. D., C. Mestdagh, J. Akhtar, N. Kreim, P. Deinhard, R. Sachidanandam, J. Treisman, and J. Y. Roignant. 2014. 'The exon junction complex controls transposable element activity by ensuring faithful splicing of the piwi transcript', *Genes Dev*, 28: 1786-99.

Marechal, A., J. M. Li, X. Y. Ji, C. S. Wu, S. A. Yazinski, H. D. Nguyen, S. Liu, A. E. Jimenez, J. Jin, and L. Zou. 2014. 'PRP19 transforms into a sensor of RPA-ssDNA after DNA damage and drives ATR activation via a ubiquitin-mediated circuitry', *Mol Cell*, 53: 235-46.

Marengo, M. S., and D. A. Wassarman. 2008. 'A DNA damage signal activates and derepresses exon inclusion in Drosophila TAF1 alternative splicing', *RNA*, 14: 1681-95.

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

Matsuoka, S., B. A. Ballif, A. Smogorzewska, E. R. McDonald, 3rd, K. E. Hurov, J. Luo, C. E. Bakalarski, Z. Zhao, N. Solimini, Y. Lerenthal, Y. Shiloh, S. P. Gygi, and S. J. Elledge. 2007. 'ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage', *Science*, 316: 1160-6.

Mayer, A., M. Heidemann, M. Lidschreiber, A. Schreieck, M. Sun, C. Hintermair, E. Kremmer, D. Eick, and P. Cramer. 2012. 'CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II', *Science*, 336: 1723-5.

McVey, M., and S. E. Lee. 2008. 'MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings', *Trends Genet*, 24: 529-38.

Mehta, Anuja, and James E. Haber. 2014. 'Sources of DNA double-strand breaks and models of recombinational DNA repair', *Cold Spring Harbor perspectives in biology*, 6: a016428-a28.

Meinhart, A., and P. Cramer. 2004. 'Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors', *Nature*, 430: 223-6.

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

Meister, G., and T. Tuschl. 2004. 'Mechanisms of gene silencing by double-stranded RNA', *Nature*, 431: 343-9.

Melander, F., S. Bekker-Jensen, J. Falck, J. Bartek, N. Mailand, and J. Lukas. 2008. 'Phosphorylation of SDT repeats in the MDC1 N terminus triggers retention of NBS1 at the DNA damage-modified chromatin', *J Cell Biol*, 181: 213-26.

Merk, Karin, Marco Breinig, Romy Böttcher, Stefan Krebs, Helmut Blum, Michael Boutros, and Klaus Förstemann. 2017. 'Splicing stimulates siRNA formation at Drosophila DNA double-strand breaks', *PLoS Genet*, 13: e1006861-e61.

Merk, Karin M. 2013. 'A genome-wide RNAi screen in Drosophila for factors involved in the endo-siRNA response to DNA double-strand breaks (Master Thesis).', *LMU Munich, Faculty for Chemistry and Pharmacy, Department of Biochemistry*.

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

Michelini, Flavia, Sethuramasundaram Pitchiaya, Valerio Vitelli, Sheetal Sharma, Ubaldo Gioia, Fabio Pessina, Matteo Cabrini, Yejun Wang, Ilaria Capozzo, Fabio Iannelli, Valentina Matti, Sofia Francia, G. V. Shivashankar, Nils G. Walter, and Fabrizio d'Adda di Fagagna. 2017. 'Damage-induced lncRNAs control the DNA damage response through interaction with DDRNAs at individual double-strand breaks', *Nat Cell Biol*, 19: 1400-11.

Miki, Daisuke, Peiying Zhu, Wencan Zhang, Yanfei Mao, Zhengyan Feng, Huan Huang, Hui Zhang, Yanqiang Li, Renyi Liu, Huiming Zhang, Yijun Qi, and Jian-Kang Zhu. 2017. 'Efficient Generation of diRNAs Requires Components in the Posttranscriptional Gene Silencing Pathway', *Scientific reports*, 7: 301-01.

Mikolaskova, B., M. Jurcik, I. Cipakova, M. Kretova, M. Chovanec, and L. Cipak. 2018. 'Maintenance of genome stability: the unifying role of interconnections between the DNA damage response and RNA-processing pathways', *Curr Genet*, 64: 971-83.

Mirkin, Ekaterina V., and Sergei M. Mirkin. 2007. 'Replication fork stalling at natural impediments', *Microbiology and molecular biology reviews : MMBR*, 71: 13-35.

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

Morimoto, S., M. Tsuda, H. Bunch, H. Sasanuma, C. Austin, and S. Takeda. 2019. 'Type II DNA Topoisomerases Cause Spontaneous Double-Strand Breaks in Genomic DNA', *Genes (Basel)*, 10.

Mount, S. M., C. Burks, G. Hertz, G. D. Stormo, O. White, and C. Fields. 1992. 'Splicing signals in Drosophila: intron size, information content, and consensus sequences', *Nucleic Acids Res*, 20: 4255-62.

Moy, R. H., B. S. Cole, A. Yasunaga, B. Gold, G. Shankarling, A. Varble, J. M. Molleston, B. R. tenOever, K. W. Lynch, and S. Cherry. 2014. 'Stem-loop recognition by DDX17 facilitates miRNA processing and antiviral defense', *Cell*, 158: 764-77.

Neish, A. S., S. F. Anderson, B. P. Schlegel, W. Wei, and J. D. Parvin. 1998. 'Factors associated with the mammalian RNA polymerase II holoenzyme', *Nucleic Acids Res*, 26: 847-53.

Nelson, M. R., H. Luo, H. K. Vari, B. J. Cox, A. J. Simmonds, H. M. Krause, H. D. Lipshitz, and C. A. Smibert. 2007. 'A multiprotein complex that mediates translational enhancement in Drosophila', *J Biol Chem*, 282: 34031-8.

Nikolakaki, E., C. Du, J. Lai, T. Giannakouros, L. Cantley, and L. Rabinow. 2002. 'Phosphorylation by LAMMER protein kinases: determination of a consensus site, identification of in vitro substrates, and implications for substrate preferences', *Biochemistry*, 41: 2055-66.

Nilsson, K., C. Wu, and S. Schwartz. 2018. 'Role of the DNA Damage Response in Human Papillomavirus RNA Splicing and Polyadenylation', *Int J Mol Sci*, 19.

Nitschko, V., S. Kunzelmann, T. Fröhlich, G. J. Arnold, and K. Förstemann. 2020. 'Trafficking of siRNA precursors by the dsRBD protein Blanks in Drosophila', *Nucleic Acids Res*, 48: 3906-21.

Nolan, T., L. Braccini, G. Azzalin, A. De Toni, G. Macino, and C. Cogoni. 2005. 'The post-transcriptional gene silencing machinery functions independently of DNA methylation to repress a LINE1-like retrotransposon in Neurospora crassa', *Nucleic Acids Res*, 33: 1564-73.

Ohle, C., R. Tesorero, G. Schermann, N. Dobrev, I. Sinning, and T. Fischer. 2016. 'Transient RNA-DNA Hybrids Are Required for Efficient Double-Strand Break Repair', *Cell*, 167: 1001-13.e7.

Oikemus, S. R., J. Queiroz-Machado, K. Lai, N. McGinnis, C. Sunkel, and M. H. Brodsky. 2006. 'Epigenetic telomere protection by Drosophila DNA damage response pathways', *PLoS Genet*, 2: e71.

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

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

Okamura, K., A. Ishizuka, H. Siomi, and M. C. Siomi. 2004. 'Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways', *Genes Dev*, 18: 1655-66.

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

Okamura, Katsutomo, Joshua W. Hagen, Hong Duan, David M. Tyler, and Eric C. Lai. 2007. 'The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila', *Cell*, 130: 89-100.

Panier, S., and D. Durocher. 2013. 'Push back to respond better: regulatory inhibition of the DNA double-strand break response', *Nat Rev Mol Cell Biol*, 14: 661-72.

Pankotai, T., C. Bonhomme, D. Chen, and E. Soutoglou. 2012. 'DNAPKcs-dependent arrest of RNA polymerase II transcription in the presence of DNA breaks', *Nat Struct Mol Biol*, 19: 276-82.

Pankotai, T., and E. Soutoglou. 2013. 'Double strand breaks: hurdles for RNA polymerase II transcription?', *Transcription*, 4: 34-8.

Park, J. W., K. Parisky, A. M. Celotto, R. A. Reenan, and B. R. Graveley. 2004. 'Identification of alternative splicing regulators by RNA interference in Drosophila', *Proc Natl Acad Sci U S A*, 101: 15974-9.

Pederiva, Chiara, and Marianne Farnebo. 2018. 'RNF8 - The Achilles heel of DNA repair when splicing rules', *Cell cycle (Georgetown, Tex.)*, 17: 137-38.

Portz, Bede, Feiyue Lu, Eric B. Gibbs, Joshua E. Mayfield, M. Rachel Mehaffey, Yan Jessie Zhang, Jennifer S. Brodbelt, Scott A. Showalter, and David S. Gilmour. 2017. 'Structural heterogeneity in the intrinsically disordered RNA polymerase II C-terminal domain', *Nat Commun*, 8: 15231.

Prado, Félix, and Andrés Aguilera. 2005. 'Impairment of replication fork progression mediates RNA polII transcription-associated recombination', *The EMBO journal*, 24: 1267-76.

Prados-Carvajal, R., A. Lopez-Saavedra, C. Cepeda-Garcia, S. Jimeno, and P. Huertas. 2018. 'Multiple roles of the splicing complex SF3B in DNA end resection and homologous recombination', *DNA Repair (Amst)*, 66-67: 11-23.

Puget, N., K. M. Miller, and G. Legube. 2019. 'Non-canonical DNA/RNA structures during Transcription-Coupled Double-Strand Break Repair: Roadblocks or Bona fide repair intermediates?', *DNA Repair (Amst)*, 81: 102661.

Rabinow, L., and J. A. Birchler. 1989. 'A dosage-sensitive modifier of retrotransposon-induced alleles of the Drosophila white locus', *EMBO J*, 8: 879-89.

Rabinow, L., S. L. Chiang, and J. A. Birchler. 1993. 'Mutations at the Darkener of apricot locus modulate transcript levels of copia and copia-induced mutations in Drosophila melanogaster', *Genetics*, 134: 1175-85.

Rabinow, L., and M. L. Samson. 2010. 'The role of the Drosophila LAMMER protein kinase DOA in somatic sex determination', *J Genet*, 89: 271-7.

Renkawitz, J., C. A. Lademann, and S. Jentsch. 2014. 'Mechanisms and principles of homology search during recombination', *Nat Rev Mol Cell Biol*, 15: 369-83.

Reuter, L. M., and K. Strasser. 2016. 'Falling for the dark side of transcription: Nab2 fosters RNA polymerase III transcription', *Transcription*, 7: 69-74.

Romano, N., and G. Macino. 1992. 'Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences', *Mol Microbiol*, 6: 3343-53.

Rother, S., and G. Meister. 2011. 'Small RNAs derived from longer non-coding RNAs', *Biochimie*, 93: 1905-15.

Russo, J., A. W. Harrington, and M. Steiniger. 2016. 'Antisense Transcription of Retrotransposons in Drosophila: An Origin of Endogenous Small Interfering RNA Precursors', *Genetics*, 202: 107-21.

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

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

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

Serpinskaya, A. S., K. Tuphile, L. Rabinow, and V. I. Gelfand. 2014. 'Protein kinase Darkener of apricot and its substrate EF1gamma regulate organelle transport along microtubules', *J Cell Sci*, 127: 33-9.

Shkreta, L., and B. Chabot. 2015. 'The RNA Splicing Response to DNA Damage', *Biomolecules*, 5: 2935-77.

Sikorski, Timothy W., Scott B. Ficarro, John Holik, TaeSoo Kim, Oliver J. Rando, Jarrod A. Marto, and Stephen Buratowski. 2011. 'Sub1 and RPA associate with RNA polymerase II at different stages of transcription', *Mol Cell*, 44: 397-409.

Skourti-Stathaki, K., K. Kamieniarz-Gdula, and N. J. Proudfoot. 2014. 'R-loops induce repressive chromatin marks over mammalian gene terminators', *Nature*, 516: 436-9.

Sollier, J., and K. A. Cimprich. 2015. 'Breaking bad: R-loops and genome integrity', *Trends Cell Biol*, 25: 514-22.

Sollier, J., C. T. Stork, M. L. García-Rubio, R. D. Paulsen, A. Aguilera, and K. A. Cimprich. 2014. 'Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability', *Mol Cell*, 56: 777-85.

Symington, L. S., and J. Gautier. 2011. 'Double-strand break end resection and repair pathway choice', *Annu Rev Genet*, 45: 247-71.

Székvölgyi, Lóránt, Kunihiro Ohta, and Alain Nicolas. 2015. 'Initiation of meiotic homologous recombination: flexibility, impact of histone modifications, and chromatin remodeling', *Cold Spring Harbor perspectives in biology*, 7: a016527.

Taliaferro, J. M., J. L. Aspden, T. Bradley, D. Marwha, M. Blanchette, and D. C. Rio. 2013. 'Two new and distinct roles for Drosophila Argonaute-2 in the nucleus: alternative pre-mRNA splicing and transcriptional repression', *Genes Dev*, 27: 378-89.

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

Tresini, M., J. A. Marteijn, and W. Vermeulen. 2016. 'Bidirectional coupling of splicing and ATM signaling in response to transcription-blocking DNA damage', *RNA Biol*, 13: 272-8.

Tresini, M., D. O. Warmerdam, P. Kolovos, L. Snijder, M. G. Vrouwe, J. A. Demmers, IJcken W. F. van, F. G. Grosveld, R. H. Medema, J. H. Hoeijmakers, L. H. Mullenders, W. Vermeulen, and J. A. Marteijn. 2015. 'The core spliceosome as target and effector of non-canonical ATM signalling', *Nature*, 523: 53-8.

Uanschou, C., T. Siwiec, A. Pedrosa-Harand, C. Kerzendorfer, E. Sanchez-Moran, M. Novatchkova, S. Akimcheva, A. Woglar, F. Klein, and P. Schlogelhofer. 2007. 'A novel plant gene essential for meiosis is related to the human CtIP and the yeast COM1/SAE2 gene', *EMBO J*, 26: 5061-70.

Ulvila, J., M. Parikka, A. Kleino, R. Sormunen, R. A. Ezekowitz, C. Kocks, and M. Ramet. 2006. 'Double-stranded RNA is internalized by scavenger receptor-mediated endocytosis in Drosophila S2 cells', *J Biol Chem*, 281: 14370-5.

Uziel, Tamar, Yaniv Lerenthal, Lilach Moyal, Yair Andegeko, Leonid Mittelman, and Yosef Shiloh. 2003. 'Requirement of the MRN complex for ATM activation by DNA damage', *The EMBO journal*, 22: 5612-21.

Valadkhan, S., and Y. Jaladat. 2010. 'The spliceosomal proteome: at the heart of the largest cellular ribonucleoprotein machine', *Proteomics*, 10: 4128-41.

van der Krol, A. R., L. A. Mur, M. Beld, J. N. Mol, and A. R. Stuitje. 1990. 'Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression', *The Plant cell*, 2: 291-99.

van Gent, D. C., J. H. Hoeijmakers, and R. Kanaar. 2001. 'Chromosomal stability and the DNA double-stranded break connection', *Nat Rev Genet*, 2: 196-206.

Vermeulen, Wim, and Maria Fousteri. 2013. 'Mammalian transcription-coupled excision repair', *Cold Spring Harbor perspectives in biology*, 5: a012625-a25.

Volanakis, Adam, Monica Passoni, Ralph D. Hector, Sneha Shah, Cornelia Kilchert, Sander Granneman, and Lidia Vasiljeva. 2013. 'Spliceosome-mediated decay (SMD) regulates expression of nonintronic genes in budding yeast', *Genes Dev*, 27: 2025-38.

Wang, Z., V. Murigneux, and H. Le Hir. 2014. 'Transcriptome-wide modulation of splicing by the exon junction complex', *Genome Biol*, 15: 551.

Wang, Z., D. Wu, Y. Liu, X. Xia, W. Gong, Y. Qiu, J. Yang, Y. Zheng, J. Li, Y. F. Wang, Y. Xiang, Y. Hu, and X. Zhou. 2015. 'Drosophila Dicer-2 has an RNA interference-independent function that modulates Toll immune signaling', *Sci Adv*, 1: e1500228.

Wei, W., Z. Ba, M. Gao, Y. Wu, Y. Ma, S. Amiard, C. I. White, J. M. Rendtlew Danielsen, Y. G. Yang, and Y. Qi. 2012. 'A role for small RNAs in DNA double-strand break repair', *Cell*, 149: 101-12.

Wickramasinghe, V. O., and A. R. Venkitaraman. 2016. 'RNA Processing and Genome Stability: Cause and Consequence', *Mol Cell*, 61: 496-505.

Will, C. L., and R. Luhrmann. 2011. 'Spliceosome structure and function', *Cold Spring Harb Perspect Biol*, 3.

Wilson, Ross C., and Jennifer A. Doudna. 2013. 'Molecular mechanisms of RNA interference', *Annual review of biophysics*, 42: 217-39.

Wright, W. D., S. S. Shah, and W. D. Heyer. 2018. 'Homologous recombination and the repair of DNA double-strand breaks', *J Biol Chem*, 293: 10524-35.

Wyman, C., and R. Kanaar. 2006. 'DNA double-strand break repair: all's well that ends well', *Annu Rev Genet*, 40: 363-83.

Xiong, X. P., K. Kurthkoti, K. Y. Chang, G. Lichinchi, N. De, A. Schneemann, I. J. MacRae, T. M. Rana, N. Perrimon, and R. Zhou. 2013. 'Core small nuclear ribonucleoprotein particle splicing factor SmD1 modulates RNA interference in Drosophila', *Proc Natl Acad Sci U S A*, 110: 16520-5.

Xiong, X. P., G. Vogler, K. Kurthkoti, A. Samsonova, and R. Zhou. 2015. 'SmD1 Modulates the miRNA Pathway Independently of Its Pre-mRNA Splicing Function', *PLoS Genet*, 11: e1005475.

Yang, J. S., P. Smibert, J. O. Westholm, D. Jee, T. Maurin, and E. C. Lai. 2014. 'Intertwined pathways for Argonaute-mediated microRNA biogenesis in Drosophila', *Nucleic Acids Res*, 42: 1987-2002.

Yang, Q., Q. A. Ye, and Y. Liu. 2015. 'Mechanism of siRNA production from repetitive DNA', *Genes Dev*, 29: 526-37.

Yano, Tamaki, Sonia López de Quinto, Yasuhisa Matsui, Anna Shevchenko, Andrej Shevchenko, and Anne Ephrussi. 'Hrp48, a *Drosophila* hnRNPA/B Homolog, Binds and Regulates Translation of *oskar* mRNA', *Developmental Cell*, 6: 637-48.

Yu, A. M., and M. McVey. 2010. 'Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions', *Nucleic Acids Res*, 38: 5706-17.

Yun, B., R. Farkas, K. Lee, and L. Rabinow. 1994. 'The Doa locus encodes a member of a new protein kinase family and is essential for eye and embryonic development in Drosophila melanogaster', *Genes Dev*, 8: 1160-73.

Yun, B., K. Lee, R. Farkas, C. Hitte, and L. Rabinow. 2000. 'The LAMMER protein kinase encoded by the Doa locus of Drosophila is required in both somatic and germline cells and is expressed as both nuclear and cytoplasmic isoforms throughout development', *Genetics*, 156: 749-61.

Zaratiegui, Mikel, Stephane E. Castel, Danielle V. Irvine, Anna Kloc, Jie Ren, Fei Li, Elisa de Castro, Laura Marín, An-Yun Chang, Derek Goto, W. Zacheus Cande, Francisco Antequera, Benoit Arcangioli, and Robert A. Martienssen. 2011. 'RNAi promotes heterochromatic silencing through replication-coupled release of RNA Pol II', *Nature*, 479: 135-38.

Zhang, Z., J. Wang, N. Schultz, F. Zhang, S. S. Parhad, S. Tu, T. Vreven, P. D. Zamore, Z. Weng, and W. E. Theurkauf. 2014. 'The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing', *Cell*, 157: 1353-63.

Zhang, Zhenyu, Qiuying Yang, Guangyan Sun, She Chen, Qun He, Shaojie Li, and Yi Liu. 2014. 'Histone H3K56 acetylation is required for quelling-induced small RNA production through its role in homologous recombination', *The Journal of biological chemistry*, 289: 9365-71.

Zhao, S., D. Chen, Q. Geng, and Z. Wang. 2013. 'The highly conserved LAMMER/CLK2 protein kinases prevent germ cell overproliferation in Drosophila', *Dev Biol*, 376: 163-70.

Zhao, Y., C. Cocco, S. Domenichini, M. L. Samson, and L. Rabinow. 2015. 'The IMD innate immunity pathway of Drosophila influences somatic sex determination via regulation of the Doa locus', *Dev Biol*, 407: 224-31.

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

Zhou, Z., and X. D. Fu. 2013. 'Regulation of splicing by SR proteins and SR protein-specific kinases', *Chromosoma*, 122: 191-207.

6 List of Figures

7 Acknowledgements

Am Anfang meiner Promotion standen ein spannendes Thema, die Vorfreude auf die weitere Zusammenarbeit mit Klaus und der "neuen" AG Förstemann an diesem Projekt, und natürlich…ein Einstandskuchen. Denn "Kuchen backen und wissenschaftliches Arbeiten haben Einiges gemeinsam"[1] und können sich durchaus gegenseitig positiv beeinflussen[2].

Wie beim Einstandskuchen, der für gut befunden wurde, basiert auch der Erfolg im und außerhalb des Labors auf den richtigen "Zutaten", ohne die diese Doktorarbeit nicht hätte gelingen können. Daher hier, mit einem großen Dankeschön an alle namentlich genannten und inhaltlich gemeinten "Zutaten", das Erfolgsrezept für die vorliegende Dissertation:

- ✓ **Klaus als Doktorvater:**

  Vielen Dank für die Möglichkeit, nach der Masterarbeit bei dir im Team/Labor an diesem Projekt weiterzuarbeiten. Danke für deine Unterstützung in fachlichen und sonstigen Fragen, für dein Interesse an meinen Ideen und Ergebnissen und die Bereitschaft, diese zu diskutieren. Herzlichen Dank für dein Vertrauen in meine Arbeit und mich (die Konferenz in Heidelberg gehört für mich zu den absoluten Höhepunkten meiner Promotionszeit). Ein besonderes Dankeschön auch für die Geduld bis zur Fertigstellung und deine Vorab-Korrektur der Arbeit.

- ✓ **3+x tolle Kollegen und motivierte Studenten in der AG Förstemann:**

  Romy, vielen Dank für die perfekte Laborорganisation, die Unterstützung/Vorbereitung einiger Versuche und Reagenzien sowie das bereitwillige Weitergeben deiner praktischen Erfahrung an uns Doktoranden. Danke auch für das "Anstupsen" von gemeinsamen Pausen und Feiern.

  Ines, ganz herzlichen Dank für deine Freundschaft, den persönlichen und fachlichen Austausch zu allen möglichen Problemen, schönen Dingen und allem dazwischen, die gemeinsame (Uhr-)Zeit im Labor und die vielfältigen Ideen und Unternehmungen außerhalb des Labors.

  Stefan, Dankeschön für die Zusammenarbeit im Labor, das gegenseitige Aushelfen und Weiterbringen bei Versuchen und Auswertungen, sowie natürlich für die hilfreichen Kletter- und Wander-Tipps und die gemeinsamen Ausflüge in die Berge.

  Volker, Tiana und Isabella, vielen Dank für die Weiterführung bzw. eure Beiträge zu den Methoden und Projekten. Danke auch an alle Mitglieder der AG Förstemann für die schöne gemeinsame Zeit und für alles, vom Kuchen essen bis zum Burger grillen und von der Foto Session bis zum Weihnachtsmarkt.

[1] KF, 2013
[2] Noch nicht statistisch belegt, aber durch zahlreiche erfolgreiche Einzelversuche der AG-Mitglieder sehr wahrscheinlich.

- ✓ **2 TAC-Mitglieder:**

  Dr. Gregor Witte und Dr. Julia Philippou-Massier, vielen Dank, dass ihr mich bei meiner Promotion mit Ideen und Rat begleitet und die TAC-Meetings zu einem konstruktiven und ermutigenden Austausch gemacht habt.

- ✓ **2 Experten-Teams für RNAi-Screening und Deep-sequencing:**

  Vielen Dank an die AG Boutros und insbesondere an Dr. Marco Breinig für die Unterstützung beim und zum genomweiten RNAi-screen, der den Startpunkt für dieses Promotionsprojekt und die daraus resultierende Publikation gebildet hat.

  Danke auch an Stefan Krebs und das LAFUGA-Team (AG Blum) für die Durchführung der Illumina-Sequenzierung und die Bereitstellung der Daten. Stefan, dir ein besonderes Dankeschön für dein offenes Ohr und deine Hilfe bei der Optimierung der Vorbereitungen und für die Spike-ins.

- ✓ **1 Zweitgutachter und 4 weitere Mitglieder der Prüfungskommission:**

  Vielen Dank an Prof. Roland Beckmann für die Übernahme des Zweitgutachtens und allen Mitgliedern der Prüfungskommission für die Bereitschaft, diese Aufgabe zu übernehmen und die investierte Zeit.

- ✓ **1-2 Handvoll guter Freunde und Kollegen, für die Aufgeben keine Option ist:**

  Ein ganz besonderer Dank gilt den Freunden, die mich durch die Hochs und Tiefs der Promotionszeit begleitet haben und auch lange nach der Laborzeit noch an mich geglaubt und mich ermutigt haben, es vollends durchzuziehen.

  Dankeschön auch den ehemaligen und jetzigen Kollegen, die die zahlreichen Versuche, die Arbeit vollends zu schreiben auf unterschiedliche Weise unterstützt haben.

  Danke besonders an Sonja, Ines, Torsten, Ann-Kathrin und Stephie, die diese Arbeit oder einzelne Teile und Versionen Korrekturgelesen haben.

- ✓ **Zu guter Letzt, meine Familie:**

  Ihr seid einfach die Besten wenn es darum geht, mich mit allem Möglichen abzulenken und auf andere Gedanken zu bringen, neue ungewöhnliche Beschreibungen für meine Forschung zu finden und mich auf die unterschiedlichste Art und mit den ausgefallensten Ideen zu motivieren und zu unterstützen. Danke für alles!

Appendix 1: DsRNA sequences of validation screens

Target gene	Validation I	DesignA (sense/antisense)	DesignB (sense/antisense)
blanks	BKN31903	gtgtggatagtcggttccaaa gatggcctcttatgccattca	gtggtggttaggatttcctgc gttgaagaaaactgctatgggc
RfC4	BKN22927	gctgcagcagcgagttgatac gcaagttctggccaagctcat	gctttcagtttcggttgctttg gggaattcagatatagccggg
Doa	BKN31879	gcctgcatcatccgttctttg gggatcaaaagggatcaatgg	gtttttgaaatgttggggctc gcaccccagctccagtatgac
Hrb27C	BKN29074	ggagcctcgtgatggatctg gagtacatgtcagtggcggtg	ggaccggacccggtaattc gaaaaacgaacagacgacgct
RfC3	BKN20762	gcgttcagataactgaggacgg gaggagtcctccaagctactgc	gtccaaacagttgagctgcct gtttcacttctccattgaaacca
l(1)10Bb	BKN46079	gtttaaacaaatcggaacgca gccatgtggacgcattattgg	ggcgaacacagagacagcgt gtagctaaactcccactcccg
CG8142	BKN20043	gctaagatgatgcagcagcca gctacagcgttggccaaatga	gcgacagctgcgagaagtttt gtttttgaaaacgggaaaatca
Gnf1	BKN22954	gcgcaagcatgtgctcataa ggtccgcggtgaagacttttt	ggctgggcaagaattccaagt gcggaccctcgtccaagtaat
CG7185	BKN28487	gctgactgtgcgatttctgct gacgtgaacccagcattcttt	gggctgcatgatggagttca gcaatggggtgtaccatcaga
CG4294	BKN23609	gggaggcggctgtgtcttag gtagatccagccagcatcaaa	gtggcagatttaagggctatca ggcggcaagactatcaccataa
CG11777	BKN28775	gctgatcgaggaatggtttcca ggccaaaggattggcatgtat	gggaaagaagacgaacatctgg gtatgaagcggcatacactgg
nbs	BKN21109	ggagcattttcgttctcctca gcatttccgaactggtggtg	gattgattgattcagtggcctg ggtcacttcgaactggaccga
SCAR	BKN22532	ggcagatcgatggtagcacct gtatcattggcattctggacg	gaaggaaattgaacgaaacgc gataccacgcttggtttcgac
Zasp66	BKN32973	ggggacaacaaaatcgcctac gacgtcttgttgatggtccct	gggtgccaccgagtactagaaa gatttcgatgccaattgattacc
CG42724	BKN27003	ggggcttgaacttgcgattta gaccacacttgctacccaacc	gatgattagcgacttggctgg ggaaatctgggtggaaaccaa
CG16903	BKN20586	gtgtctcgcttgtctatcaatca gaaagatcggtctggaaaccc	gcacgacagcagagcaggat ggctgatctccacctgggtta
CG2063	BKN22284	gctctgggaattggaactattgc gccaaacaggacgaagaaagc	gccagggtcacaacagtttca gcgtggtggaagcctttaagc
CG3689	BKN29319	gaggagccgctctttgagaag gaccgggcaatttgaagaac	ggtcgtcaggacggcgtaa gtttcatgtagctggacgagg
mre11	BKN21299	gcactcgcgtggaggagttag ggtcacatctagctggccctt	gggacgcgatcgaaaagtg gatatgcacatcatcgtcctcc
CG5872	BKN45284	gaacgcagacgactcgcttac gcttttcaatatcctcgggca	gtccacaatgaaaagctgcac gttggtgctcaaattgggact

Target gene	Validation I	DesignA (sense/antisense)	DesignB (sense/antisense)
RnpS1	BKN21156	gcactgaaaacgctactcgca ggtccttctctttggcgacct	gagttccggacaaaattcgac ggcgctgaagtggcgatac
CkIalpha	BKN27574	gatcaagccggataacttcct gcatgtaatctggctgctcctc	ggccgtggagttttgtgaatc gaaaataaaacagttgcgttttcg
pzg	BKN45389	gatcctcaagtcgacgcgtat ggccttcctggcccatatc	gccagttcaagtcaacggaca gcgggctcatcgcttagtaaa
Vha68-2	BKN28767	gggtatcatgggcagcatctt gcctggacacaggggaagag	gctgaggtactgcgtgacttcc gggggttacccaagcacttaac
Vha16-1	BKN29332	gcaatcaagccgtagagaccc gcaggggcttcattcacttg	gcgtcaggctcaatgcaaat gggaccgtttgtacatagggaga
Trxr-1	BKN45110	ggcgttctattgtgctgcgt gctctggcttgccatagatga	ggagaacggtggtattgggat gcacacgcttctccagcctat
CG17912	BKN29387	gatgagtgtaccgccgcctat gctgtacgctggaaatgttgg	ggtctgcacgctaaccaatca gggtcgccagttaatcgga
loqs	BKN20112	gcgctcatcgacaagctgat ggaatcgatgggagtctcctg	gattacatcaagctgctgggc gatgtaacatcgccactgcat
CG17168	BKN41284	ggccgtagtcaaagtccacga gaaccagaaagcaactctgcc	gctgcatattcgccaacttctc gtttttccgttttgccgtaga
Chro	BKN24506	gttcccgctacctggtaaaca gtacaaactcctcagggtgcc	gattgtttggattgaccacagtc gttcgtggaagagtttcaggc
CG2685	BKN46253	ggcagcaccttctcgttcag gatatcagcctgcgcgtgtat	gtatgtacgaaacaccacagcg gcgtccatcaggggaatca
pUf68	BKN27426	gtacacattgttagtcggcgg gaagcagacggagaacgagg	gctttctttgactttgttgtgctg gcgtgagaaatccccagaaaa
Bap55	BKN20365	gctgctccaatccccacaac gccgcaaaatatacacagacgg	gtaaaattggaattcgattgcg gtctttgacaatgcaatgctg
c12.1	BKN22470	ggcatcatcgtcatcggaatc gggcaccagcgatgagaac	gctcccaaaattggcaagc gcagctcctccagctcttgtc
Srp54	BKN22462	ggtgtgatccaggtgaccaac gggtactccggcagattgaa	ggatcacgccgctctaggtc gatgacttggcggaaacgat
CG7028	BKN50938	gtagctcctcgcgcttcttt ggcaacatggaggacctacga	ggctttcgctgcttgaactg gctacgactccagcaacaaca
pea	BKN21969	gactgacggtatgttgctccg gggagtacaccggcaggataa	gggaggcgggcaagatctatt gtcctagtcacacgtttcctgg
tex	BKN21381	ggccatatatcttaacgttgcca gtccgcaccatatcgtttagc	gacaaggagacttgtgcgtcc ggttcagcttcgaggtcaacg
snf	BKN27649	ggatgctacccaaccaaacga gggttttggtggcttgacctt	ggcttcaaggaggtgcgtct gcgactggagttgcgttcact
hoip	BKN28514	ggagattctgctccatttgcc ggactagcagtcgctcgatctc	gaattgttctgcaattgccct gagtccgcgattgagggtct
CG4266	BKN20202	gctccattccggagctgttta gccatttccaaagcgaagatg	gcatcggaaattgtcccctta gtccgatacctttggcgaata

Target gene	Validation I	DesignA (sense/antisense)	DesignB (sense/antisense)
Nipped-A	BKN51011	gttcgcgaatagaaggaacttg gtccacaaccgaatgcttctt	gggaggagccccaattagttt gatggaatcagacgatcaatgg
SF2	BKN29395	gattagcttggaaaacgtgcag ggagcaggaaacaggaaaagc	gcgcatcttcgaactcaacaa gatactgcgtcggccattt
cactin	BKN20603	gagaaactggaggcggagaat gatatagcgcgggctgtagtt	ggagattacggtatccccggt gtcgatcagtgtcctctaacgag
CG8920	BKN24173	gagcagaagaccagcaagtcc gtggtgccttttgttctcaa	gcatgtgctcaccacggag gcttgacgccatcgacagtct
CkIIalpha	BKN50398	gcgttgtcaaggacccagttt gaatctaccagtaattcgggacc	gtaagtgctccgcctaccacc gaatcgagttagccatgccac
rad50	BKN20510	gtcaattgaatcgctgtccatc gcgcttagcgagtcctgctt	ggattcacagccaggtgaaca gccagtgccaagcggataata
Mtr3	BKN45761	gctttacatagggatcctcccg ggagcacggcctgatcattac	gagctcctcaaccttggcact gaacatgggtgttctgaattgc
tll	BKN46350	gacccgccgaaattctgat gcttccaggactcctccagc	gccttcgacatctcggctaca ggtgtcatggtggttgcagtg
IntS6	BKN45482	gggggtcgcttcattggtt gcgacgttgagcaacaatcac	gatgtggagctcacaaagctg ggccaatgaagtcgccaaat
CG13800	BKN23634	ggcaaacggcagggaatact gggaggtgctctcctgcttta	gacgtctcttgaccgaggaaa ggccagcaggtctatgtcgag
E(Pc)	BKN20790	gatggaagaggatgaagccg gtcgagcgtagtcctgagctt	gacagcagcatcaaaagctga gattattgttgttgctcgtggg
Rrp6	BKN20686	gctctcgtcctctgaccgc gctgccaaagccttgaatatg	gggcgcttgtctcagttcct ggcagtagcttccccacaaac
DMAP1	BKN22326	gattgtgatggcagatcgttg ggaacttgatgccgctgctac	gacggaggacatttgcacct ggtctggggaatctgagtgga
BubR1	BKN50563	gaaaccgagccataagccc gcgcagctactgaaatgtttgc	gggaaaggactgctgatacacc gccgttcctgttccgaaaata
Dref	BKN24401	gaacttctttggtgcactgcc ggtggcgaagacattccctt	gatgaactccagtttgacccg gagaaattgcgcgctaatacac
CG3511	BKN28490	gggaggaatctgtgtcatccg gcagtgcttgtacggcatctc	gcatctacgagaacgtggtgc gatgtcgtcgtagggcttgtc
snRNP-U1-C	BKN21376	gattaattgttttcgctatcggg ggcctccactgggcactatac	gtagtggcgtcaatcaaatgc gacccttgtcgaaatacggtc
CG6066	BKN28878	ggcttctgtcgctcctcctt ggaggaagatgtttggctgga	gaagaacccttcaccagctcc gagctccacagggaaaggact
feo	BKN28553	gctccttctcctccttgagca gctgtgggacagactccagg	ggatcttttcgcgcatatcgt gggagctgaagctggacaaaa
CG2577	BKN50153	gaagtgcgaatcggcaactat gcagaagcagaacggtcttga	gccaagcctctggagattgac gagttgcatgataattgaaggagg
Pitslre	BKN28225	gcagcgaaaacagacggctc gacccacgctaagtggactgt	gaggatacaccgagctcccag gttattccagcgttcaggaca

Target gene	Validation I	DesignA (sense/antisense)	DesignB (sense/antisense)
CG9667	BKN29247	gggcgtatttctccagcagtt gatgaaatatatccccctttcgggg	gggccacccaatgaggata ggccgcaaaagaagtggaat
Hsp60B	BKN30688	gacctctagtcgctgtttcgc gagagggtggctacttgttgg	gcaagaaggacaagttcccca gatgggaatcctcgcatctc
CG16838	BKN21251	ggcagcactcatggaggacc gcttgcgttcgttccattctt	ggcacttgcttgatgtgatgg gtggagacgatcaccaatgtc
CG14641	BKN27491	gtgatgttttgctcacatagcg gggagtacggactgccgatt	gacagagtacaccaccggtca gtcccagggaaatgaaaagaa
CG9797	BKN26226	gcgccggaatgtatcagcata ggagagcgaccgtttgcat	ggttgctccgctcgaataact ggttgacaagcgacgcgttta
CG33109	BKN24267	gataattggtgcgaaaggttcc ggcaaccagttggaggacatt	gattcagcaaattctgtggagc ggcaaaggtgtcgaaagttcag
pbl	BKN22310	gcactccggtttcgtttgatt gctggagcatctgcggaaa	gtatggtgtccaatatgccca gagcatcacgaatacaccgaa
crn	BKN20508	gcaatcaagcgcaaatagtcg gctgcgctacaaggagatcg	gggttgccactccatccag gagatctggagattctgccgc
pav	BKN20803	gcaatcgacacatgttcgtcc gataggggaccttcttgggtg	ggattcggactcggattcagg gatgaacaccaaccacaagcg
CG15432	BKN28733	gaagaaaaacgtaaacagatcgga gttgcccttgtgcttcttctt	gcctcctgaacactttcacctg gcgggcaacatagaatcagc
Rack1	BKN29073	gcttggagtccttgccacct ggaccatcaagctgtggaaca	gcaaatcccacaggcgaag gaactcaagatgtccgagaccc
Snp	BKN45478	gttaacggggagatcctagcc gtggtactaaggtttcgggca	gggcatccagcagaagactgt gtaagctgctgatatggcgtc
CG15525	BKN45014	ggtgctagaatctcgccatcg ggtcggcaagctgacacaag	gcatgagcttcatgtccgaat gcgctaaaggatatgaactggg
CG10754	BKN27781	gtttctgtggtaatcggcttg gctaagatcctgccgccac	gcagatcgattgtctccaggg gtcatgaaattttaccaacgca
CG30089	BKN31714	ggcaacagcagcaacaacagt ggcagtggcaccgtcgata	
ct	BKN45881	ggcatgcttctgacctcgc gttcagcaactgcaatcgttc	gcgatggacggggtcattat gttcttgggtatcagcgatttg
CG8237	BKN45253	gggcggagaccatcgatatg gggcatcactgcttccacat	ggatgagggctctgctttatcc gcacacgacagcagtacataacg
fand	BKN21802	gagcaggtagctgccaatgag gctgcaccgtttccgtatcat	ggtgatccccgcatcactg ggtcctcctcctcatcgtcct
Bx42	BKN21649	gcccgctgcaggtttctatc gattccactggacaagcgtct	gccttttgctccttgaccgt gaaatgttggcaagaagtcgg
kz	BKN26359	gcacgtatgttgcagagcagg gtaggagctgcagagtacgcc	ggtaagcctccgccacatagt ggtatgcccagcacaagatga
ial	BKN32558	gctggccgtgttctccttg gagccggagaacattctgcta	ggccaagtagacacgtccaaa gaaaacgcgcaaacaaaagg

Target gene	Validation I	DesignA (sense/antisense)	DesignB (sense/antisense)
Prp19	BKN21386	ggcagcgaaaggccagtaac gcgccgaggtcatacagaaac	gcggttgtggtattgccgt ggaaaagcgggtgattgagaa
Mo25	BKN45886	gtcatcacggtgaaattgtgc gatcgatttcgagggcaaga	gtttttcggcgttacgctatt gaagcccattctggacatcct
CG4980	BKN29267	gtaacttgcttcatgctttcgc ggcagtagctccagcaatcca	gagtcgagcattagctccagg ggcgaagttattgtggatgctc
Rtf1	BKN21260	gctcgccaagaagcgaaag gccatgaggtcatcatcgaac	gcctgatgaaggagcgagatg gggtcaatgtcaaagtcatgcag
CG15309	BKN45574	gttggcagtgcaagttctgtt gaggaaaagcagcgagttcac	ggctcaccgagcactttgac gttattttggcctcccctctc
CG6443	BKN21026	ggcagaggaacaggagcca gcctggtccttttcggtttta	gaaagatgcggagagggagtt gctgtctctggcatcggttc
snRNP-U1-70K	BKN45552	gcattgtcgggattgggatt gagcgcagaaaacgacgag	ggtccgcgatctcgttcac gagttctacgggcccatcaag
CG13298	BKN22482	gcgcttgttcatgtccggt ggccaccctgttcaattttgt	ggtcggagggttcaccagac gcccatcatggcatccttgta
CG15747	BKN22378	ggacgaggtggcatacaaacc gctcttgcgccagaagagtgt	gcctcgaactttaggcactcg gttgccatatgagaacgagga
CG9752	BKN20542	gaaccgtctttatttgtgttgcc gcgatcagataaccatgttcgc	gcaaaagcaaaagttcagggtg gcctgttacagcgccaaaagt
Brf	BKN23074	ggcaggtcatgtacacgcaag gggctcggtattggaggattc	gccaccacaactgctggct gaggccattgaaaagatgctg
l(2)k01209	BKN51139	gtctggccaactcctacaagg gaccgatctcagccatcaaaa	gatccgtatgccgacctaagc gcctgctcatgctgtttctca
slik	BKN20098	ggccgctgcttgaactcat ggacgatgagaacggacgaat	gtcgatggaaaaggcctcata ggccaacatgtcgtttatcacc
nonA	BKN28107	gggcaagcataatgcacaaaa gggtttccagcgcgattatt	ggcgatgaacagcaccaatatg gacgaacggtgtgagattgct
RfC38	BKN22321	gatgaaaaatgtgcttggcg gtcgtgtccatcttgcagaac	gcggaatgatccgagatgtg gcaagcagagcgattttcctc
Trf4-1	BKN22316	gtcgaatcattcacatcaccg gtccttgctatcctcggagtc	gcagcagcaccaagagacttc gtgctgctgttgctgcaatta
CG14230	BKN45873	ggaacagaagcgcaaggagtc gttggcgacttaccgtcctta	gtgcaacttctgttgctgctg gactgagaaaaatcatgccga
U2af38	BKN28210	ggtctgcgagaaagtgggttt gcgtgcaaaatacgaccagac	gggggtttgtacataggtgg gatgcacttgaagcccatctc
ZC3H3	BKN22095	gagcttttcttcgccacctaa gcaagggcaagtgcgaattac	gcacttgccacgactatggg ggcgttaacaaatccctgagc
RhoL	BKN31715	gggtcacgaacttctccgagt gtgcttcctgttgtgctattcg	gcagtgtcccagagggtcaga gttaacaaagacgaacgggga
CG6613	BKN20192	gggtcctgggagtatgtgagc gagtggttcaacaggctggag	gtgcaatagctgcatggatgt gggtgctgatgtgaaagggat

Target gene	Validation I	DesignA (sense/antisense)	DesignB (sense/antisense)
Prp18	BKN27465	gcgatccttcttgctgacgta gacctgctcaaccgcaactat	gctcgacggactcctgcttta gaggcatttgttttggttactcc
faf	BKN20871	gcgttatattgggcaagacgc ggaaaccccgatcatctcgt	gagttccaaaagggcaaggac gctgctcgtacagcaactccc
Saf-B	BKN20574	gacccacttgattgtgccg gagcagcttctgctcgtcct	gatcctcgtcgagcaccaa gcgcgtctatagctcgatgtg
CG2807	BKN21741	gccgagactccaaagccagat ggagtgcgcagtggaacatag	ggctatgtgtcgccctttttc gcatggtgctgtccttgctaa
CG2021	BKN27671	gatcctagtgcaccacgggac gcttctctaccacctccggc	gcaaggacaaggagaagcagaa gatgtgtgtgtgtgggtgtgtt
Sin3A	BKN20796	ggcggtagaagctttggattg gtcgactcgaaatgccaaga	gtatggcgtgattgaactcca gtaaatctaatgacccacggc
CG9186	BKN28771	gccacacaaaggcagtggagt ggacgatgagatggccagg	gtggagtactttagggcggtg gagaggccagtattcgggag
phr	BKN30452	gatgaagcgtacgaaggcg gattccagcttgagagccaga	ggacatggaagtggacgaggt gaatggcctcctcgcaaaac
ssx	BKN29594	gcccagacaaacaggagctatc gcagccgctaaagaggttgta	gcggatatagcttcggctacg gcggtcgtccagtgagcttat
CG31368	BKN20937	ggctaaagttttatacccgcttcg gacgttctcgtcccagatgat	gggacacgaaagcagatcacc gaatgccacatggacaaggag
ubl	BKN42070	ggtttgtgccgcgattagttt gaagggtatttgacaaccctaagc	gagacacaggaacggtattccc gggcacaaagcacgagaagat
Wnk	BKN50794	gtttggatgtacgctctgctg gcagcaacgttgtcagtggtt	gctcggtttctgggaacgtgt gttggtgcaacagcaacaact
Nlp	BKN30946	ggtgaacacacccaaggactc gaaacctcgatggatgattgg	gagttccttttgaacagctccttt gaggaagtccggcttagcttt
srpk79D	BKN23592	gactgaggacattcagactcgc ggcttaccgcggaagatcac	gccatcgagaagcagaccaag gtgattttcttctccgttcggt
Tango4	BKN29170	gcagcatacccaagcccaagt gggtgatgatctgcggattg	gggtgatccccacgaactg ggtggtaccgggatcattgtaa
DNApol-delta	BKN29129	gggactccagtggcacacc gagcagctcacttgaaggtca	gcacaacacctggctgcctat gggcttcacgtctcctcaaaa
CG10492	BKN27409	ggacatggtgaactgccagaa gctgctgcggtcctggtat	gttctgcccagctctgatg ggcctgcagtggaacgttaat
Prp8	BKN20780	gtgcgagtcataatcgccc gcgtcgattcacactttggtg	gcgtcgtatacgttcccgatt gaagctgatgaggcagattcg
kay	BKN32849	ggatttgcagtgccaaaggtg ggtgagctcgttggtctggtc	gcagcaagcgcatcaccttt ggaggagctgttcggtacgag
mad2	BKN28038	gtgatcttgggatccttggac gggctccgctcagattattgt	gattgttaggaccaccaaccg gagttcctccacgcaagttct
Sam-S	BKN46638	gcgaaaccgaggaatgtatgc gagagaccggccttaaccagt	ggtactgctggcggttctctt gcatccgttgtcatacgcctt

Target gene	Validation I	DesignA (sense/antisense)	DesignB (sense/antisense)
l(1)G0007	BKN22970	gggacgccctgcagatctat gcggtagaagatcgatggcac	gggcagcagcttaaggacatc ggcttgcgctcctccatct
Rim	BKN45626	ggaccagtcagctccaccaac gagcagcagctactatcccca	ggatgttgttgcggcataaga ggataagcggcctaagctgg
CG14220	BKN22791	gttgtagacctcgttcatgcg gcgagttctgaacccaccaac	gaacaaaaagatgctgcgctt gctgctcttgcttcggcttta
Incenp	BKN26980	gggaggtctccaagaaggtcc ggctcatgtgtacgggtaggc	ggtgccgtcactaaataaaagca gaaagtgtacacgagtcgcgg
egh	BKN27178	gcgtggatccttggactgaat gtaatgataccacgcaccgaa	gccaaaaaccaaattaagcaagg gaatggagagttttcgcgtgt
CG1620	BKN50407	ggacgttgttgtcctcgtcct gtctgtggaagaagagtgagcg	gacaaccgcagtggacacac gccgggaattgtacgactactgt
Taf10b	BKN22306	gcaatgcatgccattatctactga gcccatatgcaaaccaccaat	gccttggagtgctgcagtg ggggataggaccacaccatctt
CG8677	BKN22027	gggcaaatcaacgtgagctatc gcggatgcggactgtgaagta	gcaacgcaaaggagagaccat gctgaccaagaggagcggata
Baldspot	BKN29085	gtgccttcccctaacagaatc gactcctgccttcagtttgga	ggatgaacttggagctctcgc ggtatcaccacatcaccgtgc
Skp2	BKN28837	ggagtcaaacaaaacgcggc gtgacctgagtatggcttcca	gcgagtttgaattgccgctac gagaaccacgtctctgttgtgg
Tip60	BKN27642	gagtatggtttgcgcccagta gccgatccgtttctcttctacg	ggggtggagccacagagttt ggcttggacgagtgggtca
Psi	BKN21933	gcgaccaccctgttgctgac gcaggtcatgcaggatcaaga	gtcgtttttgtgcgatttctg gatctggtggccagagcagta
CG4400	BKN28153	gcgctagctcctgccattg ggagaaacacatggcattgga	ggcagcttgttcctcggact gagcgccaatttaacgagttg
Bin1	BKN21416	gtgattactccggaagttggg ggtcctatgctgctccgtgtc	gctcgcggtcaatctgctt gcgaaaagagcaaagagaagcc
His2Av	BKN28652	ggcaacgcatcgaaggact gcctgcgacagaatgacgtt	gcctactaagccagtcggcaa gtttgtttcatactcaacctcgc
sip1	BKN28960	gctacaagtccgtggaggagg ggagcagctcccgaaagagt	gatctggcagccggtgtaata ggcaatgctatacccaatttgc
SRPK	BKN50500	gcatgtgcgctcactagcaac gtctttttggctttgtcagtgc	gcgtccacgtaaagatcgcc gcgtaaggaacgaagcgaatg
Upf3	BKN45942	gtgaactcctggtagtggggt gctgaaacttttgccgaggaa	ggccgtcatccttttttcttgt gcatgggcgacgtcaagat
mus209	BKN28382	gaggaatgaagtgaacgtggg ggtcgatgtgcgcagatgtt	gagattcatgcccatggagag gaaagtcacagccgcggta
Pcf11	BKN20818	gcttttgttccacgacggg ggcgttacctttggctagcat	gaggtctgctgtctggcgt gggctccaaggagatacgctt
Cbp20	BKN29448	gcttggacgaccgtctgattc gaactagcgattgtccgtgttg	gggcagatgctgatgaacctt gaaactgggcagatcttcgg

Target gene	Validation I	DesignA (sense/antisense)	DesignB (sense/antisense)
CSN5	BKN20306	gcacaaaagacctgggagctg gtcctctaccttgccaagcat	gtaatattgggtgcatttccct gccacgcttcaatttgtttgt
Rpb5	BKN20359	gtcgataggatatgtaccgtccc gaacatcacggagcacgaact	gaagagattcatgctggcgat gagtctacgtgggcagcattt
Ndc80	BKN23993	gcggagaacttctgcctgtta gcttcgaagaggctagcaaatg	gtactgtagcttttgcattgcg ggatcgggccagtcaaatg
E2f	BKN20400	ggctactgccgttgttgctct gatgtggacgtggagctgaa	ggttgttgttgatgctggtgg gagagagcggagaagagagagg
CSN1b	BKN21757	gctcccgggaaatgctctt ggccccatgtaaccactctgt	gaggtcgccacaactgaggta ggacgtctgcccggttttag
MED4	BKN28661	gctactgcgaatcgctggaac ggtagcgcaggaagtgggagt	ggctgttctggtgggtcact gaccgtgagatccagaagctg
Jon25Bi	BKN31014	gccagtccagctggttggtaa gcgtcgacctgcagatcatta	gacggcccaatagttgtcgta gccctacaccgtgggtctg
mRNA-cap	BKN60320	gaggttcctgctctccacaaa gtggtgcctggactatgatga	gaggcggtagtgcaactgg ggacacagcaccgagaagaact
Cnot4	BKN22063	gtcaagccaatgcaatcattc gcgacaaaaggcgaattcagt	gcgttcgttgagacccaaaag ggcattccagttgacacccat
tho2	BKN22082	gctctcctttagcatgtgggg gcatgctcgagaccgtgatg	gctggcgcacaatggctatac gacctgccatgcactacgac
Cdk8	BKN45143	gctgttgctgctgtcgcttat gaacgctccacatactccacc	gatgatgttccggcatcttct gcatcgtgacctgaagccg
Rcd5	BKN21572	gtacttttcctgctggtgcg gttcgcttttgatcgttttgc	gccacctcccagaattatcca gcaggctcatacagcagagca
puc	BKN46004	gtatagaaacacaccccgcct gcctcctggaagtactgcttga	ggaagatgcacggaaaacgg gcgctgtccacatcatcgtaa
Cdk12	BKN21793	ggaagactcaagcgacccagt gggttgctgatcgactgctta	gactcacgaagtcggcaatct gattgctgttctcctcttgcc
Top2	BKN21781	gcgtcttcttgcgcagacttt gcaggctgatcgcctcagt	gcttcaagccatccaccagac ggctgtcattctactcgctaccc
SC35	BKN32717	ggagtggctaccaggcgag gtatgacaaacgtgatgccga	ggctctcacgtgtgtagcgat ggttgttccagctctaacgtcg
Uba2	BKN22286	gcgtcagacgactccaaaatct ggtcgtagcagatgcctcaca	gatgtgggatcgcacgttg gggaaggaaaaaggcgaagag
brat	BKN33206	gaacagcagcagcgacaact gctgttaaccttggccgtgta	gcggtaaatgcccagaatgac ggcagatactggccctcgtag
JIL-1	BKN50657	ggggtcaaatcgttcgtgc gggggaacgtttaatcagcg	gttaatgccgttgaagaaggg gtgtacatagccgaggtggtg
Bre1	BKN20850	gtgcctgtggtagttggtgtc gaggttaatgcaacgcaatga	ggctgccgcgtttcatatcta gcgcgtcgattgagaaagagt
Droj2	BKN27834	gtacttggcgtgaagcctaatg gcgctccattgtacagctcct	gcatggcctacgatgaggac gagatacggttctagtggcgg

Target gene	Validation I	DesignA (sense/antisense)	DesignB (sense/antisense)
CG42550	BKN33975	gtcttctggaggtagtggtgg gctgcagcagttgatcgagtt	gttgtttctctttcagttggacg gtgcttgtccttgctgtagga
smt3	BKN21837	gctgctgctggtaaacctcg gaaagaagggaggtgagaccg	gtgcttcttgatcttgaactgga gacgctccgcaaaagaaaaa
endos	BKN28642	gccccttctgcagcctttt gggaagaaaacagcaacagcc	gggctggctgctccaactta gggcatttcgcgacagttta
mask	BKN20625	gagagcaggataccaacagcg gctctccatgctgctgcttct	gagaagaacaagaagcaggcg gattcaaagtggtggcactcg
CstF-64	BKN27450	ggacaagtttgtcaacgtttgct gccactttcccgatcaaagac	ggtcccatggatcccaggtta gtatgcatgagcgaaatggtg
nito	BKN28409	gaacatgaagcgtagcgcc gatattccggtctggttgtgg	gcaccctgagatccggaatg gagctggattttgcgtttcag
mkg-p	BKN40366	ggtaaatccagcgtggtgagc gaaatgctttgcggatgtctt	gaacccgtaaccagtgaacca gctgcctcggaatcaacttct
Trf2	BKN27538	gtaccaggaatcaaatcacgg gaatccgatttttgatttggc	gaatcctgcttgtagatgccc gatcgctggctatactgtccg
Sym	BKN21288	ggccttaaatcggggcact gccatgacccccactgatatt	gcgcctgtacggaggatctac ggtgaattcttcggcggtagt
wds	BKN20216	gagagagagcagccggtgac gattcggactgaacttgaccg	gagtggcaaatagctcacggt gcagaacgtgctcagccaaac
Sse	BKN20595	gcttttccacggccaaattc gcagtacgtcaatggtcgcat	ggagttgaccaaatccgcatc ggggtacccaccaacgtgttt
CG7200	BKN28118	gtctaaacgagcgagcaatcc ggggccaatgaagctgtttt	gaatggtgtaatatcgcacgttg gccaaatgccgggctataact
nonC	BKN50537	ggcaggaagttctggaacagg ggcactcacacgactcacatacg	gtagtttcaatacggcctcgc gccacgacatctgccttacaa
Caf1-180	BKN27324	gtcctcctcttcctcctcgtc ggaggtggactccaagaacga	gttcagtttcgtgtggcctg gatacatcaaacaaagcacgca
CG3857	BKN30209	gcatgttcctagacaggcccc gcgaccgacatcaagtcttttt	ggtcggcccgtataagtaggc gcggtccctggactacgacta
gwl	BKN51041	gtttggaaaagtcttcctggg gtcgcatatcgattttgctca	gatccgtggtgtgctaaaggt ggaccgtgaccttgcttcag
CG31709	BKN25255	gcgcaccattactgacaatgc ggttaacttgctatttttgcgcc	ggcattcccttgatccttctg gtagagccaagtggaaatcagc
Pp1-87B	BKN31383	ggtcggactgcattgtggac gtatctaccgaccgaccaacc	gacttttacgcttgtcggct gatcttctgctgccacggt
Cpsf73	BKN20534	gtccagatttatttccgcacg gatgtctgttgtaacccaacgc	gagtagcagttcttgggcacg gatgccacgaaagccatttac
CG14418	BKN31933	ggctgctgctgtcgattgat gcaatggcaacaacagtttgc	gaggcaggtctgacacacgtt gttttcgacaagtagaaagcgaa
Cenp-C	BKN23772	ggatagtgaccatgggcaagc gccctcgattgcttagagctac	gcacaccatggatcctagcaa gcttcctgctcgtctctcagg

Target gene	Validation I	DesignA (sense/antisense)	DesignB (sense/antisense)
CstF-50	BKN27473	gcttcggtaccagaaacgtgc gatcttccggtatggattcactg	gtctacgactgggacatgctg gtttctatggatttttcacggc
CG5514	BKN23035	ggtcctggacaagggagtgg gtcaccgaatccgagttattca	gaaacagaattggaggcaagg gctgtgttccatcctcacttaacc
CG1109	BKN20307	gcaagcggaaaaagaggcat gagtgccttgatgatggacg	ggatcgcgaacgagaagga gggcagaaaacagtgttggtg
Spc105R	BKN26153	gccaataaccgagatcgttcg gtcgatacttctgcttctcatcc	ggtttcggcggagcctatc gggtcttgcacatctcttcctg
Mes-4	BKN20996	ggcgcaactggatttgcttta gaacgcaagcggaccaagtat	gaaagtccttttccacaccca ggcaggaacaggacatggaac
Upf2	BKN20728	gtttggtgcttgaacttgtgc gagactcgtaccaagaacggc	gattccgcgtaaactccagtg gcaccacccaatgcttcgt
Fas1	BKN29572	ggcgtacacccagaaacgact gcggcatcatagaccgactct	ggtgcccactgtcaacaacaa gcgaaatagccaaatgacgct
stwl	BKN45199	gatcttctgacgggggttcat gagtcgcccataaacaccaaa	ggcttaaggttgtccagcgaa ggtagcagcaatccacccaat
Pp1alpha-96A	BKN28922	gaatcgacgtcactaaccaacg gcgattgcagcaactgctattt	gttccagatattgaaacccgc gtttcatttcattcgtttcgtttt
Su(var)2-10	BKN28790	gtgtcattcctacaaaccgacc gtgaccaggttgtcataaatgg	gtacttccaggaggtgttggg gcactgtttgacgttgatgtgg
Rm62	BKN32269	ggccaccagaatgttggactt ggccccagatcagaaagatcg	gcacggtgatctcctgctctt ggatcgcgactttggtcaca
Clp	BKN27764	ggtgggctgcagttcaagg gggacactctttgttgtggca	gaggacgctctgctctgtatga gcgttctactggcagttctaggg
Upf1	BKN27828	gactccgactccagtgactcc gccaaatggtgggacagtca	gagaagtacaaggcatgacgga gcggatttgttcgggacacta
Cpsf100	BKN40434	ggacacggctttcgaaaagat gattctatctgcgacttggcg	gggtagagaagacgcgctca gggaaagactgctaaaatgtccg
Cpsf160	BKN23031	ggtcatcggttcctccgttt gtatcccaagggtcacctgaa	gaagtatcggagcgcacctt gatacgtttgcgctgaaaactc
CG12640	BKN30871	gacctgtgcggatcatgaaa gcagcctgaatccgagtcc	gttcaggcgtcttcagattgt gtattgtacgtgcggaacctg
bel	BKN28143	gcgcttgtcgggttcatagac gagcagtatccgttgggtctg	gccagctcgtacatctggttga gcggcaacagaagctacaaca
tsr	BKN28706	ggtatcgggacaccacgacat gagaaggacaaaaagcatcgc	gaggcttcggaaagatcagtg gcctctgagagttcaaagaagca
Smg5	BKN22183	ggttcccgtggcactggttat gggagaccattatccgggag	ggatcagctgcatcaatgtcg gcggagtacgaggaacgctt
CG2186	BKN31161	gacccaaccagtcaccacaac ggccattggcatagttcgg	ggcatgtcgcgagtccatt gacgcgtacgatattctcggt

Appendix 2: Detailed results from validation screens

Data published in (Merk et al. 2017). Acceptance criteria and Ago2/Dcr-2 values are listed at end of table. Values outside the validation range are marked in red.

Gene information							dsRNA screening constructs			new dsRNA designs A and B								
							genome-wide screen DSB reporter	validation DSB reporter	validation high copy reporter	DSB reporter			miR-277 reporter			high-copy reporter		
CG-Nr.	Symbol	Name	GeneID	BKNID	Hit	validated	zscore_ loess	average fold change (n=2)	average fold change (n=2)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)
CG10630	Blanks	blanks	FBgn0035608	BKN31903	pos	yes	**38,14**	**3,73**	**1,64**	3,36	2,89	**3,12**	1,11	1,14	**1,13**	1,59	1,41	**1,50**
CG14999	RfC4	Replication factor C subunit 4	FBgn0260985	BKN22927	pos	yes	**25,95**	**4,97**	**0,74**	2,54	1,81	**2,18**	1,15	1,12	**1,13**	0,98	0,77	**0,87**
CG42320	Doa	Darkener of apricot	FBgn0259220	BKN31879	pos	yes	**25,09**	**2,16**	**1,45**	1,35	1,33	**1,34**	1,25	1,05	**1,15**	1,40	1,69	**1,55**
CG10377	Hrb27C	Heterogeneous nuclear ribonucleo-protein at 27C	FBgn0004838	BKN29074	pos	yes	**25,07**	**4,21**	**1,68**	2,58	1,76	**2,17**	1,50	1,56	**1,53**	1,41	1,44	**1,42**
CG5313	RfC3	Replication factor C subunit 3	FBgn0032244	BKN20762	pos	yes	**24,43**	**2,20**	**1,24**	2,35	2,92	**2,64**	1,08	1,05	**1,07**	0,81	0,89	**0,85**
CG1639	l(1)10Bb	lethal (1) 10Bb	FBgn0001491	BKN46079	pos	yes	**23,26**	**4,10**	**2,08**	4,00	3,79	**3,89**	1,13	1,29	**1,21**	2,36	2,23	**2,29**
CG8142	CG8142	NA	FBgn0030871	BKN20043	pos	yes	**20,11**	**4,58**	**0,74**	2,84	1,79	**2,32**	1,14	1,07	**1,10**	0,91	0,98	**0,94**
CG1119	Gnf1	Germ line trans-cription factor 1	FBgn0004913	BKN22954	pos	yes	**17,94**	**0,79**	**1,16**	2,86	3,28	**3,07**	1,51	1,39	**1,45**	0,81	1,03	**0,92**
CG7185	CG7185	NA	FBgn0035872	BKN28487	pos	yes	**17,63**	**2,32**	**1,45**	1,91	2,46	**2,18**	1,43	1,31	**1,37**	1,32	1,33	**1,32**
CG4294	CG4294	NA	FBgn0034742	BKN23609	pos	yes	**17,28**	**2,50**	**1,12**	2,09	2,07	**2,08**	0,59	1,08	**0,83**	1,43	1,37	**1,40**
CG11777	CG11777	NA	FBgn0033527	BKN28775	pos	no	**17,10**	**1,12**	**0,81**	1,02	1,03	**1,03**	1,05	1,03	**1,04**	0,95	1,14	**1,05**
CG6754	nbs	nbs	FBgn0261530	BKN21109	pos	yes	**15,17**	**1,06**	**0,82**	1,58	2,12	**1,85**	1,08	1,04	**1,06**	0,88	0,94	**0,91**
CG4636	SCAR	SCAR	FBgn0041781	BKN22532	pos	no	**14,69**	**5,37**	**1,27**	1,10	0,76	**0,93**	0,64	1,11	**0,87**	0,93	0,84	**0,89**
CG6416	Zasp66	Z band alternatively spliced PDZ-motif protein 66	FBgn0035917	BKN32973	pos	no	**12,27**	**1,03**	**1,02**	0,95	1,09	**1,02**	1,14	1,06	**1,10**	0,74	0,85	**0,79**
CG42724	CG42724	NA	FBgn0261641	BKN27003	pos	yes	**11,57**	**1,20**	**1,30**	1,92	1,39	**1,66**	1,21	1,05	**1,13**	1,19	1,33	**1,26**
CG16903	CG16903	NA	FBgn0040394	BKN20586	pos	yes	**11,53**	**3,95**	**1,64**	1,50	1,78	**1,64**	1,05	1,19	**1,12**	2,07	1,14	**1,61**
CG2063	CG2063	NA	FBgn0033400	BKN22284	pos	yes	**11,34**	**3,61**	**2,74**	1,40	1,64	**1,52**	1,08	1,08	**1,08**	2,15	2,03	**2,09**
CG3689	CG3689	NA	FBgn0035987	BKN29319	pos	yes	**11,32**	**1,43**	**0,98**	1,56	1,54	**1,55**	1,23	1,13	**1,18**	1,30	1,20	**1,25**

Gene information							dsRNA screening constructs			new dsRNA designs A and B								
							genome-wide screen DSB reporter	validation DSB reporter	validation high copy reporter	DSB reporter			miR-277 reporter			high-copy reporter		
CG-Nr.	Symbol	Name	GeneID	BKNID	Hit	validated	zscore_ loess	average fold change (n=2)	average fold change (n=2)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)
CG16928	mre11	meiotic recombination 11	FBgn0020270	BKN21299	pos	yes	**10,69**	**1,19**	**0,78**	1,80	1,08	**1,44**	1,00	1,15	**1,08**	0,92	0,78	**0,85**
CG5872	CG5872	NA	FBgn0036991	BKN45284	pos	yes	**10,47**	**1,62**	**0,97**	1,83	3,56	**2,69**	1,10	1,06	**1,08**	0,89	0,87	**0,88**
CG16788	RnpS1	RNA-binding protein S1	FBgn0037707	BKN21156	pos	yes	**10,12**	**2,84**	**1,17**	2,02	1,25	**1,64**	1,01	1,14	**1,07**	1,35	1,26	**1,31**
CG2028	Cklalpha	Casein kinase lalpha	FBgn0015024	BKN27574	pos	yes	**9,99**	**3,85**	**1,34**	3,25	2,23	**2,74**	1,23	1,30	**1,26**	1,17	1,20	**1,18**
CG7752	pzg	putzig	FBgn0259785	BKN45389	pos	yes	**9,82**	**4,91**	**0,84**	2,69	2,54	**2,61**	1,13	1,12	**1,13**	0,83	0,80	**0,81**
CG3762	Vha68-2	V-ATPase 69 kDa subunit 2	FBgn0263598	BKN28767	pos	yes	**9,74**	**1,98**	**1,06**	1,99	1,55	**1,77**	1,08	1,12	**1,10**	0,93	1,04	**0,99**
CG3161	Vha16-1	Vacuolar H[+] ATPase subunit 16-1	FBgn0262736	BKN29332	pos	yes	**9,08**	**2,22**	**1,00**	2,22	1,53	**1,88**	1,06	1,06	**1,06**	0,87	0,97	**0,92**
CG2151	Trxr-1	Thioredoxin reductase-1	FBgn0020653	BKN45110	pos	no	**9,07**	**1,67**	**1,33**	1,11	0,69	**0,90**	1,01	1,11	**1,06**	0,91	0,70	**0,80**
CG17912	CG17912	NA	FBgn0032600	BKN29387	pos	yes	**8,72**	**1,97**	**1,63**	1,66	1,74	**1,70**	1,10	1,24	**1,17**	1,56	1,63	**1,60**
CG6866	loqs	loquacious	FBgn0032515	BKN20112	pos	yes	**8,66**	**2,80**	**0,93**	2,76	1,28	**2,02**	1,02	1,09	**1,06**	0,92	0,61	**0,76**
CG17168	CG17168	NA	FBgn0039943	BKN41284	pos	yes	**8,56**	**1,92**	**1,52**	1,86	2,16	**2,01**	1,08	1,26	**1,17**	1,76	1,78	**1,77**
CG10712	Chro	Chromator	FBgn0044324	BKN24506	pos	yes	**8,36**	**5,64**	**0,74**	2,69	2,20	**2,44**	1,18	1,22	**1,20**	0,67	0,67	**0,67**
CG2685	CG2685	NA	FBgn0024998	BKN46253	pos	yes	**8,07**	**1,83**	**1,75**	1,51	1,95	**1,73**	1,17	1,22	**1,19**	1,18	1,27	**1,23**
CG12085	pUf68	poly U binding factor 68kD	FBgn0028577	BKN27426	pos	yes	**7,99**	**2,45**	**1,22**	1,89	0,84	**1,36**	0,99	1,13	**1,06**	1,28	0,86	**1,07**
CG6546	Bap55	Brahma associated protein 55kD	FBgn0025716	BKN20365	pos	yes	**7,99**	**1,09**	**1,04**	1,33	1,26	**1,29**	1,13	1,06	**1,09**	0,82	0,90	**0,86**
CG12135	c12.1	c12.1	FBgn0040235	BKN22470	pos	yes	**7,54**	**2,54**	**1,46**	2,08	0,77	**1,43**	1,07	1,08	**1,07**	1,91	2,08	**2,00**
CG4602	Srp54	Srp54	FBgn0024285	BKN22462	pos	yes	**7,54**	**2,68**	**1,10**	2,09	1,57	**1,83**	0,63	1,07	**0,85**	1,31	1,12	**1,22**
CG7028	CG7028	NA	FBgn0027587	BKN50938	pos	yes	**7,49**	**1,58**	**1,33**	1,71	1,52	**1,62**	1,17	1,08	**1,13**	0,84	0,89	**0,86**
CG8241	pea	peanuts	FBgn0086895	BKN21969	pos	yes	**7,49**	**4,88**	**2,54**	3,25	4,52	**3,89**	1,19	1,14	**1,16**	2,58	3,42	**3,00**
CG9615	tex	tex	FBgn0037569	BKN21381	pos	yes	**7,42**	**1,97**	**1,25**	0,89	1,25	**1,07**	1,06	1,05	**1,05**	0,76	0,71	**0,73**
CG4528	snf	sans fille	FBgn0003449	BKN27649	pos	yes	**7,40**	**1,21**	**1,46**	1,61	2,66	**2,13**	1,20	1,10	**1,15**	1,04	1,36	**1,20**
CG3949	hoip	hoi-polloi	FBgn0015393	BKN28514	pos	no	**7,36**	**low viab.**	**2,93**	low viab.	low viab.	**low viab.**	1,53	1,98	**1,75**	low viab.	low viab.	**low viab.**

Gene information							dsRNA screening constructs			new dsRNA designs A and B								
							genome-wide screen DSB reporter	validation DSB reporter	validation high copy reporter	DSB reporter			miR-277 reporter			high-copy reporter		
CG-Nr.	Symbol	Name	GeneID	BKNID	Hit	validated	zscore_ loess	average fold change (n=2)	average fold change (n=2)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)
CG4266	CG4266	NA	FBgn0034598	BKN20202	pos	yes	**7,27**	**1,24**	**1,30**	1,02	1,73	**1,38**	1,22	1,21	**1,21**	1,28	1,37	**1,32**
CG33554	Nipped-A	Nipped-A	FBgn0053554	BKN51011	pos	yes	**7,25**	**1,54**	**1,01**	2,00	1,55	**1,78**	1,14	1,14	**1,14**	1,21	1,06	**1,13**
CG6987	SF2	SF2	FBgn0040284	BKN29395	pos	no	**7,25**	**2,98**	**1,08**	1,09	1,22	**1,15**	1,10	1,16	**1,13**	1,04	0,97	**1,01**
CG1676	cactin	cactin	FBgn0031114	BKN20603	pos	yes	**6,99**	**3,26**	**2,21**	2,64	3,10	**2,87**	1,08	1,23	**1,16**	2,30	2,19	**2,24**
CG8920	CG8920	NA	FBgn0027529	BKN24173	pos	yes	**6,92**	**1,32**	**0,72**	1,27	1,04	**1,16**	1,01	1,04	**1,03**	0,86	0,87	**0,87**
CG17520	CkIIalpha	casein kinase IIalpha	FBgn0264492	BKN50398	pos	yes	**6,89**	**1,31**	**1,03**	1,50	1,55	**1,53**	0,94	1,07	**1,01**	1,04	1,21	**1,12**
CG6339	rad50	rad50	FBgn0034728	BKN20510	pos	no	**6,85**	**1,30**	**0,77**	1,22	1,10	**1,16**	1,14	1,00	**1,07**	0,63	0,80	**0,72**
CG8025	Mtr3	Mtr3	FBgn0036916	BKN45761	pos	yes	**6,75**	**1,82**	**0,91**	2,21	1,83	**2,02**	1,10	1,03	**1,07**	1,08	1,24	**1,16**
CG1378	tll	tailless	FBgn0003720	BKN46350	pos	yes	**6,67**	**1,67**	**1,41**	6,57	1,33	**3,95**	1,05	1,10	**1,07**	1,06	0,93	**0,99**
CG3125	IntS6	Integrator 6	FBgn0261383	BKN45482	pos	no	**6,66**	**1,04**	**0,78**	1,00	0,99	**1,00**	1,04	1,02	**1,03**	0,61	0,65	**0,63**
CG13800	CG13800	NA	FBgn0035338	BKN23634	pos	no	**6,65**	**0,90**	**0,76**	0,91	1,09	**1,00**	1,02	1,17	**1,10**	0,93	0,92	**0,93**
CG7776	E(Pc)	Enhancer of Polycomb	FBgn0000581	BKN20790	pos	yes	**6,64**	**2,15**	**1,80**	2,97	1,09	**2,03**	1,46	1,28	**1,37**	0,93	1,09	**1,01**
CG7292	Rrp6	Rrp6	FBgn0038269	BKN20686	pos	yes	**6,46**	**4,16**	**1,10**	low viab.	1,27	**1,27**	low viab.	0,82	**0,82**	1,76	1,67	**1,72**
CG11132	DMAP1	DMAP1	FBgn0034537	BKN22326	pos	yes	**6,40**	**1,00**	**1,06**	1,84	1,43	**1,63**	1,14	1,15	**1,14**	0,91	0,93	**0,92**
CG7838	BubR1	Bub1-related kinase	FBgn0263855	BKN50563	pos	yes	**6,33**	**4,26**	**0,67**	1,57	1,79	**1,68**	1,65	1,52	**1,59**	0,81	0,94	**0,87**
CG5838	Dref	DNA replication-related element factor	FBgn0015664	BKN24401	pos	yes	**6,24**	**1,73**	**0,58**	1,84	2,59	**2,21**	1,09	1,03	**1,06**	0,75	0,75	**0,75**
CG3511	CG3511	NA	FBgn0035027	BKN28490	pos	yes	**6,19**	**2,18**	**1,32**	2,58	1,42	**2,00**	1,14	1,09	**1,11**	1,12	1,40	**1,26**
CG5454	snRNP-U1-C	small ribonucleoprotein particle U1 subunit C	FBgn0261792	BKN21376	pos	yes	**6,08**	**4,43**	**1,65**	1,29	2,11	**1,70**	1,08	1,19	**1,14**	0,95	1,28	**1,11**
CG6066	CG6066	NA	FBgn0039488	BKN28878	pos	yes	**6,02**	**1,07**	**1,13**	2,49	1,92	**2,21**	1,16	1,33	**1,25**	0,78	0,83	**0,81**
CG11207	feo	fascetto	FBgn0030241	BKN28553	pos	no	**6,02**	**2,05**	**1,28**	0,96	1,23	**1,10**	1,88	1,67	**1,78**	0,86	0,95	**0,90**
CG2577	CG2577	NA	FBgn0030384	BKN50153	pos	yes	**5,98**	**1,51**	**0,79**	1,01	2,20	**1,60**	1,11	1,11	**1,11**	0,82	0,97	**0,90**

Gene information							dsRNA screening constructs			new dsRNA designs A and B								
							genome-wide screen DSB reporter	validation DSB reporter	validation high copy reporter	DSB reporter			miR-277 reporter			high-copy reporter		
CG-Nr.	Symbol	Name	GeneID	BKNID	Hit	validated	zscore_ loess	average fold change (n=2)	average fold change (n=2)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)
CG4268	Pitslre	Pitslre	FBgn0016696	BKN28225	pos	yes	**5,97**	**1,68**	**2,01**	2,53	2,14	**2,34**	1,20	1,07	**1,13**	1,86	2,40	**2,13**
CG9667	CG9667	NA	FBgn0037550	BKN29247	pos	yes	**5,95**	**2,36**	**1,39**	1,34	1,87	**1,60**	1,05	1,05	**1,05**	1,07	1,54	**1,31**
CG2830	Hsp60B	Heat shock protein 60 related	FBgn0011244	BKN30688	pos	no	**5,91**	**2,94**	**1,58**	1,08	1,08	**1,08**	1,12	1,15	**1,14**	0,83	0,84	**0,84**
CG16838	CG16838	NA	FBgn0036574	BKN21251	pos	yes	**5,87**	**1,41**	**0,82**	1,37	1,26	**1,31**	1,02	1,11	**1,06**	0,88	0,82	**0,85**
CG14641	CG14641	NA	FBgn0037220	BKN27491	pos	yes	**5,81**	**2,72**	**2,35**	3,11	2,53	**2,82**	1,27	1,18	**1,22**	2,89	2,90	**2,89**
CG9797	CG9797	NA	FBgn0037621	BKN26226	pos	no	**5,80**	**1,95**	**1,07**	0,96	1,10	**1,03**	1,13	1,01	**1,07**	0,69	0,79	**0,74**
CG33109	CG33109	NA	FBgn0053109	BKN24267	pos	no	**5,79**	**2,09**	**0,76**	1,10	0,90	**1,00**	1,05	1,04	**1,04**	0,87	0,91	**0,89**
CG8114	pbl	pebble	FBgn0003041	BKN22310	pos	no	**5,75**	**1,15**	**0,82**	0,85	1,14	**1,00**	1,42	1,73	**1,57**	0,89	0,84	**0,87**
CG3193	crn	crooked neck	FBgn0000377	BKN20508	pos	no	**5,66**	**low viab.**	**3,90**	low viab.	low viab.	**low viab.**	low viab.	1,47	**1,47**	2,14	3,38	**2,76**
CG1258	pav	pavarotti	FBgn0011692	BKN20803	pos	yes	**5,54**	**1,32**	**0,80**	0,55	1,30	**0,93**	2,34	3,12	**2,73**	1,14	0,83	**0,99**
CG15432	CG15432	NA	FBgn0031603	BKN28733	pos	yes	**5,48**	**1,87**	**1,34**	1,53	1,21	**1,37**	1,04	1,11	**1,08**	1,42	0,92	**1,17**
CG7111	Rack1	Receptor of activated protein kinase C 1	FBgn0020618	BKN29073	pos	yes	**5,43**	**2,25**	**1,24**	1,19	2,08	**1,64**	1,09	1,04	**1,07**	0,92	1,09	**1,01**
CG42257	Snp	Snipper	FBgn0259142	BKN45478	pos	yes	**5,36**	**2,09**	**0,77**	1,77	1,42	**1,60**	1,07	1,09	**1,08**	1,16	1,01	**1,09**
CG15525	CG15525	NA	FBgn0039732	BKN45014	pos	yes	**5,34**	**1,31**	**1,00**	1,27	1,20	**1,23**	1,04	1,06	**1,05**	1,17	1,06	**1,11**
CG10754	CG10754	NA	FBgn0036314	BKN27781	pos	yes	**5,34**	**2,83**	**1,37**	0,93	1,58	**1,25**	1,12	1,17	**1,14**	1,15	1,62	**1,39**
CG30089	CG30089	NA	FBgn0050089	BKN31714	pos	no	**5,20**	**1,07**	**0,90**	1,20	0,66	**0,93**	1,06	1,15	**1,11**	1,11	0,73	**0,92**
CG11387	ct	cut	FBgn0004198	BKN45881	pos	no	**5,19**	**0,83**	**0,93**	0,89	0,75	**0,82**	1,35	1,73	**1,54**	0,86	0,60	**0,73**
CG8237	CG8237	NA	FBgn0033350	BKN45253	pos	no	**5,14**	**1,08**	**0,89**	1,89	1,08	**1,48**	0,99	0,95	**0,97**	0,77	0,84	**0,80**
CG6197	fand	fanfango	FBgn0033859	BKN21802	pos	yes	**5,09**	**3,80**	**3,29**	2,80	2,60	**2,70**	1,19	1,21	**1,20**	2,74	2,79	**2,76**
CG8264	Bx42	Bx42	FBgn0004856	BKN21649	pos	no	**5,04**	**low viab.**	**2,37**	low viab.	low viab.	**low viab.**	1,07	low viab.	**1,07**	2,31	2,35	**2,33**
CG3228	kz	kurz	FBgn0001330	BKN26359	pos	no	**5,03**	**0,89**	**0,91**	1,50	1,09	**1,29**	1,05	1,04	**1,04**	0,91	0,84	**0,87**
CG6620	ial	Ipll-aurora-like kinase	FBgn0024227	BKN32558	pos	yes	**5,02**	**2,33**	**0,83**	1,86	1,43	**1,64**	2,45	1,39	**1,92**	0,70	0,72	**0,71**

Gene information							dsRNA screening constructs			new dsRNA designs A and B								
							genome-wide screen DSB reporter	validation DSB reporter	validation high copy reporter	DSB reporter			miR-277 reporter			high-copy reporter		
CG-Nr.	Symbol	Name	GeneID	BKNID	Hit	validated	zscore_ loess	average fold change (n=2)	average fold change (n=2)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)
CG5519	Prp19	Prp19	FBgn0261119	BKN21386	pos	yes	**5,00**	**2,94**	**2,26**	1,70	2,18	**1,94**	1,45	1,20	**1,33**	1,49	2,35	**1,92**
CG4083	Mo25	Mo25	FBgn0017572	BKN45886	pos	yes	**4,99**	**1,17**	**1,00**	1,08	1,56	**1,32**	1,04	1,06	**1,05**	0,80	1,03	**0,91**
CG4980	CG4980	NA	FBgn0039558	BKN29267	pos	yes	**4,96**	**1,91**	**1,08**	1,35	1,08	**1,21**	1,13	1,05	**1,09**	1,03	1,24	**1,13**
CG10955	Rtf1	Rtf1	FBgn0034722	BKN21260	pos	yes	**4,94**	**3,03**	**1,14**	1,60	1,93	**1,76**	1,10	1,12	**1,11**	1,07	0,86	**0,96**
CG15309	CG15309	NA	FBgn0030183	BKN45574	pos	no	**4,90**	**1,04**	**0,92**	1,00	0,86	**0,93**	1,05	1,11	**1,08**	0,90	0,94	**0,92**
CG6443	CG6443	NA	FBgn0032290	BKN21026	pos	yes	**4,87**	**2,68**	**1,35**	2,04	1,59	**1,81**	1,19	1,07	**1,13**	1,25	1,13	**1,19**
CG8749	snRNP-U1-70K	small ribonucleo-protein particle U1 subunit 70K	FBgn0016978	BKN45552	pos	no	**4,79**	**2,43**	**0,78**	0,59	0,54	**0,57**	0,94	1,15	**1,05**	0,56	0,44	**0,50**
CG13298	CG13298	NA	FBgn0035692	BKN22482	pos	yes	**4,74**	**3,19**	**1,83**	2,15	2,74	**2,44**	1,14	1,85	**1,49**	1,68	2,13	**1,90**
CG15747	CG15747	NA	FBgn0030474	BKN22378	pos	yes	**4,70**	**2,62**	**1,18**	2,73	1,11	**1,92**	1,31	1,13	**1,22**	1,28	1,02	**1,15**
CG9752	CG9752	NA	FBgn0034614	BKN20542	pos	no	**4,63**	**3,07**	**2,74**	0,90	1,18	**1,04**	1,11	0,95	**1,03**	0,62	0,78	**0,70**
CG31256	Brf	Brf	FBgn0038499	BKN23074	pos	no	**4,62**	**1,30**	**1,01**	0,99	0,78	**0,88**	1,02	1,02	**1,02**	0,78	0,76	**0,77**
CG4798	l(2)k01209	lethal (2) k01209	FBgn0022029	BKN51139	pos	no	**4,62**	**2,09**	**0,71**	0,83	0,91	**0,87**	0,79	1,09	**0,94**	0,85	0,84	**0,84**
CG4527	slik	Sterile20-like kinase	FBgn0035001	BKN20098	pos	no	**4,59**	**1,52**	**1,03**	1,03	0,83	**0,93**	0,77	1,06	**0,92**	0,87	0,91	**0,89**
CG4211	nonA	no on or off transient A	FBgn0004227	BKN28107	pos	yes	**4,58**	**1,60**	**1,20**	1,31	1,48	**1,39**	1,05	1,03	**1,04**	0,99	1,09	**1,04**
CG6258	RfC38	Replication factor C 38kD subunit	FBgn0028700	BKN22321	pos	yes	**4,55**	**1,49**	**0,72**	1,31	2,13	**1,72**	1,19	1,16	**1,17**	0,49	0,87	**0,68**
CG11265	Trf4-1	NA	FBgn0030049	BKN22316	pos	yes	**4,50**	**0,82**	**0,90**	2,26	1,28	**1,77**	1,53	1,35	**1,44**	0,95	1,23	**1,09**
CG14230	CG14230	NA	FBgn0031062	BKN45873	pos	no	**4,50**	**1,38**	**0,82**	1,07	1,21	**1,14**	1,04	1,13	**1,08**	0,80	1,09	**0,94**
CG3582	U2af38	U2 small nuclear riboprotein auxiliary factor 38	FBgn0017457	BKN28210	pos	no	**4,47**	**1,52**	**1,02**	1,13	1,12	**1,12**	1,16	1,13	**1,14**	0,71	0,89	**0,80**
CG6694	ZC3H3	ZC3H3	FBgn0035900	BKN22095	pos	yes	**4,47**	**3,65**	**0,86**	3,33	1,21	**2,27**	1,02	0,97	**1,00**	1,67	1,94	**1,80**
CG9366	RhoL	rho-like	FBgn0014380	BKN31715	pos	no	**4,47**	**0,69**	**0,97**	0,99	0,85	**0,92**	0,93	0,99	**0,96**	0,78	0,91	**0,85**
CG6613	CG6613	NA	FBgn0034694	BKN20192	pos	no	**4,46**	**1,00**	**0,90**	0,74	0,72	**0,73**	1,07	1,01	**1,04**	0,77	0,87	**0,82**

Gene information							dsRNA screening constructs			new dsRNA designs A and B								
							genome-wide screen DSB reporter	validation DSB reporter	validation high copy reporter	DSB reporter			miR-277 reporter			high-copy reporter		
CG-Nr.	Symbol	Name	GeneID	BKNID	Hit	validated	zscore_ loess	average fold change (n=2)	average fold change (n=2)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)
CG6011	Prp18	Prp18	FBgn0027784	BKN27465	pos	yes	**4,45**	**2,67**	**2,00**	2,19	2,04	**2,12**	1,21	1,07	**1,14**	1,87	1,84	**1,86**
CG1945	faf	fat facets	FBgn0005632	BKN20871	pos	no	**4,45**	**1,16**	**0,76**	1,20	0,82	**1,01**	0,99	1,12	**1,05**	0,90	0,76	**0,83**
CG6995	Saf-B	Scaffold attachment factor B	FBgn0039229	BKN20574	pos	yes	**4,37**	**1,73**	**0,72**	1,17	2,11	**1,64**	1,17	1,21	**1,19**	0,87	1,02	**0,94**
CG2807	CG2807	NA	FBgn0031266	BKN21741	pos	no	**4,35**	**1,55**	**3,10**	low viab.	low viab.	**low viab.**	2,03	2,09	**2,06**	2,05	2,91	**2,48**
CG2021	CG2021	NA	FBgn0035271	BKN27671	pos	no	**4,35**	**0,87**	**0,83**	0,95	1,26	**1,10**	1,02	1,13	**1,07**	0,97	0,77	**0,87**
CG8815	Sin3A	Sin3A	FBgn0022764	BKN20796	pos	yes	**4,33**	**3,25**	**0,48**	3,10	low viab.	**3,10**	0,89	low viab.	**0,89**	0,92	1,40	**1,16**
CG9186	CG9186	NA	FBgn0035206	BKN28771	pos	no	**4,29**	**1,71**	**1,41**	1,13	0,91	**1,02**	1,04	1,05	**1,05**	0,76	0,84	**0,80**
CG11205	phr	photorepair	FBgn0003082	BKN30452	pos	no	**4,27**	**2,96**	**0,67**	1,07	1,13	**1,10**	1,11	1,18	**1,14**	0,86	0,88	**0,87**
CG3056	ssx	sister-of-Sex-lethal	FBgn0024987	BKN29594	pos	yes	**4,26**	**1,83**	**0,73**	1,55	1,09	**1,32**	1,00	1,02	**1,01**	0,90	0,80	**0,85**
CG31368	CG31368	NA	FBgn0051368	BKN20937	pos	no	**4,23**	**3,69**	**2,01**	1,11	0,79	**0,95**	1,05	1,03	**1,04**	1,03	1,09	**1,06**
CG3450	ubl	ubiquitin like	FBgn0022224	BKN42070	pos	no	**4,22**	**0,98**	**1,10**	0,84	1,21	**1,02**	1,11	1,12	**1,11**	1,09	1,06	**1,08**
CG7177	Wnk	NA	FBgn0037098	BKN50794	pos	yes	**4,19**	**2,14**	**0,92**	1,35	1,73	**1,54**	1,05	1,40	**1,22**	0,84	0,73	**0,78**
CG7917	Nlp	Nucleoplasmin	FBgn0016685	BKN30946	pos	yes	**4,19**	**2,50**	**0,97**	1,82	1,15	**1,48**	1,02	1,03	**1,03**	0,83	0,82	**0,82**
CG11489	srpk79D	serine-arginine protein kinase at 79D	FBgn0025702	BKN23592	pos	no	**4,15**	**1,28**	**1,03**	0,99	0,91	**0,95**	1,01	1,26	**1,13**	0,96	0,78	**0,87**
CG1796	Tango4	Transport and Golgi organization 4	FBgn0030365	BKN29170	pos	yes	**4,14**	**3,13**	**1,99**	2,29	2,03	**2,16**	1,33	1,37	**1,35**	2,09	1,96	**2,03**
CG5949	DNApol-delta	DNA-polymerase-delta	FBgn0263600	BKN29129	pos	yes	**4,14**	**1,16**	**0,66**	1,84	1,30	**1,57**	1,22	1,11	**1,16**	0,89	0,74	**0,81**
CG10492	CG10492	NA	FBgn0032748	BKN27409	pos	no	**4,11**	**1,13**	**0,90**	1,23	1,49	**1,36**	1,23	1,38	**1,31**	0,83	0,64	**0,73**
CG8877	Prp8	pre-mRNA processing factor 8	FBgn0033688	BKN20780	pos	yes	**4,09**	**1,85**	**2,22**	2,44	2,31	**2,38**	1,50	1,33	**1,41**	2,18	2,19	**2,19**
CG33956	kay	kayak	FBgn0001297	BKN32849	pos	yes	**4,09**	**2,84**	**0,62**	4,17	2,32	**3,24**	1,26	1,64	**1,45**	0,93	0,78	**0,85**
CG17498	mad2	mad2	FBgn0035640	BKN28038	pos	yes	**4,08**	**4,45**	**0,92**	1,53	1,62	**1,57**	1,05	1,35	**1,20**	0,96	1,02	**0,99**
CG2674	Sam-S	S-adenosylmethio-nine Synthetase	FBgn0005278	BKN46638	pos	no	**4,07**	**0,71**	**1,08**	1,07	1,15	**1,11**	0,95	0,96	**0,95**	0,92	0,98	**0,95**
CG32604	l(1)G0007	lethal (1) G0007	FBgn0026713	BKN22970	pos	yes	**4,06**	**1,98**	**1,21**	1,57	1,24	**1,40**	1,06	1,11	**1,08**	1,28	1,19	**1,23**

Gene information							dsRNA screening constructs			new dsRNA designs A and B								
							genome-wide screen DSB reporter	validation DSB reporter	validation high copy reporter	DSB reporter			miR-277 reporter			high-copy reporter		
CG-Nr.	Symbol	Name	GeneID	BKNID	Hit	validated	zscore_ loess	average fold change (n=2)	average fold change (n=2)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)
CG33547	Rim	Rim	FBgn0053547	BKN45626	pos	yes	**4,04**	**1,17**	**0,98**	1,21	2,60	**1,91**	1,11	1,24	**1,18**	1,10	1,02	**1,06**
CG14220	CG14220	NA	FBgn0031036	BKN22791	pos	yes	**4,03**	**1,77**	**0,81**	2,95	1,59	**2,27**	1,14	1,05	**1,09**	1,08	1,00	**1,04**
CG12165	Incenp	Inner centromere protein	FBgn0260991	BKN26980	pos	yes	**3,98**	**2,78**	**0,97**	1,59	2,27	**1,93**	1,56	1,85	**1,71**	0,80	0,72	**0,76**
CG9659	egh	egghead	FBgn0001404	BKN27178	pos	no	**3,90**	**0,96**	**0,75**	0,96	0,77	**0,86**	1,22	1,05	**1,13**	0,72	0,88	**0,80**
CG1620	CG1620	NA	FBgn0033183	BKN50407	pos	no	**3,89**	**1,28**	**1,09**	0,89	1,04	**0,96**	1,06	1,08	**1,07**	0,92	0,83	**0,87**
CG3069	Taf10b	TBP-associated factor 10b	FBgn0026324	BKN22306	pos	no	**3,86**	**1,18**	**0,71**	0,97	1,06	**1,01**	1,11	1,20	**1,15**	0,69	0,78	**0,73**
CG8677	CG8677	NA	FBgn0026577	BKN22027	pos	no	**3,75**	**1,10**	**0,73**	0,86	1,94	**1,40**	0,94	0,99	**0,97**	0,82	0,94	**0,88**
CG3971	Baldspot	Baldspot	FBgn0260960	BKN29085	pos	no	**3,72**	**1,24**	**1,12**	0,90	1,10	**1,00**	1,05	1,06	**1,06**	0,86	0,92	**0,89**
CG9772	Skp2	NA	FBgn0037236	BKN28837	pos	no	**3,72**	**1,30**	**0,93**	1,13	0,94	**1,03**	1,16	1,03	**1,09**	0,87	0,80	**0,84**
CG6121	Tip60	Tip60	FBgn0026080	BKN27642	pos	no	**3,71**	**1,50**	**1,22**	1,12	1,19	**1,16**	1,57	1,22	**1,40**	0,79	0,95	**0,87**
CG8912	Psi	P-element somatic inhibitor	FBgn0014870	BKN21933	pos	yes	**3,62**	**1,57**	**0,80**	0,89	1,79	**1,34**	1,05	1,07	**1,06**	0,74	0,92	**0,83**
CG4400	CG4400	NA	FBgn0030434	BKN28153	pos	yes	**3,61**	**2,04**	**0,77**	1,36	1,36	**1,36**	0,59	1,09	**0,84**	0,98	1,04	**1,01**
CG6046	Bin1	Bicoid interacting protein 1	FBgn0024491	BKN21416	pos	no	**3,56**	**1,42**	**0,97**	1,07	0,99	**1,03**	1,13	1,06	**1,09**	0,76	1,01	**0,89**
CG5499	His2Av	Histone H2A variant	FBgn0001197	BKN28652	pos	no	**3,47**	**1,22**	**1,07**	0,78	0,91	**0,85**	3,74	0,93	**2,34**	0,80	0,79	**0,79**
CG7238	sip1	septin interacting protein 1	FBgn0024191	BKN28960	pos	yes	**3,39**	**1,46**	**1,00**	1,48	0,98	**1,23**	1,06	1,00	**1,03**	0,77	0,81	**0,79**
CG8174	SRPK	SRPK	FBgn0026370	BKN50500	pos	no	**3,29**	**1,06**	**0,95**	1,17	0,97	**1,07**	1,17	1,05	**1,11**	0,91	0,93	**0,92**
CG11184	Upf3	Upf3	FBgn0034923	BKN45942	neg	no	**-2,86**	**0,84**	**0,46**	0,86	1,08	**0,97**	1,16	1,18	**1,17**	0,51	0,53	**0,52**
CG9193	mus209	mutagen-sensitive 209	FBgn0005655	BKN28382	neg	no	**-3,12**	**1,10**	**2,30**	0,84	0,71	**0,78**	1,10	1,16	**1,13**	0,80	0,95	**0,87**
CG10228	Pcf11	Pcf11	FBgn0264962	BKN20818	neg	yes	**-3,14**	**0,45**	**0,72**	0,46	1,72	**1,09**	1,11	1,79	**1,45**	0,76	0,66	**0,71**
CG12357	Cbp20	cap binding protein 20	FBgn0022943	BKN29448	neg	no	**-3,48**	**1,29**	**0,62**	0,76	1,24	**1,00**	0,98	1,07	**1,03**	0,72	0,92	**0,82**
CG14884	CSN5	COP9 complex homolog subunit 5	FBgn0027053	BKN20306	neg	yes	**-3,59**	**0,50**	**0,73**	0,56	0,55	**0,56**	0,93	0,98	**0,95**	1,01	0,77	**0,89**
CG11979	Rpb5	Rpb5	FBgn0033571	BKN20359	neg	no	**-3,62**	**1,62**	**0,50**	1,25	3,78	**2,52**	1,07	1,15	**1,11**	0,86	0,77	**0,81**

Gene information							dsRNA screening constructs			new dsRNA designs A and B								
							genome-wide screen DSB reporter	validation DSB reporter	validation high copy reporter	DSB reporter			miR-277 reporter			high-copy reporter		
CG-Nr.	Symbol	Name	GeneID	BKNID	Hit	validated	zscore_ loess	average fold change (n=2)	average fold change (n=2)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)
CG9938	Ndc80	Ndc80	FBgn0030500	BKN23993	neg	yes	**-3,62**	**0,57**	**0,92**	0,86	0,58	**0,72**	1,08	1,02	**1,05**	0,72	0,78	**0,75**
CG6376	E2f	E2F transcription factor	FBgn0011766	BKN20400	neg	no	**-3,70**	**0,98**	**0,86**	10,40	2,10	**6,25**	5,18	2,26	**3,72**	0,82	0,70	**0,76**
CG3889	CSN1b	COP9 complex homolog subunit 1b	FBgn0027057	BKN21757	neg	yes	**-3,70**	**0,55**	**0,79**	0,65	0,70	**0,68**	0,99	0,92	**0,96**	0,71	0,79	**0,75**
CG8609	MED4	Mediator complex subunit 4	FBgn0035754	BKN28661	neg	no	**-3,71**	**0,64**	**0,64**	0,88	0,81	**0,85**	1,09	1,18	**1,14**	0,59	0,62	**0,61**
CG8867	Jon25Bi	Jonah 25Bi	FBgn0020906	BKN31014	neg	yes	**-3,71**	**0,56**	**0,73**	1,22	0,72	**0,97**	0,83	1,00	**0,91**	0,77	0,83	**0,80**
CG1810	mRNA-cap	mRNA-capping-enzyme	FBgn0030556	BKN60320	neg	yes	**-3,75**	**0,77**	**0,63**	0,66	0,50	**0,58**	0,95	1,01	**0,98**	0,62	0,49	**0,56**
CG31716	Cnot4	Cnot 4 homologue	FBgn0051716	BKN22063	neg	no	**-3,75**	**0,83**	**0,75**	0,77	1,05	**0,91**	1,12	1,12	**1,12**	0,72	0,73	**0,73**
CG31671	tho2	tho2	FBgn0031390	BKN22082	neg	yes	**-3,75**	**0,86**	**0,66**	0,52	0,62	**0,57**	1,12	1,09	**1,10**	0,56	0,71	**0,64**
CG10572	Cdk8	Cyclin-dependent kinase 8	FBgn0015618	BKN45143	neg	no	**-3,78**	**0,65**	**0,84**	1,16	1,02	**1,09**	1,07	1,22	**1,14**	0,92	1,19	**1,05**
CG1135	Rcd5	Reduction in Cnn dots 5	FBgn0263832	BKN21572	neg	no	**-3,82**	**1,22**	**0,68**	3,76	1,71	**2,74**	1,18	0,96	**1,07**	1,23	0,82	**1,02**
CG7850	puc	puckered	FBgn0243512	BKN46004	neg	no	**-3,82**	**0,65**	**0,83**	0,94	1,00	**0,97**	2,50	1,17	**1,84**	0,81	0,96	**0,88**
CG7597	Cdk12	NA	FBgn0037093	BKN21793	neg	no	**-3,87**	**0,85**	**0,96**	1,03	1,30	**1,17**	1,31	1,48	**1,39**	0,72	0,88	**0,80**
CG10223	Top2	Topoisomerase 2	FBgn0003732	BKN21781	neg	yes	**-3,88**	**0,70**	**0,56**	0,66	0,54	**0,60**	1,09	1,13	**1,11**	0,86	0,87	**0,86**
CG5442	SC35	SC35	FBgn0265298	BKN32717	neg	no	**-3,91**	**0,91**	**0,68**	0,77	0,67	**0,72**	1,07	1,04	**1,05**	0,67	0,65	**0,66**
CG7528	Uba2	Smt3 activating enzyme 2	FBgn0029113	BKN22286	neg	yes	**-3,93**	**0,65**	**0,71**	0,92	0,65	**0,79**	1,27	1,10	**1,18**	0,93	1,10	**1,02**
CG10719	brat	brain tumor	FBgn0010300	BKN33206	neg	no	**-3,94**	**1,23**	**0,64**	1,69	1,18	**1,43**	1,23	1,07	**1,15**	0,78	0,69	**0,74**
CG6297	JIL-1	JIL-1	FBgn0020412	BKN50657	neg	yes	**-3,96**	**0,61**	**1,11**	0,49	0,51	**0,50**	1,17	1,01	**1,09**	0,76	0,89	**0,83**
CG10542	Bre1	Bre1	FBgn0086694	BKN20850	neg	yes	**-3,98**	**0,76**	**0,74**	0,89	0,68	**0,79**	1,25	1,23	**1,24**	0,94	0,90	**0,92**
CG8863	Droj2	DnaJ-like-2	FBgn0038145	BKN27834	neg	no	**-4,00**	**1,24**	**0,86**	1,33	0,88	**1,11**	1,11	1,24	**1,17**	0,78	0,81	**0,80**
CG42550	CG42550	NA	FBgn0260723	BKN33975	neg	no	**-4,02**	**0,75**	**0,84**	1,05	1,14	**1,09**	1,04	1,13	**1,09**	1,02	0,98	**1,00**
CG4494	smt3	smt3	FBgn0264922	BKN21837	neg	yes	**-4,06**	**0,95**	**0,81**	0,71	0,54	**0,62**	1,13	1,28	**1,20**	0,94	1,03	**0,99**
CG6513	endos	endosulfine	FBgn0061515	BKN28642	neg	yes	**-4,06**	**0,74**	**0,95**	0,53	0,35	**0,44**	1,10	1,04	**1,07**	0,75	0,81	**0,78**

Gene information							dsRNA screening constructs			new dsRNA designs A and B								
							genome-wide screen DSB reporter	validation DSB reporter	validation high copy reporter	DSB reporter			miR-277 reporter			high-copy reporter		
CG-Nr.	Symbol	Name	GeneID	BKNID	Hit	validated	zscore_ loess	average fold change (n=2)	average fold change (n=2)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)
CG33106	mask	multiple ankyrin repeats single KH domain	FBgn0043884	BKN20625	neg	no	**-4,13**	**0,84**	**0,71**	0,84	1,24	**1,04**	0,77	0,70	**0,73**	0,59	0,67	**0,63**
CG7697	CstF-64	Cleavage stimulation factor 64 kD subunit	FBgn0027841	BKN27450	neg	yes	**-4,19**	**0,57**	**0,58**	0,76	0,67	**0,71**	1,23	1,03	**1,13**	0,27	0,92	**0,60**
CG2910	nito	spenito	FBgn0027548	BKN28409	neg	yes	**-4,20**	**0,76**	**0,63**	1,69	0,61	**1,15**	1,22	1,08	**1,15**	0,68	0,59	**0,63**
CG7163	mkg-p	monkey king protein	FBgn0035889	BKN40366	neg	yes	**-4,20**	**0,44**	**0,81**	0,54	0,68	**0,61**	1,11	0,99	**1,05**	0,74	0,85	**0,80**
CG18009	Trf2	TATA box binding protein-related factor 2	FBgn0261793	BKN27538	neg	no	**-4,22**	**1,35**	**0,79**	0,91	0,85	**0,88**	1,57	1,74	**1,66**	0,61	0,56	**0,58**
CG2097	Sym	Symplekin	FBgn0037371	BKN21288	neg	yes	**-4,22**	**0,78**	**0,54**	0,60	0,56	**0,58**	1,07	1,12	**1,10**	0,62	0,53	**0,57**
CG17437	wds	will die slowly	FBgn0040066	BKN20216	neg	no	**-4,23**	**0,72**	**0,85**	1,34	1,15	**1,25**	1,01	1,10	**1,06**	0,91	1,11	**1,01**
CG10583	Sse	Separase	FBgn0035627	BKN20595	neg	yes	**-4,26**	**0,61**	**0,79**	3,32	0,59	**1,96**	0,96	1,36	**1,16**	1,11	1,00	**1,05**
CG7200	CG7200	NA	FBgn0032671	BKN28118	neg	no	**-4,27**	**0,98**	**0,93**	0,80	1,01	**0,91**	1,09	0,98	**1,04**	0,78	0,86	**0,82**
CG32743	nonC	no-on-and-no-off transient C	FBgn0263968	BKN50537	neg	yes	**-4,27**	**0,56**	**0,62**	0,62	0,75	**0,68**	1,16	1,13	**1,14**	0,51	0,58	**0,55**
CG12109	Caf1-180	Caf1-180	FBgn0030054	BKN27324	neg	no	**-4,32**	**1,00**	**0,78**	0,54	0,89	**0,72**	1,13	1,35	**1,24**	0,73	0,82	**0,78**
CG3857	CG3857	NA	FBgn0023520	BKN30209	neg	yes	**-4,34**	**0,65**	**0,70**	1,05	0,72	**0,88**	1,01	1,04	**1,03**	1,11	0,81	**0,96**
CG7719	gwl	greatwall	FBgn0260399	BKN51041	neg	yes	**-4,38**	**0,57**	**0,94**	0,52	0,41	**0,46**	1,02	0,96	**0,99**	0,98	0,90	**0,94**
CG31709	CG31709	NA	FBgn0051709	BKN25255	neg	no	**-4,44**	**1,08**	**0,92**	1,22	3,60	**2,41**	1,10	1,11	**1,11**	0,85	1,06	**0,95**
CG5650	Pp1-87B	Protein phosphatase 1 at 87B	FBgn0004103	BKN31383	neg	yes	**-4,46**	**0,64**	**0,71**	0,85	0,69	**0,77**	1,08	1,18	**1,13**	0,86	0,93	**0,90**
CG7698	Cpsf73	Cleavage and polyadenylation specificity factor 73	FBgn0261065	BKN20534	neg	yes	**-4,46**	**0,70**	**0,66**	0,63	0,69	**0,66**	1,06	0,99	**1,02**	0,55	0,68	**0,62**
CG14418	CG14418	NA	FBgn0040354	BKN31933	neg	no	**-4,60**	**0,41**	**0,59**	1,38	low viab.	**1,38**	0,99	1,74	**1,36**	1,00	1,72	**1,36**
CG31258	Cenp-C	Cenp-C	FBgn0086697	BKN23772	neg	no	**-4,70**	**1,36**	**0,74**	0,79	1,00	**0,89**	1,27	1,32	**1,29**	0,69	0,75	**0,72**
CG2261	CstF-50	CstF-50	FBgn0039867	BKN27473	neg	yes	**-4,70**	**0,88**	**0,81**	0,57	0,48	**0,52**	1,17	1,19	**1,18**	0,56	0,62	**0,59**
CG5514	CG5514	NA	FBgn0039560	BKN23035	neg	yes	**-4,72**	**0,58**	**0,67**	0,48	0,52	**0,50**	1,20	1,16	**1,18**	0,70	0,68	**0,69**

Gene information							dsRNA screening constructs			new dsRNA designs A and B								
							genome-wide screen DSB reporter	validation DSB reporter	validation high copy reporter	DSB reporter			miR-277 reporter			high-copy reporter		
CG-Nr.	Symbol	Name	GeneID	BKNID	Hit	validated	zscore_ loess	average fold change (n=2)	average fold change (n=2)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)	DesignA average fold change (n=2)	DesignB average fold change (n=2)	DesignA and B combined (n=2 each)
CG1109	CG1109	NA	FBgn0046222	BKN20307	neg	yes	**-4,74**	**0,32**	**0,38**	0,22	0,54	**0,38**	1,13	0,97	**1,05**	0,39	0,45	**0,42**
CG11451	Spc105R	Spc105-related	FBgn0037025	BKN26153	neg	no	**-4,78**	**1,35**	**0,74**	1,38	1,49	**1,43**	1,27	1,43	**1,35**	0,87	0,89	**0,88**
CG4976	Mes-4	Mes-4	FBgn0039559	BKN20996	neg	no	**-4,86**	**0,68**	**0,67**	0,99	0,80	**0,89**	0,66	1,10	**0,88**	0,90	0,89	**0,90**
CG2253	Upf2	Upf2	FBgn0029992	BKN20728	neg	yes	**-5,24**	**0,67**	**0,50**	0,55	0,61	**0,58**	1,21	1,30	**1,25**	0,47	0,56	**0,52**
CG6588	Fas1	Fasciclin 1	FBgn0262742	BKN29572	neg	no	**-5,32**	**0,97**	**0,76**	0,82	0,84	**0,83**	1,14	1,03	**1,09**	0,73	0,89	**0,81**
CG3836	stwl	stonewall	FBgn0003459	BKN45199	neg	yes	**-5,34**	**0,61**	**0,81**	0,97	0,56	**0,77**	1,36	1,11	**1,24**	1,08	0,78	**0,93**
CG6593	Pp1alpha-96A	Protein phosphatase 1alpha at 96A	FBgn0003134	BKN28922	neg	yes	**-5,39**	**0,47**	**0,72**	0,73	0,87	**0,80**	1,12	0,98	**1,05**	0,72	0,68	**0,70**
CG8068	Su(var)2-10	Suppressor of variegation 2-10	FBgn0003612	BKN28790	neg	yes	**-5,48**	**0,55**	**1,14**	0,64	0,44	**0,54**	1,26	1,12	**1,19**	0,81	0,81	**0,81**
CG10279	Rm62	Rm62	FBgn0003261	BKN32269	neg	yes	**-5,73**	**0,53**	**0,49**	0,34	0,42	**0,38**	1,05	1,11	**1,08**	0,56	0,52	**0,54**
CG3642	Clp	Clipper	FBgn0015621	BKN27764	neg	yes	**-5,77**	**0,40**	**0,47**	0,46	1,91	**1,19**	1,13	1,07	**1,10**	0,52	0,68	**0,60**
CG1559	Upf1	Upf1	FBgn0030354	BKN27828	neg	yes	**-5,86**	**0,75**	**0,44**	0,50	0,76	**0,63**	1,14	1,40	**1,27**	0,54	0,39	**0,46**
CG1957	Cpsf100	Cleavage and polyadenylation specificity factor100	FBgn0027873	BKN40434	neg	yes	**-5,98**	**0,76**	**0,64**	0,57	0,91	**0,74**	1,07	1,09	**1,08**	0,59	0,67	**0,63**
CG10110	Cpsf160	Cleavage and polyadenylation specificity factor160	FBgn0024698	BKN23031	neg	yes	**-6,55**	**0,48**	**0,59**	0,27	0,31	**0,29**	1,08	1,18	**1,13**	0,53	0,37	**0,45**
CG12640	CG12640	NA	FBgn0030202	BKN30871	neg	no	**-6,56**	**0,77**	**0,79**	1,07	0,99	**1,03**	0,98	1,07	**1,02**	0,80	0,97	**0,89**
CG9748	bel	belle	FBgn0263231	BKN28143	neg	yes	**-6,56**	**0,59**	**0,57**	0,84	0,55	**0,69**	1,31	1,11	**1,21**	0,59	0,48	**0,54**
CG4254	tsr	twinstar	FBgn0011726	BKN28706	neg	no	**-7,37**	**1,27**	**0,90**	0,93	0,91	**0,92**	1,23	1,08	**1,15**	0,94	1,03	**0,99**
CG8954	Smg5	Smg5	FBgn0019890	BKN22183	neg	yes	**-7,51**	**0,91**	**0,44**	0,55	0,53	**0,54**	1,10	1,22	**1,16**	0,46	0,42	**0,44**
CG2186	CG2186	NA	FBgn0030243	BKN31161	neg	no	**-8,01**	**0,51**	**0,58**	1,37	0,93	**1,15**	1,29	1,19	**1,24**	0,76	0,71	**0,73**
exclusion criteria positive candidates						63-33 >=1.6 false positive		**<1,16**		**<1,24**	**<1,24**	**<1,24**	**>1,25**	**>1,25**	**>1,25**			
exclusion criteria negative candidates								**>0,84**		**>0,76**	**>0,76**	**>0,76**	**<0,85**	**<0,85**	**<0,85**			
positive controls:	Dcr-2	Dicer-2						4,10	2,22			3,08			1,36			2,23
	Ago2	Argonaute-2						15,23	1,97			8,84			5,08			2,24