

RESEARCH ARTICLE

# Splicing stimulates siRNA formation at *Drosophila* DNA double-strand breaks

Karin Merk<sup>1</sup>, Marco Breinig<sup>2</sup>, Romy Böttcher<sup>1</sup>, Stefan Krebs<sup>3</sup>, Helmut Blum<sup>3</sup>, Michael Boutros<sup>2</sup>, Klaus Förstemann<sup>1\*</sup>

**1** Gene Center and Dept. of Biochemistry, Ludwig-Maximilians-Universität München, München, Germany, **2** Division Signaling and Functional Genomics, German Cancer Research Center (DKFZ) and Heidelberg University, Im Neuenheimer Feld 580, Heidelberg, Germany, **3** Laboratory for Functional Genome Analysis (LAFUGA), Gene Center, Ludwig-Maximilians-Universität München, München, Germany

\* [Foerstemann@Imb.uni-muenchen.de](mailto:Foerstemann@Imb.uni-muenchen.de)



**OPEN ACCESS**

**Citation:** Merk K, Breinig M, Böttcher R, Krebs S, Blum H, Boutros M, et al. (2017) Splicing stimulates siRNA formation at *Drosophila* DNA double-strand breaks. *PLoS Genet* 13(6): e1006861. <https://doi.org/10.1371/journal.pgen.1006861>

**Editor:** Rene F. Ketting, Institute of Molecular Biology, GERMANY

**Received:** December 5, 2016

**Accepted:** June 9, 2017

**Published:** June 19, 2017

**Copyright:** © 2017 Merk et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The original screening data has been deposited at <http://www.genomernai.org>. The deep sequencing datasets are available at the European Nucleotide archive (ENA) under the accession number PRJEB20896.

**Funding:** This work was supported by grant SFB646/TPA13 from the Deutsche Forschungsgemeinschaft (DFG) to KF. The work in the lab of MB was in part supported by a DFG DRiC grant. The funders had no role in study design,

## Abstract

DNA double-strand breaks trigger the production of locus-derived siRNAs in fruit flies, human cells and plants. At least in flies, their biogenesis depends on active transcription running towards the break. Since siRNAs derive from a double-stranded RNA precursor, a major question is how broken DNA ends can generate matching sense and antisense transcripts. We performed a genome-wide RNAi-screen in cultured *Drosophila* cells, which revealed that in addition to DNA repair factors, many spliceosome components are required for efficient siRNA generation. We validated this observation through site-specific DNA cleavage with CRISPR-cas9 followed by deep sequencing of small RNAs. DNA breaks in intron-less genes or upstream of a gene's first intron did not efficiently trigger siRNA production. When DNA double-strand breaks were induced downstream of an intron, however, this led to robust siRNA generation. Furthermore, a downstream break slowed down splicing of the upstream intron and a detailed analysis of siRNA coverage at the targeted locus revealed that unspliced pre-mRNA contributes the sense strand to the siRNA precursor. Since splicing factors are stimulating the response but unspliced transcripts are entering the siRNA biogenesis, the spliceosome is apparently stalled in a pre-catalytic state and serves as a signaling hub. We conclude that convergent transcription at DNA breaks is stimulated by a splicing dependent control process. The resulting double-stranded RNA is converted into siRNAs that instruct the degradation of cognate mRNAs. In addition to a potential role in DNA repair, the break-induced transcription may thus be a means to cull improper RNAs from the transcriptome of *Drosophila melanogaster*. Since the splicing factors identified in our screen also stimulated siRNA production from high copy transgenes, it is possible that this surveillance mechanism serves in genome defense beyond DNA double-strand breaks.

## Author summary

DNA damage ultimately threatens the integrity of our genome; in addition, it is an impediment for transcription and thus affects cells independently of the mutagenic effects. DNA double-strand breaks in transcribed regions can induce the production of corresponding

data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

small interfering RNAs, a class of regulators that is well known for controlling gene expression. The reason why they are generated at sites of DNA damage is not understood, putative roles in DNA repair remain controversial. Thus, the function of DNA damage-induced siRNAs is a major open question. In an unbiased approach, we screened *Drosophila* cells genome-wide to detect those factors that are required to produce damage-induced siRNAs. In addition to DNA repair proteins, we found that many splicing factors and spliceosome components were necessary. The need for an intron upstream of the break was confirmed by *cas9*-CRISPR mediated cleavage, followed by small RNA sequencing. A detailed look at the siRNA coverage revealed nonetheless that unspliced transcripts give rise to siRNAs. Hence, the spliceosome must be stalled and signal to the transcription machinery. The new role of damage-induced siRNAs in local RNA surveillance should be considered on par with their putative function during DNA repair.

## Introduction

DNA is constantly challenged by mutagenic processes of extrinsic and intrinsic origin. Of these damages, DNA double-strand breaks are particularly problematic lesions because they disrupt the continuity of genetic information. Their repair can either proceed via end-joining activities or through homology-directed repair [1]. A detailed mechanistic understanding of these repair processes is not only important for the prevention and treatment of diseases, such as cancer, but also to help researchers direct the outcome of genome editing experiments with precision [2, 3]. The information stored in DNA is read-out by the process of transcription into RNA. DNA damage is therefore not only an impediment for replication, but also a hindrance for RNA biosynthesis. If DNA damage has occurred in an actively transcribed region, concomitant action of DNA repair factors and the transcription machinery is not possible; access to the DNA must thus be regulated [4]. Stalling of an RNA polymerase upstream of the damaged site may lead to extensive RNA-DNA hybrids, called R-loops, which can themselves cause genomic instability [5–7]. Consequently, a domain of specific chromatin states is assembled around a DNA double-strand break and transcriptional silencing can occur [8]. On the other hand, RNA polymerases stalled by certain types of base damage serve as sensors and thus promote repair of the lesion during transcription-couple repair [9, 10]. It is also established that some RNA binding proteins are recruited to sites of DNA damage [5] and non-coding transcription may play an important role in DNA repair [11, 12]. Finally, the Prp19 component of the spliceosome can interact with RPA bound to single-stranded DNA and reinforces activation of the protein kinase ATR in a manner that is independent of its function during the splicing reaction [13].

Recently, formation of locus-specific siRNAs has been observed at DNA double-strand breaks [14–17]. They may promote repair via homologous recombination in mammalian cells and plants [15, 18], but the molecular mechanisms through which siRNAs promote homologous recombination are not fully established; in mammalian cells they may involve targeting of Rad51 to the damaged site via protein-protein interactions with Ago2 [18], but it is challenging to exclude indirect effects via perturbed miRNA biogenesis in these experiments. In *Neurospora crassa*, their biogenesis is even dependent on, rather than important for, homologous recombination [19–21]. In *Drosophila* the generation of DNA damage-induced siRNAs depends on transcription and is limited to only one side of the broken DNA, the region between a transcription start site and the DNA end. It was thus proposed that the DNA ends serve as transcription initiation sites to generate corresponding antisense transcripts to

generate dsRNA [16]. Consistently, transcription initiation at DNA breaks has now been directly observed in *S. pombe* [12]. However, no DNA repair defects could be observed in *dcr2* mutant *Drosophila melanogaster* flies where the siRNA pathway is completely inactivated but the miRNA pathway remains mostly unperturbed [22]. Although the significance of the siRNAs for DNA repair is thus a matter of debate, their presence demonstrates that transcripts running towards a DNA DSB are subject to some sort of surveillance and, as a consequence, at least partially converted into double-stranded RNA (the precursor of siRNAs).

To shed light on the mechanistic details of this process, we conducted a genome-wide RNA interference screen in cultured *Drosophila* cells using siRNAs generated from a linearized plasmid. These siRNAs control expression of a reporter gene. In addition to DNA double-strand break repair proteins, we discovered that many splicing factors, in particular components from the Prp19 and Prp19-related spliceosome sub-complexes are important for siRNA generation. Consistently, the presence of upstream introns greatly stimulated siRNA generation at chromosomal DNA DSBs induced by CRISPR-*cas9*. An intron-less gene, on the other hand, only generated few siRNAs upon cleavage. We propose that the perturbed transcript maturation that ensues when RNA polymerase II encounters a DNA double-strand break is sensed with participation of the spliceosome. As a consequence, double-stranded RNA is generated and processed into siRNAs. Intriguingly, the splicing factors identified in our screen for break-induced siRNA generation were also important for siRNA generation from high-copy transgenes. It is thus conceivable that the transcript maturation surveillance mechanism serves in genome defense beyond DNA double-strand breaks.

## Results

### A genome-wide RNAi screen for triggers of DNA double-strand break derived siRNAs

We had previously described a reporter system that allows for a dual luciferase-based readout of DNA double-strand break derived siRNA activity [16]. In short, a linearized plasmid with either a truncated or an inverted coding sequence of *Renilla* luciferase is co-transfected with a mix of circular expression vectors for *Renilla* and firefly luciferase to control for transfection efficiency. The siRNAs generated from the linearized plasmid can then repress full-length *Renilla* luciferase expression *in trans*. When combined with prior experimental RNAi, this assay system has a high signal-to-noise ratio and can easily be scaled up. Note that the promoters in all reporter constructs contain an intron in the 5'-UTR, which precedes the *Renilla* or firefly luciferase CDS. We thus performed a genome-wide RNAi screen in *Drosophila* S2-cells to identify factors that are required for the generation of DNA double-strand break derived siRNAs (see S1 Fig and S2 Fig for details). Two independent biological replicates of the entire screen were performed and averaged.

After removing likely false positives, such as retracted gene models or genes that are not expressed in S2-cells, we selected a total of 142 positive and 66 negative candidates from the screen for further validation (S1 Fig and S1 Table). We re-screened the original dsRNA trigger of our candidates for a third biological replicate and then generated two independent, non-overlapping dsRNAs for each candidate to identify false positives due to off-target RNAi. Only those candidates that scored positive for at least two out of the three distinct RNAi triggers (= screened dsRNA and two validation constructs) were retained. We also counter-screened the entire set of candidates with a cell line where a GFP reporter is repressed by two perfect matches to miR-277 in its 3'-UTR; expression of this reporter is driven by the same promoter as the *Renilla* luciferase in the screen. Since miR-277 is processed by Dcr-1, then loaded into Ago2 via the Dcr-2/R2D2 complex, we could distinguish between core RNAi pathway

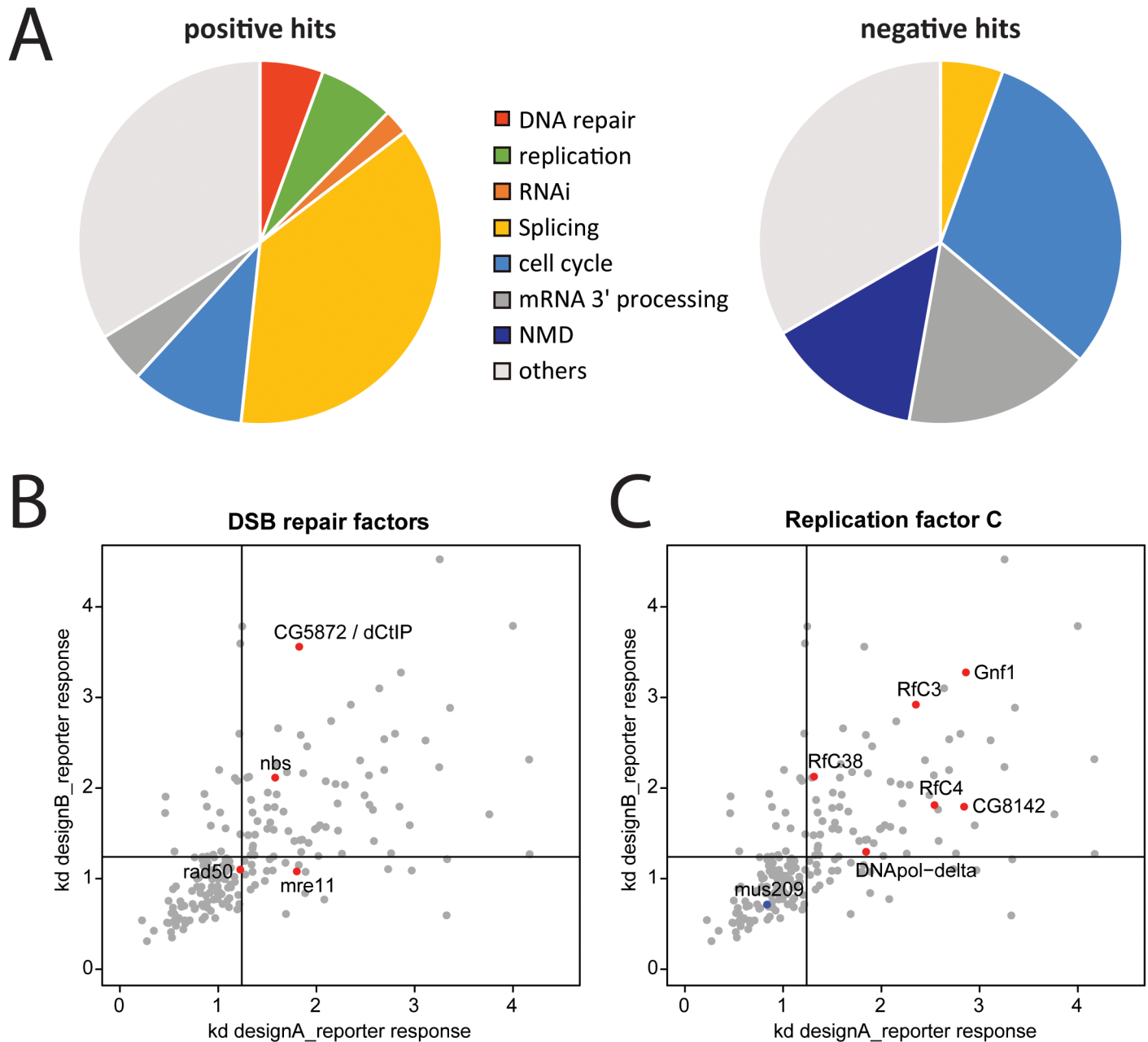
components or factors that non-specifically activate transcription of our reporter and those factors that are specifically required for DNA damage-induced siRNA biogenesis. After this stringent validation process (summarized in [S1 Table](#)), we retained a set of 89 genes that promote DSB-derived siRNA generation or function and 36 candidates that are potential repressors of DSB-derived siRNA production.

## Processing of the DNA ends stimulates siRNA generation

To obtain an initial overview of the biological processes involved in siRNA generation at the DNA break we performed a gene ontology analysis of the validated candidates ([Fig 1A](#)) and calculated significances using g:profiler [23]. As expected for a screen with linearized DNA, we identified a series of DNA replication/repair factors among the positive candidates (GO-term enrichment of “DNA metabolic process” with a  $p$ -value of  $9.0 \times 10^{-5}$ ). DNA double-strand breaks are recognized by the Mre11-Rad50-Nbs1 (MRN) complex; all components of this complex were among the initial candidates and 2 out of 3 passed our validation experiments. In addition to DNA damage signaling, the MRN complex initiates 5' to 3' resection of the break and thus commits the site for homology-directed repair. The 3' single-stranded end is subsequently extended further by CtIP/Sae2, promoting homologous recombination [24]. Bioinformatic analysis has suggested that the *Drosophila* CtIP homolog is CG5872 [25] and this gene was also required for correct re-localization of a heterochromatic DSB [26]. We identified CG5872 as a strong candidate in our screen; CG5872 is thus most likely the *Drosophila* homolog of CtIP/Sae2 and we propose that it should be called *dCtIP* ([Fig 1B](#)). For further repair, DNA synthesis carried out by the replicative polymerases DNA-pol $\delta$  and DNA-pol $\epsilon$  after Rad51-mediated annealing of the exposed 3' single-stranded regions with a homologous template. We identified DNA-pol $\delta$  and all subunits of replication factor C (RfC) as stimulators of break-induced siRNA generation. RfC normally loads the processivity clamp proliferating cell nuclear antigen (PCNA, called *mus209* in *Drosophila*) to ensure long-range DNA synthesis. However, the PCNA-homolog *mus209* was initially among the negative candidates but did not pass our validation criteria ([Fig 1C](#)). NHEJ factors, such as the Ku70/Ku80 complex, were not identified in the screen. Taken together, our screening efforts demonstrate that recognition and processing of the DNA double-strand break for homology-directed repair promotes and thus precedes siRNA generation.

## Splicing stimulates siRNA generation

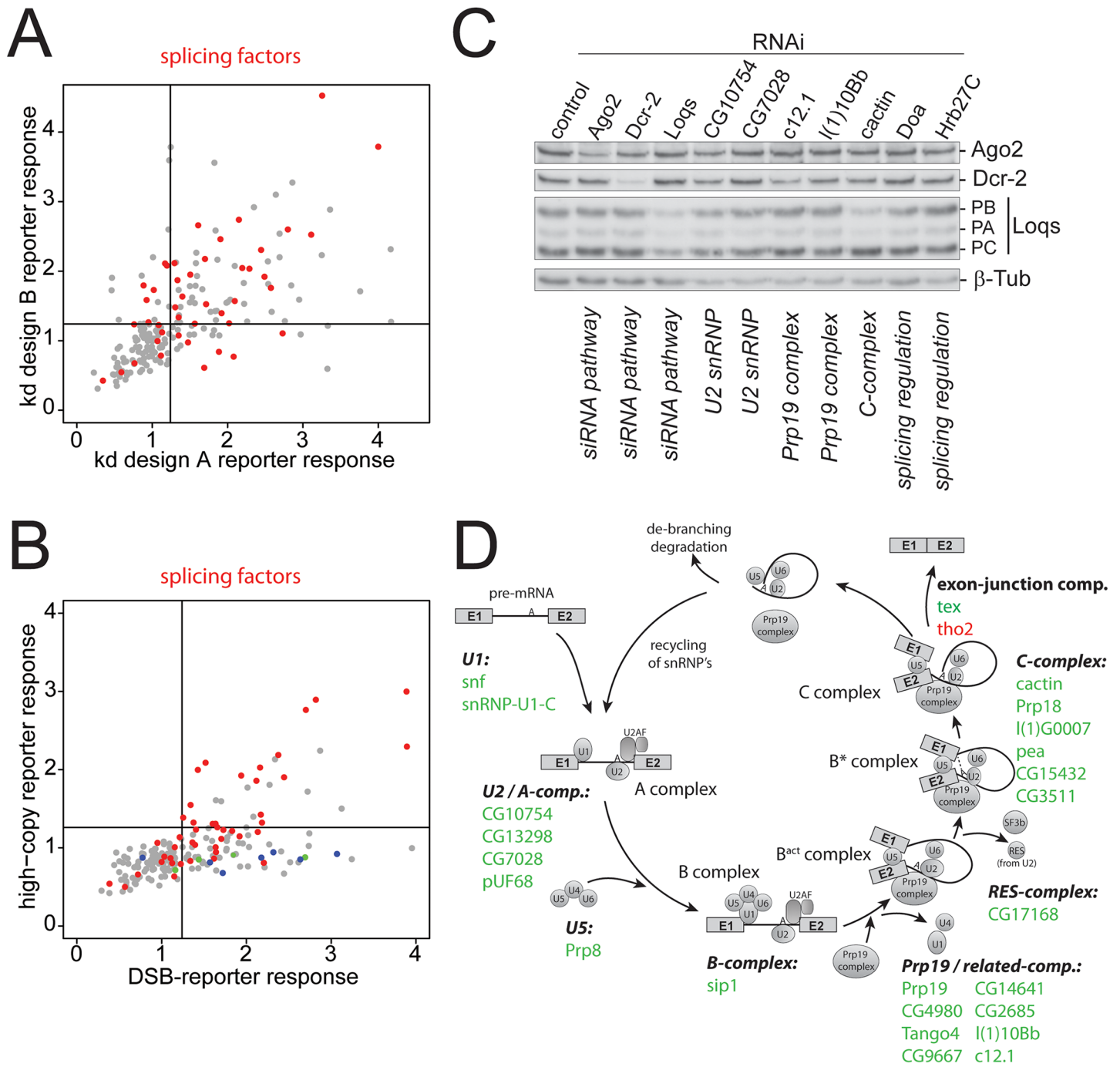
Much more striking than the recovery of the DNA repair factors, however, was the enrichment of splicing factors (see [Fig 1A](#), positive candidates). For example, the GO-term “RNA splicing” was found enriched among the candidates with a  $p$ -value of  $2 \times 10^{-29}$ . Among the potential repressors of DSB-derived siRNAs, we found that mRNA 3'-end processing activities were enriched (e.g. GO-term “mRNA cleavage”  $p$ -value  $4.7 \times 10^{-14}$ ). Although GO-term analyses must be interpreted with caution, this overview is consistent with the hypothesis that DSB-derived siRNAs are generated with a contribution of the mRNA splicing reaction. On the other hand, canonical transcript termination via mRNA cleavage appears to remove the trigger for siRNA generation. The candidates with the GO-term association “splicing” showed a validation success rate comparable to other candidates ([Fig 2A](#)). Since the involvement of splicing is reminiscent of a recently proposed model for transposon recognition in the fungus *Cryptococcus neoformans* [27], we tested our candidates for their requirement to repress a genomically integrated, endo-siRNA generating high-copy transgene analogous to a previously published system [28]. In the current study, we measured a cell line where the high-copy integrated *Renilla* luciferase responds more strongly to an impaired RNAi pathway than the firefly



**Fig 1. Analysis of the screening data for siRNA induction at a DNA double-strand break.** (A) Pie-charts of selected gene ontology terms among the 89 positive and 36 negative candidates obtained after screening and validation; the genes were analyzed with GOrilla [55]. (B) Specific labeling of the MRE11-Rad50-Nbs1 (MRN) complex in the context of the validation dataset. The putative *Drosophila* CtIP homolog, CG5872, was also recovered and confirmed in our screen. (C) Specific labeling of the replication factor C (Rfc) subunits and DNA polymerase  $\delta$  in the context of the validation dataset. Note that the *Drosophila* PCNA homolog, *mus209*, could not be confirmed as a negative hit in our validation experiments. For (A) and (C), the threshold lines were set at the mean plus 1x standard deviation of our negative controls (knock-down of GFP/DsRed).

<https://doi.org/10.1371/journal.pgen.1006861.g001>

luciferase integrated at low copy-number; since identical plasmid constructs were used this reporter system allows for direct comparison with the screening data (Fig 2B). This comparison defined two groups of candidates: The first is required for the generation of DNA break induced as well as high-copy transgene induced siRNAs and comprises, among others, the splicing factors. The second group is specific for DNA double-strand break induced siRNAs



**Fig 2. Analysis of splicing factors recovered in the screen.** (A) All factors with the GO-term annotation “splicing” were labeled in red in the context of the validation dataset. These factors had a similar distribution with respect to effect strength and validation success as other candidates. (B) Direct comparison of the effect strength obtained with splicing factors on DNA DSB induced siRNAs (abscissa) and high-copy transgene induced siRNAs (ordinate). Splicing factors are marked in red, while DNA repair factors and RfC components (see Fig 1) are marked in green and blue, respectively. The full dataset is contained in S1 Table. (C) Western Blot analysis to determine whether knock-down of splicing factors leads to impaired expression of core RNAi components. The knock-down is indicated above each lane, the monoclonal antibody used for detection is indicated on the right. All RNAi antibodies were kind gifts of Dr. Mikiko Siomi, the  $\beta$ -tubulin loading control was obtained from the developmental studies hybridoma bank. (D) A map of all spliceosome components identified in the screen onto the various complexes that assemble along the course of a splicing reaction. See also S3 Fig for more detail.

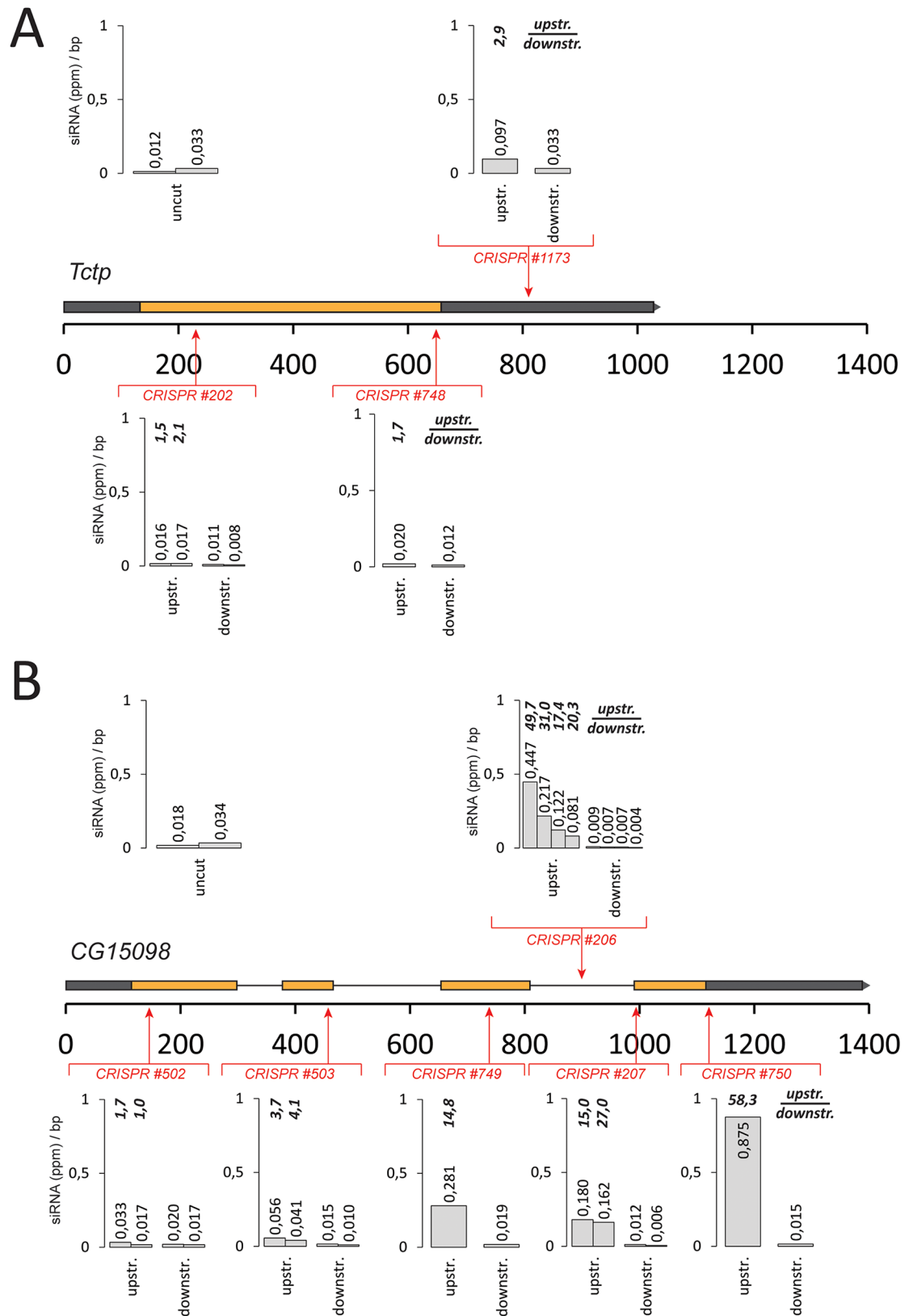
<https://doi.org/10.1371/journal.pgen.1006861.g002>

and comprises the homologous recombination factors discussed above (including dCtIP) as well as RfC. Perturbed mRNA splicing may thus be a common trigger for siRNA biogenesis at DNA double-strand breaks as well as high-copy transgenes.

On the other hand, RNAi-mediated depletion of splicing factors could indirectly affect siRNA biogenesis; for example, altered splicing efficiencies could result in lower protein levels of core RNAi factors. Our analysis with the miR-277 perfect match reporter cell line suggested that upon knockdown of the splicing factors, the core RNAi pathway is unaffected and reporter expression is unchanged (S1 Table). In addition, we tested several candidates involved in the splicing reaction for de-regulation of RNAi factors via Western blot and saw no consistent protein level changes for core RNAi components (Fig 2C). We constructed a map of our candidates on the various spliceosome complexes that assemble along the path of a splicing reaction (Fig 2D). Although we identified factors from all complexes, the recovery was particularly prominent among members of the Prp19- and Prp19-related complexes (8 out of 16 vs. 26 out of 138 for all spliceosome components,  $p < 0.04$   $\chi^2$ -test, see also S3 Fig). This complex has a pivotal role in enabling the transitions from the pre-catalytic spliceosome into the catalytic phases.

We wanted to confirm the importance of splicing for break-induced siRNA generation by creating DNA breaks at specific positions relative to splice sites. To this end, we employed the *cas9*-CRISPR nuclease technology and programmed the enzyme to cleave chromosomal sites before or after an intron, then deep sequenced the small RNAs and mapped them back to the cleaved locus. To obtain a quantitative measure of siRNA generation at a given site, we calculated the normalized read density per base pair of the affected transcript upstream and downstream of the cleavage site. If break-derived siRNAs are efficiently generated, then the average reads/bp values are higher between the promoter and the DSB site than in the region following the break [16]. We first targeted a strongly expressed, intron-less gene (*tctp*) for cleavage at three different positions. Although cleavage at each site was detectable (as judged by a T7 endonuclease assay, S4 Fig), we observed siRNA generation that was only slightly above background (average ratios of before vs. after the cut of 1.6–2.9, Fig 3A). Analogously, when we targeted the spliced gene CG15098 (with a similar expression level) before the first intron, we observed only low levels of siRNA generation (Fig 3B). A cut close to the first intron-exon junction (82 nt downstream) resulted in rather moderate siRNA induction. Cleavage further downstream, however, led to a strong generation of siRNAs with an increasing ratio of reads upstream vs. downstream of the break (up to 58.3) as the cut was moved downstream along the gene (Fig 3B, see also S5 Fig and S6 Fig for detailed traces). Thus, splicing of the transcript affected by the DNA break greatly stimulates siRNA generation.

We extended our analysis to the CG18273 gene, which is only moderately expressed in S2-cells (S7 Fig). Upon *cas9*-mediated DSB induction, an siRNA response could be induced here as well. Interestingly, this gene showed moderate, cleavage-independent siRNA coverage in its 3'-portion. Prior to this zone, the DNA-break induced coverage was about 10-fold lower than the coverages we observed in the cleaved CG15098 gene. The strength of the DSB-induced siRNA response correlates thus with the host gene expression level, consistent with the notion that the mRNA transcript contributes the sense strand to dsRNA formation. CG18273 has a short (60 nt) first intron followed by a rather long second exon (2375 nt). Cleavage within this second exon resulted in siRNA formation, indicating that a single short intron can suffice to trigger the response. Furthermore, when we induced DNA cleavage close to the end of the CG18273 gene (4686 nt downstream of the transcription start site), we observed siRNA coverage all the way to the start of the transcriptional unit. The DSB-induced small RNA response can thus cover a window of several kbp even in moderately expressed regions.



**Fig 3. Correlation between gene structure and efficiency of DSB-induced siRNA formation.** We calculated the normalized read frequency (reads per million) and determined the per base average for the region between the



transcription start site up to the cut (“upstream”, excluding the region where reads derived from the sgRNA itself map) and for the region between the cut until the annotated transcript 3'-end (“downstream”). Detailed traces of siRNAs mapping to the respective loci are provided in S5 Fig (TCTP) and 6 (CG15098). (A) cas9-CRISPR cuts in the intronless *tctp* gene result in barely any production of siRNAs. (B) Cuts in the intron-containing gene *CG15098* can efficiently produce DNA-damage dependent siRNAs, provided that the cut is sufficiently downstream of an intron.

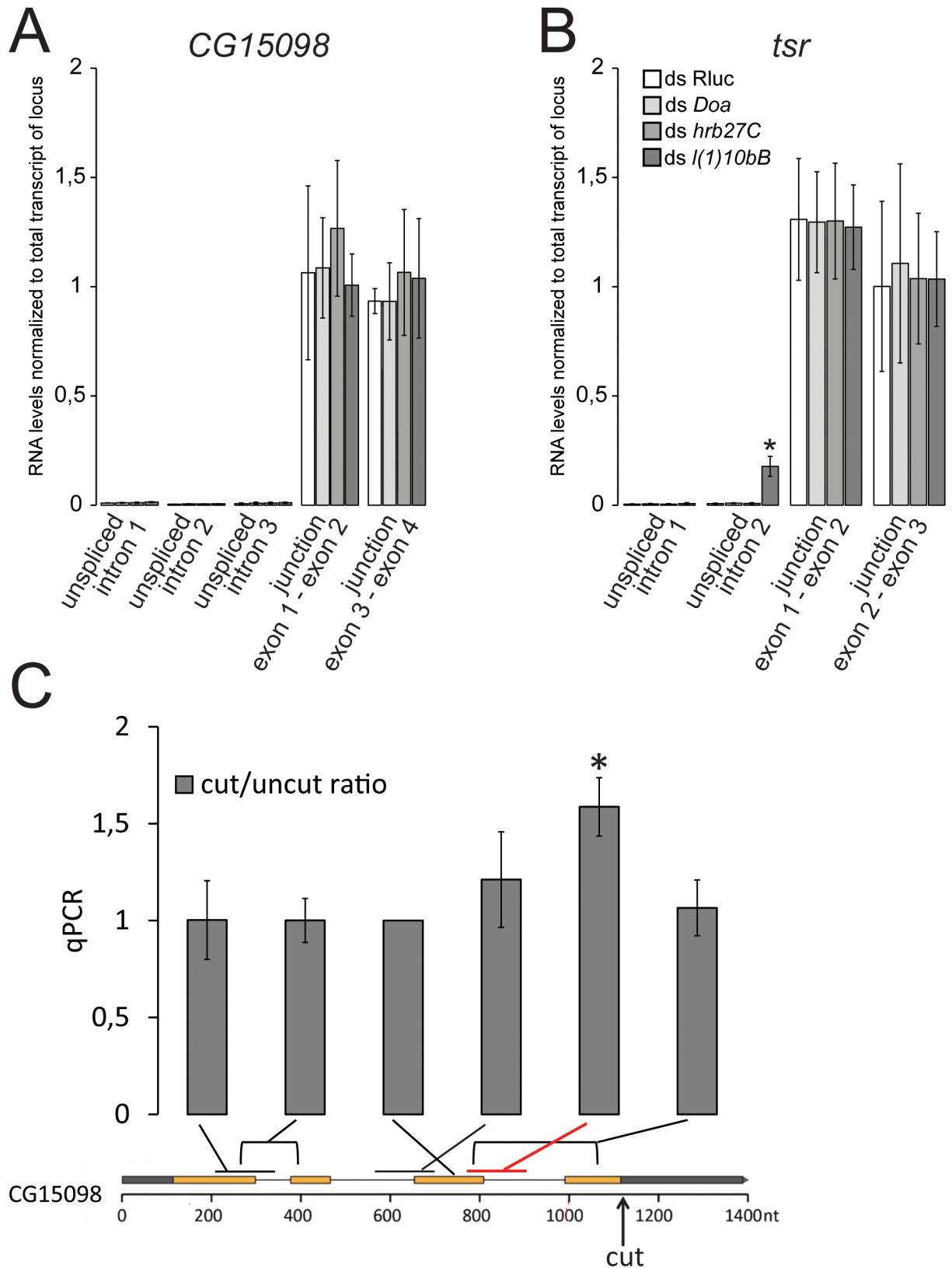
<https://doi.org/10.1371/journal.pgen.1006861.g003>

Since splicing is required to trigger the DNA-damage dependent siRNA response, we tested whether a knockdown of splicing factors identified in our screen simply reduces mRNA splicing and thereby diminishes the trigger for siRNA production. We thus determined splicing efficiencies at our CG15098 model locus and the *tsr* gene (both show strong expression in our S2 cells) following depletion of candidates recovered in our screen. After knockdown of the SR protein kinase *Doa*, the hnRNP protein *hrb27C* and the spliceosome component *l(1)10Bb* (the Bud31 homolog from the Prp19-related complex), we isolated total RNA and used qRT-PCR (random hexamer primed) to quantify the levels of unspliced pre-mRNA and spliced mRNA (Fig 4). The values were normalized to an amplicon that was internal to one of the exons and thus reported on the total amount of transcript from each locus (i.e. spliced and unspliced message). Overall, we found that the unspliced pre-mRNA did not increase relative to a control knockdown (*Renilla* luciferase). There was one exception, however: At the short intron in *tsr* the pre-mRNA became more abundant after RNAi against the spliceosome component *l(1)10Bb*. Even in this case, though, the amount of spliced message was comparable to the total amount of transcript produced, indicating that the majority of transcripts are correctly spliced. This was also seen for all other cases where we compared the level of spliced exon-exon junctions to the total amount of message, indicating that the canonically spliced mRNA accounts for essentially all of the transcripts detected at steady-state. In summary, we conclude that during the time of our knockdown-experiments there is no major change in general splicing efficiency. This argues that upon our experimental RNAi, the splicing factors and spliceosome components we identified did not yet induce major changes in mature mRNA levels. Rather than influencing the general cellular protein content, they may thus limit the signaling events that emanate from splicing reactions / spliceosomes perturbed by a nearby DNA break.

We used qRT-PCR to directly test whether a downstream DNA break perturbs progression of the splicing reaction. We induced cas9-mediated DNA cleavage downstream of the third intron of the CG15098 gene, then isolated total RNA and reverse transcribed both nascent and mature mRNA with random hexamer primers. We then interrogated nascent RNA with primers covering an exon-intron junction and mature mRNA with primers spanning an exon-exon junction. The samples were normalized to total CG15098 levels with an amplicon located inside of exon 3 as described above. Control samples were analyzed analogously and we calculated the cut/uncut ratio of each of the amplicons (Fig 4C). Indeed, we found significantly more unspliced RNA at the exon3-intron3 junction when the DNA was cut (student's t-test  $p < 0.01$ , 3 biological replicates). A downstream DNA break thus has the potential to stall progression of the splicing reaction at least transiently.

### At what stage is the splicing reaction stalled before double-stranded RNA is generated?

Since siRNA generation at DNA breaks depends on transcription levels in *Drosophila*, the sense strand of the siRNA precursor is most likely the normal transcript originating at the locus. Because upstream introns stimulate siRNA generation (Fig 3), we asked whether the splicing reaction takes place on the transcript molecule that contributes to the siRNA precursor. To this end, we analyzed the small RNA sequencing data from our cuts in the CG15098



**Fig 4. Knock-down of the screened splicing factors does not lead to a major decrease in overall splicing efficiency.** (A) Following knock-down of a selected set of identified splicing factors, the levels of unspliced pre-mRNA were determined with

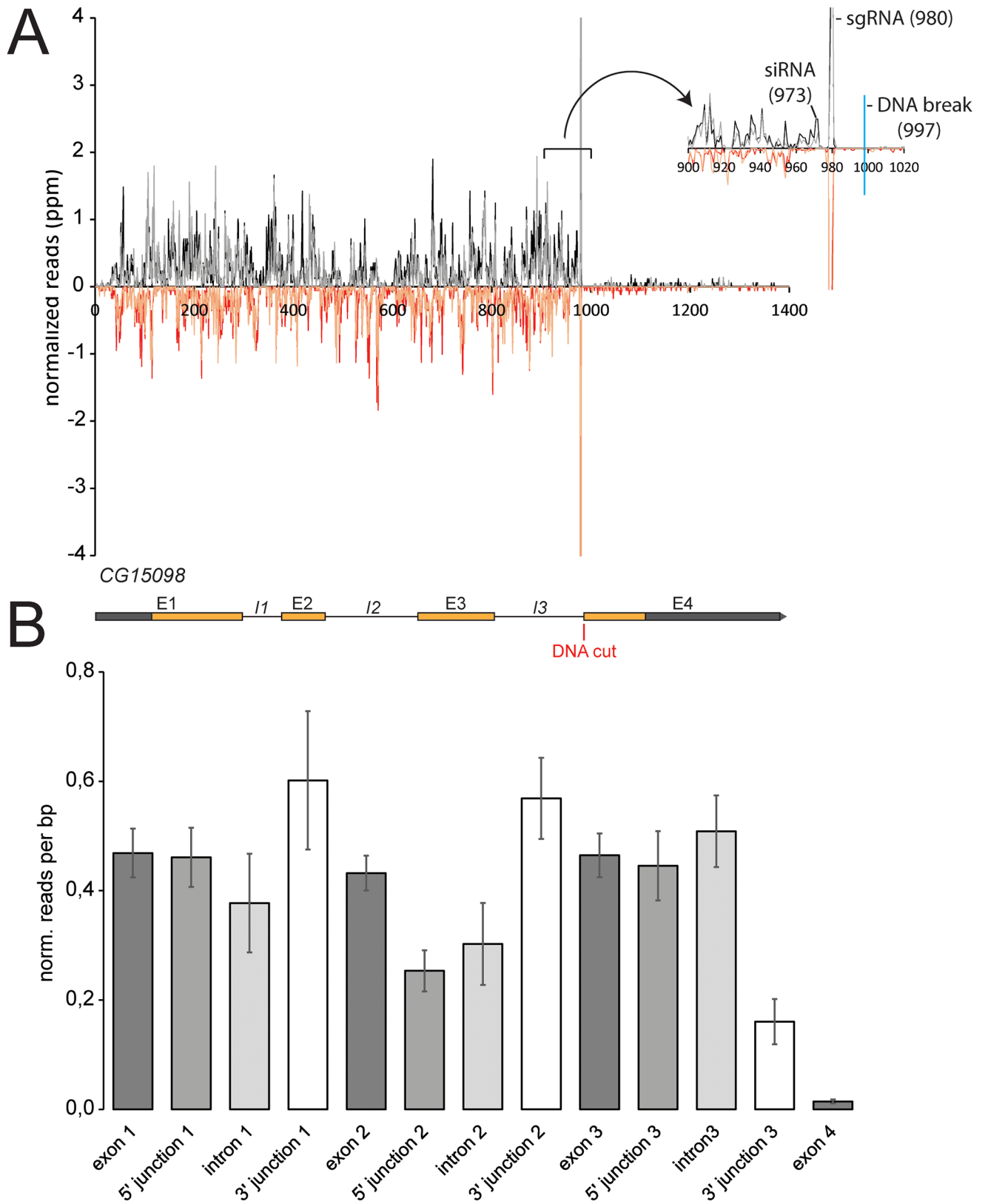
amplicons that contained the intron-exon junction, the spliced messages were specifically amplified with amplicons where one primer spanned the exon-exon junction. An amplicon confined within an exon was used to normalize to the total amount of transcript produced at the locus. The values shown are the mean  $\pm$  SD of three biological replicates. No increase of pre-mRNA levels was detected at any of the three introns in our CG15098 model locus and the levels of spliced transcript were comparable with the total transcript levels. (B) For the second intron of the *tsr* gene, we detected an increase of unspliced pre-mRNA after knock-down of *l(1)10Bb*. Nonetheless, the levels of spliced message were comparable to the total transcript levels, indicating that even in this case the majority of transcripts are still spliced. \*  $p = 0.023$ , Student's t-test (two-sided,  $n = 3$  biological replicates) (C) Induction of a downstream DNA break slows CG15098 transcript maturation. We induced cleavage at the CG15098 locus or in the TCTP gene as a control, isolated total RNA and performed qPCR analysis with random primed cDNA. The nascent RNA levels are much lower than those of mature CG15098 (see part A), but calculating the ratio of cut vs. uncut samples displays only the DNA-break induced changes. We detected a ~1.6-fold increase of nascent RNA directly upstream of the break ( $p < 0.01$ , student's t-test, 3 biological replicates) and there may be a trend towards slightly increased levels of nascent RNA further upstream at the second intron. We did not observe any changes in the level of mature CG15098 mRNA, which was to be expected given its far greater abundance and presumably slower turnover.

<https://doi.org/10.1371/journal.pgen.1006861.g004>

gene in detail. The siRNA coverage started essentially adjacent to the induced breaks and continued relatively uniformly until the beginning of the CG15098 transcription start site (see e.g. S6 Fig). The pattern of siRNA coverage was astonishingly reproducible (Fig 5A), but there was no strong correlation between gene structure and local siRNA coverage (Fig 5B, calculated based on four biological replicates). Intron 2, but not intron 1 or 3, showed a somewhat lower coverage than the exons 1–3 ( $p = 0.016$ , two-sided Student's t-test,  $n = 4$  biological replicates). Exon-exon junction spanning siRNAs were essentially absent, consistent with the notion that the siRNAs are not generated by RdRP-activities acting on mature mRNA ( $p < 10^{-4}$ , S8 Fig). Furthermore, we found that the coverage of the 3' intron-exon junctions—the site of the second transesterification reaction—was not diminished; rather, there was a trend towards slightly enhanced coverage relative to exonic reads (3'-junction 1:  $p = 0.11$ , 3'-junction 2:  $p = 0.06$ ). Clearly, the sense transcript is not fully spliced prior to siRNA generation. If the splicing reaction is stalled after the first catalytic step, then siRNAs covering the 5' exon-intron junction should be diminished. There was no change for intron 1 and 3 of CG15098. For intron 2 we observed a reduction of 5' exon-intron spanning reads when compared with the exonic coverage ( $p = 2 \times 10^{-4}$ ), but not when compared with the adjacent intron 2 coverage ( $p = 0.3$ ). The reduced coverage for intron 2 as well as its 5' splice junction remained when the DNA break was located at different positions relative to intron 2 (S9 Fig). We thus favor the interpretation that it is an inherent property of this intron (e.g. sequence-dependent). Because upstream splicing stimulates the siRNA response but the reaction is not completed, we conclude that the spliceosome is stalled most likely in a pre-catalytic state when dsRNA formation is triggered. This is also consistent with our qRT-PCR analysis (Fig 4C) where we saw an increase of nascent RNA as detected by an amplicon spanning the exon-intron junction, i.e. the site of the first catalytic step, after DNA cleavage.

## Discussion

The precursor of siRNAs in *Drosophila* is double-stranded RNA (dsRNA), which must be generated through convergent transcription since flies lack an RNA-dependent RNA polymerase. It was previously proposed that the DNA end serves as an initiation site for transcription that produces antisense RNA, followed by pairing with the normal transcript to generate dsRNA [16]. Our screening and validation experiments demonstrate that DNA end processing by the MRN-complex stimulates siRNA generation. Thus, recognition of the damage is independent of, and can precede, siRNA biogenesis. This is consistent with the observation that DNA damage signaling occurs normally in the absence of DNA damage induced siRNAs [14] and with the finding that DNA repair is unaffected when these siRNAs can no longer be made [22]. Our identification of the *Drosophila* CtIP homolog CG5872, a nuclease, in the screen further



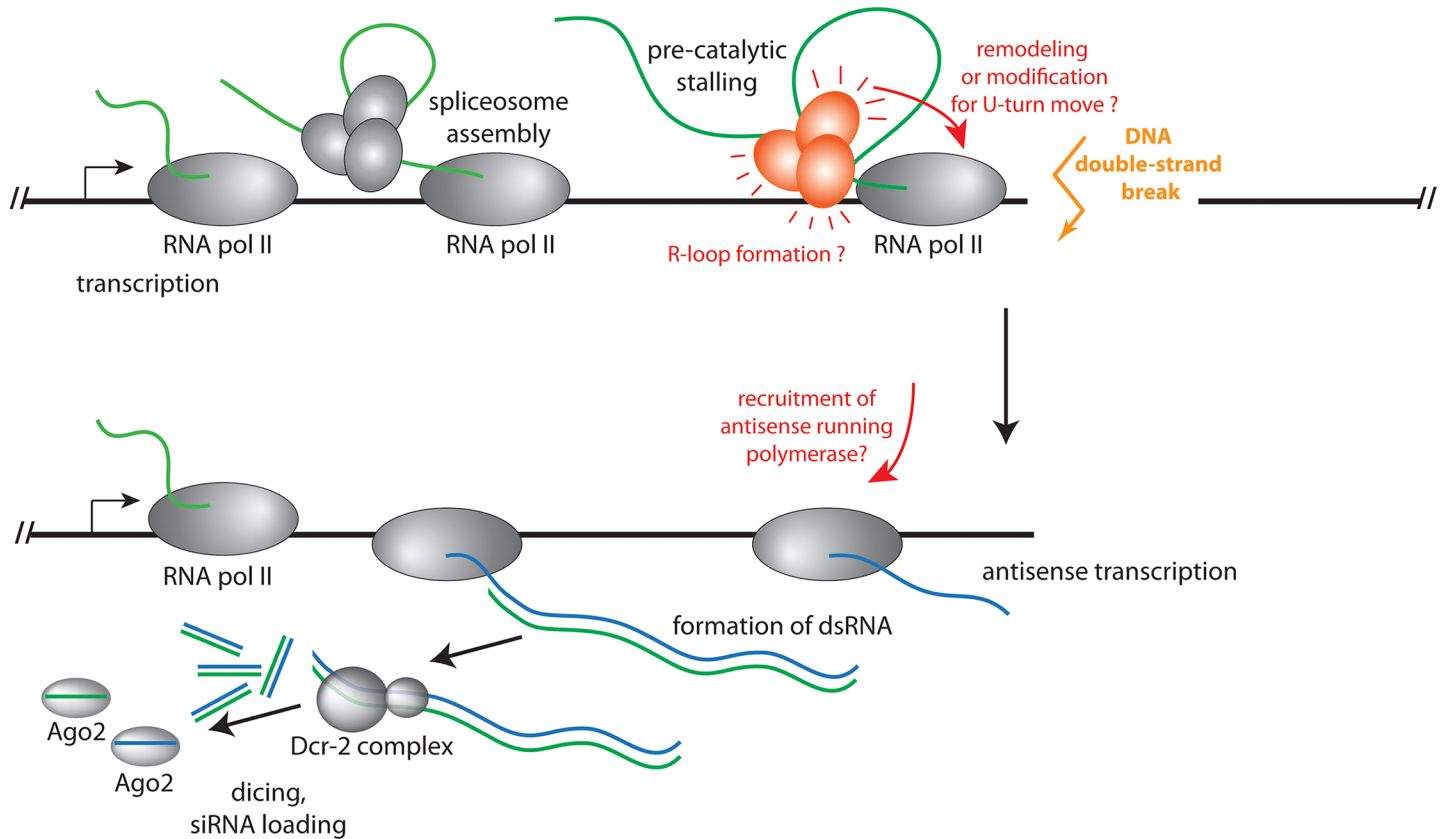
**Fig 5. Detailed analysis of the siRNA coverage along the CG15098 gene.** A set of four biological replicates was used to determine siRNA coverage along the gene. (A) Overall, the siRNA coverage was astonishingly reproducible. This diagram shows an overlay of two independent

sequencing experiments (experiment 1 colored in red and black, experiment 2 colored in orange and gray), the traces indicate the position of the 5'-end for sense-oriented siRNAs and of the 3'-end for antisense-oriented siRNAs. The read coverage thus extends 21 nt downstream of the indicated position. A close-up view of the two traces is shown in the inset. (B) The reads covering individual exons, introns and the corresponding junctions were normalized to the size of the respective feature, then the average  $\pm$  siRNA.

<https://doi.org/10.1371/journal.pgen.1006861.g005>

demonstrate that the generation of a 3'-single stranded DNA overhang facilitates the initiation of antisense transcription. Yet, siRNA coverage started adjacent to the break site, arguing that nucleolytic processing of the 5'-strand is not overly extensive. Based on our screening efforts we conclude that spliceosome components are required to trigger an siRNA response and cas9/CRISPR mediated cleavage in the genome revealed that an intron upstream of the break stimulates siRNA generation. The simplest interpretation is that spliceosomes participate in triggering this response independently of the splicing reaction. Consistently, we did not observe major splicing defects during the course of the knock-down experiments (Fig 4A and 4B), we could demonstrate that a downstream cut does slow down transcript maturation (Fig 4C) and the siRNA coverage analysis argued for a pre-catalytic stalling event (Fig 5).

What is the mechanism of dsRNA generation at the break? A straightforward hypothesis is that upon reaching the broken DNA end, the transcribing RNA polymerase II simply turns around and continues transcription of the other DNA template strand, thus forming a hairpin transcript ("U-turn move"). This phenomenon is well known when DNA templates with protruding 3'-ends are transcribed by bacteriophage T7 RNA polymerase [29]. Similarly, free RNA polymerases could spontaneously initiate transcription at the newly formed DNA end; this is also readily observed *in vitro* with DNA templates that bear a 3' single-stranded extension and was proposed to occur at DNA double-strand breaks in *S. pombe* [12]. In both cases, however, one would not predict that the presence of an intron should stimulate the generation of dsRNA; rather, DNA breaks in the intronless gene should have led to a comparable extent of siRNA generation as the ones in the intron-containing gene. Consistently, we had observed that transfected linear PCR products do not trigger an siRNA response [16]. Thus, while we cannot exclude that U-turn transcripts are formed, it is unlikely that they are the source of the majority of the siRNA precursors. It remains possible, however, that association of the spliceosome with the nascent mRNP leads to a remodeling or modification of the RNA polymerase complex, which favors the execution of a U-turn at the DNA end (see Fig 6). A formal possibility is that non-canonical enzymes are recruited to serve as RNA-dependent RNA polymerases (RdRP) acting on the transcript affected by the DNA break. For example, it has been demonstrated that RNA polymerase II can use an RNA template to create a corresponding RNA transcript in the case of human hepatitis delta virus or plant viroid replication [30–33] and that bacteriophage T7 RNA polymerase can replicate short RNA templates [34]. However, this has only been observed for RNAs with a particular secondary structure, while the DNA damage-induced siRNA response appears to be generic. Since we did not detect any exon-exon junction spanning reads in our siRNA coverage analysis, any RdRP-like activity would be limited to the unspliced, nascent transcript. Taken together, we do not consider this a likely scenario. The most parsimonious hypothesis is that RNA polymerase stalls at the break, co-transcriptional mRNA maturation is concomitantly delayed and that this induces a signaling event with participation of the spliceosome (Fig 6). Such a stalled transcript probably leads to a persistent R-loop with a corresponding displaced, single-stranded DNA region. This single-stranded DNA, together with a signal from the spliceosome, could serve as an initiation site for antisense transcription. It is also conceivable that such R-loops may be larger or more persistent if the stalling occurs in the vicinity of an engaged spliceosome without the need for a specific signaling step. This mechanism may also act when R-loops occur independently of a DNA break,



**Fig 6. Model for the sequence of events leading to DNA DSB-induced siRNA formation in *Drosophila*.** We propose that when the RNA polymerase reaches the DNA double-strand break, the co-transcriptional splicing process is adversely affected. This leads to a signaling event that emanates from the pre-catalytically stalled spliceosome, potentially augmented by formation of an R-loop, or a remodeling/modification of the RNA polymerase complex to enable a “U-turn” move and the synthesis of a long RNA hairpin. In both cases, double-stranded RNA is eventually produced when the normal mRNA transcript base-pairs with the break-induced antisense transcript. This double-stranded RNA is then converted into siRNAs by Dcr-2 and loaded into Ago2.

<https://doi.org/10.1371/journal.pgen.1006861.g006>

consistent with the observation that the splicing factors we identified were also required for the small RNA response triggered by high-copy transgenes.

Several previous publications have reported a requirement of splicing factors, but not splicing in general, for small RNA-mediated transcriptional silencing in fission yeast [35, 36]. Consistent with this, centromeric non-coding transcripts are indeed spliced in *S. pombe* [36]. However, messages coding for essential fission yeast silencing factors appear to be particularly sensitive to diminished splicing activity, thus leading to reduced silencing efficiency; a similar phenomenon can affect the expression of the *Drosophila melanogaster* piRNA factor *piwi* [37, 38]. It remains a matter of debate whether intron-less, cDNA-based rescue constructs can bypass the silencing defect induced by perturbed splicing in *S. pombe* [36, 37]. Other splicing factors, such as *smD1*, appear to function independently of their splicing activity during miRNA and siRNA RISC formation [39, 40]. We now demonstrate that for DNA double-strand break triggered siRNAs, an intron is required upstream (with respect to transcription) of the lesion to trigger siRNA formation (Fig 3). This strongly supports the notion that the splicing process acts in *cis* during siRNA generation rather than *in trans* via perturbed splicing of silencing factor mRNAs.

Based on our siRNA coverage and qRT-PCR analysis, we propose that the spliceosome is stalled at a pre-catalytic stage (see the model in Fig 6). The Prp19 complex promotes the

transition into the catalytic splicing phases and members of this complex appeared enriched among the spliceosome components we identified (Fig 2D and S3 Fig). Prp19 is of central importance for the splicing-mediated identification of transposon transcripts in *C. neoformans* [27], but in this case the reaction was stalled after the first catalytic step. The *Drosophila* response we describe is conceptually similar to the recently discovered spliceosome-mediated decay in the budding yeast *Saccharomyces cerevisiae*. There, nucleases are recruited to intronless genes as a consequence of non-productive association with the spliceosome [41]. Splicing controls piRNA biogenesis in fruit flies; this is a class of small RNAs that represses—together with endo-siRNAs—transposable elements in the germline. Here, spliced transcripts from the so-called master control loci are prevented from entering the piRNA biogenesis pathway [42]. On the other hand, the Tho/TREX complex, normally deposited on RNA as a consequence of splicing, is essential for piRNA biogenesis and must associate through an alternative route with unspliced piRNA precursors [43]. Interestingly, we identified the Tho complex component *tho2* as a potential inhibitor of both, DNA damage-induced siRNAs as well as high-copy transgene induced siRNAs (S1 Table).

Induction of DNA damage by UV-light revealed that in human cells, splicing is both a sensor for as well as a target of the DNA damage response. This depends on the formation of R-loops due to stalled polymerases and results in pleiotropic splicing changes. The phenomenon bears many parallels to our analysis, but it was limited to transcription-blocking lesions and the authors specifically excluded DNA double-strand breaks as triggers [44]. A consequence of R-loop formation is the generation of a corresponding stretch of displaced, single-stranded DNA. Potentially, this DNA is covered by replication protein A (RPA), [45], a situation that can trigger DNA damage signaling via direct interaction with Prp19 and accumulation of ATR-interacting protein (ATRIP) in mammalian cells [13]. This role of Prp19 appears to be independent of its function during the splicing reaction. Future experiments should therefore address the question whether multiple recruitment platforms rely on the conversion of the Prp19 complex from a regulator of the splicing reaction into a trigger for DNA damage associated signaling events.

Due to the focus on siRNA biogenesis in our screen, rather than their downstream function, we cannot directly conclude on the benefits of the break-derived siRNAs for the organism. We previously demonstrated that *dcr-2* and *ago2* play—at best—an accessory role during DNA repair [22], but this only addresses the importance of the siRNAs and not their precursor molecules. It is possible that conversion of the stalled transcript into double-stranded RNA or the process of antisense transcription *per se* are important for DNA repair. For example, this could limit the extent of R-loop formation behind a stalled RNA polymerase (both temporally and spatially) or set the optimal length of RPA-covered single-stranded DNA, as was recently proposed for *Schizosaccharomyces pombe* [12]. Control of R-Loop size during transcription has also been described: In the centromeric regions of fission yeast, a specific arrangement of replication origins and repetitive elements leads to frequent collisions of an RNA polymerase and the replication machinery, followed by the generation of siRNAs in an RdRP-dependent process. In this case, RNA interference prevents excessive R-loop formation by releasing RNA polymerase II and thus fosters genome integrity [46]. Other publications have described a role for RNAi during DNA repair in other organisms [14, 15, 18, 19, 47]. In the case of the damage-induced *Neurospora crassa* qiRNAs, DNA repair was even proposed to be the trigger for small RNA production [19]. We note that in the case of our model Gene CG15098, no repetitive and thus recombination-favoring sequence arrangement was required. Clearly, more results are needed to delineate common and divergent features between these experimental systems. Since the splicing-dependent surveillance mechanism is also important for siRNA generation from high-copy transgenes (Fig 2B), it is possible that the DNA double-strand

breaks fortuitously trigger the splicing-dependent mRNA maturation check-point in *Drosophila melanogaster*, while its primary importance might be to act as a surveillance mechanism against foreign DNA sequences. The list of candidates we identified in our screen provides a valuable resource to further our understanding of the molecular mechanisms and the biological significance of this surveillance mechanism.

## Materials and methods

### Cell culture, transfections and cas9-CRISPR mediated induction of DNA double-strand breaks

Cells were cultured and transfected as described [48]. DNA double-strand breaks were induced in an S2-cell clone that stably expresses the cas9 nuclease [49] with linear U6-sgRNA fusion fragments generated by overlap extension PCR as described [50]. The primer sequences for sgRNA expression are provided in S2 Table. The cell line with stable high-copy integration of luciferase reporters was generated by transfecting S2-cells with a mix containing pRB1 (Renilla luc.) in a 10-fold excess over pRB2 (firefly luc.) and phsHygro. After selection of stable resistance the cells were diluted, clones were picked and clone 4 was used for the experiments.

### Genome-wide RNAi screen

A suspension of DMe1 cells (an S2-cell isolate) in Express5 SFM was prepared (final concentration of 25,000 cells/well) and 30  $\mu$ L cell suspension were added to each well of a Greiner 384-well plate containing 5  $\mu$ L of dsRNA (conc. 50ng/ $\mu$ L). The plates were sealed and incubated at 25 C for 48 h. On each plate of the screen, we included a set of positive controls (knock-down of Ago2, Dcr-2 and Ago1) as well as a set of negative controls (knock-down of GFP and DsRed). Furthermore, we included a knock-down of *Renilla* luciferase as a positive control for potential repressors of DSB-induced siRNA generation. Finally, we knocked down the apoptosis inhibitor *thread* to set thresholds for exclusion of data from predominantly dead cells.

For transfection, 0.15  $\mu$ L/well Fugene (Promega) and 6 ng/well of pRB1, 20 ng/well of pRB2 and 30 ng/well of linearized pRB3 were pre-diluted in 14.85  $\mu$ L/well Express5 SFM and incubated at room temperature for 30 min. 15  $\mu$ L of the thus prepared transfection solution were then dispensed into each well of the assay 384-well plates containing 35  $\mu$ L of cell suspension with dsRNA. The plates were again sealed and incubated at 25°C.

96 h after knockdown, cells were lysed and luminescence was measured. Fluc and RLuc buffers were added subsequently to the cell lysate and the signals were measured 5 min after the addition of the according substrate with no filter (Fluc) and a F485 Coelenterazin filter (RLuc).

The primary analysis of the FLuc and RLuc data that was obtained from luminescence measurements was done in R using the cellHTS2 package [51]. The raw data was log-transformed and normalized on the plate median, each channel was analyzed separately and no variance adjustment was applied. The mean RLuc values were plotted against the mean FLuc values and a smoothed curve was fitted to the data using locally weighted polynomial regression (LOESS) with a smoothing parameter of 0.9. The thus predicted LOESS values were subtracted from the corresponding mean RLuc values to obtain the LOESS-residuals (resi). The residuals were then used to calculate a z-score. A comprehensive overview of the screening data and controls is given in S2 Fig, the original screening data has been deposited at <http://www.genomernai.org>.



## Validation experiments

For validation, S2 cells (lab stock) were grown in FBS-supplemented Schneider's medium and the validation screens were conducted in 96-well plates with linearized pRB4 as siRNA inducer. The knockdown constructs used for validation are provided in [S1 Table](#).

The oligonucleotide sequences for the generation of the independent dsRNA constructs for validation are provided in [S1 Table](#). *In vitro* transcription and treatment of the cells was as previously described [16].

## Generation and analysis of next generation sequencing data for the characterization of DSB-induced siRNAs

A total volume of 3 ml of stable cas9-expressing cell culture was transfected with 1500 ng of U6-sgRNA PCR product as described [50]. Five days after transfection, total RNA was isolated using Trizol and deep sequencing libraries were generated [52]. In certain cases, two sgRNAs were introduced, leading to cas9-mediated cleavage in distinct genes. Up to four libraries were barcoded, combined and analyzed on one lane of an Illumina HiSeq instrument. Subsequently, the obtained reads were sorted, trimmed and mapped to the regions of interest using bowtie [53]. Further analysis was performed with in-house Perl scripts (available on request).

## Quantitative RT-PCR

Cells were treated with dsRNA to induce knock-down of the corresponding genes under identical conditions as for screening and validation. Five days after induction of knock-down, total RNA was isolated using Trizol and the RNA was reverse transcribed with random hexamer primers and the Superscript-III enzyme. For the analysis of splicing upstream of a DNA break, we transfected the corresponding sgRNA expression PCR products into our stable cas9-expressing cell line, then isolated total RNA 3 days after transfection. Quantitative PCR was carried out on a Biometra TOptical real-time PCR cycler with the Dynamo Flash SybrGreen PCR kit (Finnzymes). Data analysis was done according to the  $2^{-\Delta\Delta C_t}$  method [54], primer sequences for qPCR are given in [S2 Table](#).

## Supporting information

**S1 Fig. The figure depicts the workflow of our genome-wide screen for DNA double-strand break induced siRNA biogenesis.** Note that the promoters driving expression of firefly (tubulin) and *Renilla* (ubiquitin) luciferase both contain an intron in the 5'-UTR. Following the screen, stringent validations were performed with two independent sets of dsRNA triggers and only those candidates were retained that scored positive in two out of three dsRNA designs (bottom diagram).

(JPG)

**S2 Fig. This diagram gives a comprehensive overview of all the screening data (including the positive and negative controls).**

(JPG)

**S3 Fig. This figure presents a comprehensive map of the spliceosome components and complexes that assemble along the splicing reaction.** The assignment is based on the information contained in the spliceosome database (<http://spliceosomedb.ucsc.edu/proteins>) as of June 2016. Please note that assignment of individual genes to a specific complex may not always be unambiguous but had to be simplified for this diagram. Furthermore, this map is

not absolutely comprehensive as some simplification and consolidation was necessary to preserve clarity. The color code depicts the effect strength in the original screening data (Z-score, corresponding numerical values can be found in [S1 Table](#)). The effect strength for a knock-down of Dcr-2 is shown as a reference below the color key.

(JPG)

**S4 Fig. Assessment of cleavage and mutagenesis at the cas9-CRISPR target sites.** We performed a T7 endonuclease assay on PCR products obtained from DNA isolated alongside with the RNA for our deep sequencing experiments. Processing by T7 endonuclease at the intended position is indicated with an asterisk and demonstrates that the corresponding site had been cleaved and subject to mutagenic repair *in vivo*. The results are grouped according to cleavage within CG15098 (A), TCTP (B) or CG18273 (C).

(JPG)

**S5 Fig. Detailed traces of siRNA reads mapping to the intronless *tctp* gene.** Examples for the two positions of cas9-CRISPR mediated cuts in the intronless *tctp* gene are depicted; the position of the cut site can be deduced from the large peak that derives from the targeting region of the sgRNA itself. Please refer to [Fig 3](#) of the manuscript for the correlation of cleavage site with gene structure.

(JPG)

**S6 Fig. Detailed traces of siRNA reads mapping to the intron-containing *CG15098* gene.** A selection of cas9-CRISPR mediated cuts are depicted; The position of the cut site can be deduced from the large peak that derives from the targeting region of the sgRNA itself. Please refer to [Fig 3](#) of the manuscript for the correlation of cleavage site with gene structure.

(JPG)

**S7 Fig. Detailed traces of siRNA reads mapping to the intron-containing *CG18273* gene.** The blue line indicates the position of the cas9-mediated cut. Note the scale change relative to the previous figure.

(JPG)

**S8 Fig. Analysis of exon-intron and exon-exon spanning siRNA coverage.** We used a dataset with four replicate experiments containing a cut at the end of the third intron in the CG15098 gene (CRISPR 207). All exons, introns and junctions that lie upstream of the cut were combined and an average  $\pm$  SD for the four replicates was calculated. In addition, we mapped the sequencing data to a CG15098 cDNA sequence, and then calculated the abundance of exon-exon junction spanning reads in our data.

(JPG)

**S9 Fig. Analysis of read coverage according to gene structure.** (A) For data derived from a cut within the 3'-UTR of CG15098 (CRISPR 750); (B) For data derived from a cut within the third exon (CRISPR 749).

(JPG)

**S1 Table. Detailed information on the candidate validation experiments.**

(XLSX)

**S2 Table. Primer sequences for sgRNA+T7 endonuclease assays and primer sequences for qPCR.**

(XLSX)

## Acknowledgments

We are indebted to Dr. Mikiko Siomi (Keio University School of Medicine, Japan) for sharing monoclonal antibodies with us. The  $\beta$ -tubulin antibody developed by Michael Klymkowsky was obtained from the Developmental Studies Hybridoma Bank, created by the NICHD of the NIH and maintained at The University of Iowa, Department of Biology, Iowa City, IA 52242.

## Author Contributions

**Conceptualization:** KF MBo KM.

**Data curation:** MBr SK HB KF.

**Formal analysis:** KM KF MBr SK.

**Funding acquisition:** HB MBo KF.

**Investigation:** KM RB.

**Methodology:** MBr KM HB SK KF MBo.

**Project administration:** MBo KF.

**Resources:** MBo SK HB.

**Software:** KF KM.

**Supervision:** MBr KF.

**Validation:** KM KF.

**Visualization:** KM KF.

**Writing – original draft:** KM KF.

**Writing – review & editing:** KM SK RB MBr MBo SK HB.

## References

1. Symington LS, Gautier J. Double-strand break end resection and repair pathway choice. *Annu Rev Genet.* 2011; 45:247–71. Epub 2011/09/14. <https://doi.org/10.1146/annurev-genet-110410-132435> PMID: 21910633.
2. Miyaoka Y, Berman JR, Cooper SB, Mayerl SJ, Chan AH, Zhang B, et al. Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Sci Rep.* 2016; 6:23549. Epub 2016/04/01. <https://doi.org/10.1038/srep23549> PMID: 27030102; PubMed Central PMCID: PMC4814844.
3. Richardson CD, Ray GJ, DeWitt MA, Curie GL, Corn JE. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nature biotechnology.* 2016; 34(3):339–44. Epub 2016/01/21. <https://doi.org/10.1038/nbt.3481> PMID: 26789497.
4. Shkreta L, Chabot B. The RNA Splicing Response to DNA Damage. *Biomolecules.* 2015; 5(4):2935–77. Epub 2015/11/04. <https://doi.org/10.3390/biom5042935> PMID: 26529031; PubMed Central PMCID: PMC4693264.
5. Wickramasinghe VO, Venkitaraman AR. RNA Processing and Genome Stability: Cause and Consequence. *Mol Cell.* 2016; 61(4):496–505. Epub 2016/02/20. <https://doi.org/10.1016/j.molcel.2016.02.001> PMID: 26895423.
6. Sollier J, Stork CT, Garcia-Rubio ML, Paulsen RD, Aguilera A, Cimprich KA. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol Cell.* 2014; 56(6):777–85. Epub 2014/12/02. <https://doi.org/10.1016/j.molcel.2014.10.020> PMID: 25435140; PubMed Central PMCID: PMC4272638.

7. Hamperl S, Cimprich KA. The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. *DNA Repair (Amst)*. 2014; 19:84–94. Epub 2014/04/22. <https://doi.org/10.1016/j.dnarep.2014.03.023> PMID: 24746923; PubMed Central PMCID: PMC4051866.
8. Pankotai T, Soutoglou E. Double strand breaks: hurdles for RNA polymerase II transcription? *Transcription*. 2013; 4(1):34–8. PMID: 23340208. <https://doi.org/10.4161/trns.22879>
9. Gaillard H, Aguilera A. Transcription coupled repair at the interface between transcription elongation and mRNP biogenesis. *Biochimica et biophysica acta*. 2013; 1829(1):141–50. Epub 2012/10/11. <https://doi.org/10.1016/j.bbagr.2012.09.008> PMID: 23046879.
10. Vermeulen W, Fouteri M. Mammalian transcription-coupled excision repair. *Cold Spring Harbor perspectives in biology*. 2013; 5(8):a012625. Epub 2013/08/03. <https://doi.org/10.1101/cshperspect.a012625> PMID: 23906714; PubMed Central PMCID: PMC3721277.
11. Ohsawa R, Seol JH, Tyler JK. At the intersection of non-coding transcription, DNA repair, chromatin structure, and cellular senescence. *Front Genet*. 2013; 4:136. Epub 2013/08/24. <https://doi.org/10.3389/fgene.2013.00136> PMID: 23967007; PubMed Central PMCID: PMC3744812.
12. Ohle C, Tesorero R, Schermann G, Dobrev N, Sinning I, Fischer T. Transient RNA-DNA Hybrids Are Required for Efficient Double-Strand Break Repair. *Cell*. 2016; 167(4):1001–13.e7. <https://doi.org/10.1016/j.cell.2016.10.001> PMID: 27881299
13. Marechal A, Li JM, Ji XY, Wu CS, Yazinski SA, Nguyen HD, et al. PRP19 Transforms into a Sensor of RPA-ssDNA after DNA Damage and Drives ATR Activation via a Ubiquitin-Mediated Circuitry. *Mol Cell*. 2014; 53(2):235–46. <https://doi.org/10.1016/j.molcel.2013.11.002> PMID: 24332808
14. Francia S, Michelini F, Saxena A, Tang D, de Hoon M, Anelli V, et al. Site-specific DICER and DROSHA RNA products control the DNA-damage response. *Nature*. 2012 Aug 9; 488(7410):231–5. <https://doi.org/10.1038/nature11179> PMID: 22722852
15. Wei W, Ba Z, Gao M, Wu Y, Ma Y, Amiard S, et al. A Role for Small RNAs in DNA Double-Strand Break Repair. *Cell*. 2012; 149(1):101–12. Epub 2012/03/27. <https://doi.org/10.1016/j.cell.2012.03.002> PMID: 22445173.
16. Michalik KM, Bottcher R, Forstemann K. A small RNA response at DNA ends in *Drosophila*. *Nucleic Acids Res*. 2012; 40(19):9596–603. Epub 2012/08/01. <https://doi.org/10.1093/nar/gks711> PMID: 22848104; PubMed Central PMCID: PMC3479179.
17. Lee HC, Chang SS, Choudhary S, Aalto AP, Maiti M, Bamford DH, et al. qiRNA is a new type of small interfering RNA induced by DNA damage. *Nature*. 2009; 459(7244):274–7. Epub 2009/05/16. <https://doi.org/10.1038/nature08041> PMID: 19444217; PubMed Central PMCID: PMC2859615.
18. Gao M, Wei W, Li MM, Wu YS, Ba Z, Jin KX, et al. Ago2 facilitates Rad51 recruitment and DNA double-strand break repair by homologous recombination. *Cell research*. 2014; 24(5):532–41. Epub 2014/03/26. <https://doi.org/10.1038/cr.2014.36> PMID: 24662483; PubMed Central PMCID: PMC4011338.
19. Zhang Z, Chang SS, Xue Z, Zhang H, Li S, Liu Y. Homologous recombination as a mechanism to recognize repetitive DNA sequences in an RNAi pathway. *Genes Dev*. 2013; 27(2):145–50. Epub 2013/01/17. <https://doi.org/10.1101/gad.209494.112> PMID: 23322299; PubMed Central PMCID: PMC3566306.
20. Zhang Z, Yang Q, Sun G, Chen S, He Q, Li S, et al. Histone H3K56 acetylation is required for quelling-induced small RNA production through its role in homologous recombination. *J Biol Chem*. 2014; 289(13):9365–71. <https://doi.org/10.1074/jbc.M113.528521> PMID: 24554705; PubMed Central PMCID: PMC3979385.
21. Yang Q, Ye QA, Liu Y. Mechanism of siRNA production from repetitive DNA. *Genes Dev*. 2015; 29(5):526–37. <https://doi.org/10.1101/gad.255828.114> PMID: 25691092; PubMed Central PMCID: PMC3979385.
22. Schmidts I, Bottcher R, Mirkovic-Hosle M, Forstemann K. Homology directed repair is unaffected by the absence of siRNAs in *Drosophila melanogaster*. *Nucleic Acids Res*. 2016. Epub 2016/06/30. <https://doi.org/10.1093/nar/gkw570> PMID: 27353331.
23. Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, et al. g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res*. 2016. Epub 2016/04/22. <https://doi.org/10.1093/nar/gkw199> PMID: 27098042.
24. Panier S, Durocher D. Push back to respond better: regulatory inhibition of the DNA double-strand break response. *Nat Rev Mol Cell Biol*. 2013; 14(10):661–72. Epub 2013/09/05. <https://doi.org/10.1038/nrm3659> PMID: 24002223.
25. Uanschou C, Siwiec T, Pedrosa-Harand A, Kerzendorfer C, Sanchez-Moran E, Novatchkova M, et al. A novel plant gene essential for meiosis is related to the human CtIP and the yeast COM1/SAE2 gene. *Embo J*. 2007; 26(24):5061–70. Epub 2007/11/17. <https://doi.org/10.1038/sj.emboj.7601913> PMID: 18007598; PubMed Central PMCID: PMC2140101.

26. Chiolo I, Minoda A, Colmenares SU, Polyzos A, Costes SV, Karpen GH. Double-strand breaks in heterochromatin move outside of a dynamic HP1a domain to complete recombinational repair. *Cell*. 2011; 144(5):732–44. Epub 2011/03/01. <https://doi.org/10.1016/j.cell.2011.02.012> PMID: 21353298; PubMed Central PMCID: PMC3417143.
27. Dumesic PA, Natarajan P, Chen C, Drinnenberg IA, Schiller BJ, Thompson J, et al. Stalled spliceosomes are a signal for RNAi-mediated genome defense. *Cell*. 2013; 152(5):957–68. Epub 2013/02/19. <https://doi.org/10.1016/j.cell.2013.01.046> PMID: 23415457; PubMed Central PMCID: PMC3645481.
28. Hartig JV, Esslinger S, Bottcher R, Saito K, Forstemann K. Endo-siRNAs depend on a new isoform of loquacious and target artificially introduced, high-copy sequences. *Embo J*. 2009; 28(19):2932–44. PMID: 19644447. <https://doi.org/10.1038/emboj.2009.220>
29. Rong M, Durbin RK, McAllister WT. Template strand switching by T7 RNA polymerase. *The Journal of biological chemistry*. 1998; 273(17):10253–60. PMID: 9553077.
30. Lehmann E, Brueckner F, Cramer P. Molecular basis of RNA-dependent RNA polymerase II activity. *Nature*. 2007; 450(7168):445–9. Epub 2007/11/16. <https://doi.org/10.1038/nature06290> PMID: 18004386.
31. Lai MM. RNA replication without RNA-dependent RNA polymerase: surprises from hepatitis delta virus. *J Virol*. 2005; 79(13):7951–8. Epub 2005/06/16. <https://doi.org/10.1128/JVI.79.13.7951-7958.2005> PMID: 15956541; PubMed Central PMCID: PMC1143735.
32. Tseng CH, Lai MM. Hepatitis delta virus RNA replication. *Viruses*. 2009; 1(3):818–31. Epub 2009/12/01. <https://doi.org/10.3390/v1030818> PMID: 21994571; PubMed Central PMCID: PMC3185533.
33. Rackwitz HR, Rohde W, Sanger HL. DNA-dependent RNA polymerase II of plant origin transcribes viroid RNA into full-length copies. *Nature*. 1981; 291(5813):297–301. Epub 1981/05/28. PMID: 7231549.
34. Konarska MM, Sharp PA. Structure of RNAs replicated by the DNA-dependent T7 RNA polymerase. *Cell*. 1990; 63(3):609–18. 1 PMID: 169966899668.
35. Bayne EH, Portoso M, Kagansky A, Kos-Braun IC, Urano T, Ekwall K, et al. Splicing factors facilitate RNAi-directed silencing in fission yeast. *Science*. 2008; 322(5901):602–6. Epub 2008/10/25. <https://doi.org/10.1126/science.1164029> PMID: 18948543; PubMed Central PMCID: PMC2585287.
36. Bayne EH, Bijos DA, White SA, de Lima Alves F, Rappsilber J, Allshire RC. A systematic genetic screen identifies new factors influencing centromeric heterochromatin integrity in fission yeast. *Genome biology*. 2014; 15(10):481. Epub 2014/10/03. <https://doi.org/10.1186/s13059-014-0481-4> PMID: 25274039; PubMed Central PMCID: PMC4210515.
37. Kallgren SP, Andrews S, Tadeo X, Hou H, Moresco JJ, Tu PG, et al. The proper splicing of RNAi factors is critical for pericentric heterochromatin assembly in fission yeast. *PLoS Genet*. 2014; 10(5):e1004334. Epub 2014/05/31. <https://doi.org/10.1371/journal.pgen.1004334> PMID: 24874881; PubMed Central PMCID: PMC4038458.
38. Malone CD, Mestdagh C, Akhtar J, Kreim N, Deinhard P, Sachidanandam R, et al. The exon junction complex controls transposable element activity by ensuring faithful splicing of the piwi transcript. *Genes Dev*. 2014; 28(16):1786–99. PMID: 25104425. <https://doi.org/10.1101/gad.245829.114>
39. Xiong XP, Kurthkoti K, Chang KY, Lichinchi G, De N, Schneemann A, et al. Core small nuclear ribonucleoprotein particle splicing factor SmD1 modulates RNA interference in *Drosophila*. *Proc Natl Acad Sci U S A*. 2013; 110(41):16520–5. Epub 2013/09/27. <https://doi.org/10.1073/pnas.1315803110> PMID: 24067655; PubMed Central PMCID: PMC3799365.
40. Xiong XP, Vogler G, Kurthkoti K, Samsonova A, Zhou R. SmD1 Modulates the miRNA Pathway Independently of Its Pre-mRNA Splicing Function. *PLoS Genet*. 2015; 11(8):e1005475. Epub 2015/08/27. <https://doi.org/10.1371/journal.pgen.1005475> PMID: 26308709; PubMed Central PMCID: PMC4550278.
41. Volanakis A, Passoni M, Hector RD, Shah S, Kilchert C, Granneman S, et al. Spliceosome-mediated decay (SMD) regulates expression of nonintronic genes in budding yeast. *Genes Dev*. 2013; 27(18):2025–38. Epub 2013/09/26. <https://doi.org/10.1101/gad.221960.113> PMID: 24065768; PubMed Central PMCID: PMC3792478.
42. Zhang Z, Wang J, Schultz N, Zhang F, Parhad SS, Tu S, et al. The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. *Cell*. 2014; 157(6):1353–63. Epub 2014/06/07. <https://doi.org/10.1016/j.cell.2014.04.030> PMID: 24906152; PubMed Central PMCID: PMC4167631.
43. Hur JK, Luo Y, Moon S, Ninova M, Marinov GK, Chung YD, et al. Splicing-independent loading of TREX on nascent RNA is required for efficient expression of dual-strand piRNA clusters in *Drosophila*. *Genes Dev*. 2016; 30(7):840–55. Epub 2016/04/03. <https://doi.org/10.1101/gad.276030.115> PMID: 27036967; PubMed Central PMCID: PMC4826399.

44. Tresini M, Warmerdam DO, Kolovos P, Snijder L, Vrouwe MG, Demmers JAA, et al. The core spliceosome as target and effector of non-canonical ATM signalling. *Nature*. 2015; 523(7558):53–+. <https://doi.org/10.1038/nature14512> PMID: 26106861
45. Sikorski TW, Ficarro SB, Holik J, Kim T, Rando OJ, Marto JA, et al. Sub1 and RPA associate with RNA polymerase II at different stages of transcription. *Mol Cell*. 2011; 44(3):397–409. Epub 2011/11/08. <https://doi.org/10.1016/j.molcel.2011.09.013> PMID: 22055186; PubMed Central PMCID: PMC3227220.
46. Zaratiegui M, Castel SE, Irvine DV, Kloc A, Ren J, Li F, et al. RNAi promotes heterochromatic silencing through replication-coupled release of RNA Pol II. *Nature*. 2011; 479(7371):135–8. Epub 2011/10/18. <https://doi.org/10.1038/nature10501> PMID: 22002604.
47. Lee H-C, Chang S-S, Choudhary S, Aalto AP, Maiti M, Bamford DH, et al. qiRNA is a new type of small interfering RNA induced by DNA damage. *Nature*. 2009; 459(7244):274–7. <https://doi.org/10.1038/nature08041> PMID: 19444217
48. Shah C, Forstemann K. Monitoring miRNA-mediated silencing in *Drosophila melanogaster* S2-cells. *Biochim Biophys Acta*. 2008; 1779(11):766–72. Epub 2008/07/19. <https://doi.org/10.1016/j.bbagr.2008.06.008> PMID: 18634912.
49. Bottcher R, Hollmann M, Merk K, Nitschko V, Obermaier C, Philippou-Massier J, et al. Efficient chromosomal gene modification with CRISPR/cas9 and PCR-based homologous recombination donors in cultured *Drosophila* cells. *Nucleic Acids Res*. 2014; 42(11):e89. Epub 2014/04/22. <https://doi.org/10.1093/nar/gku289> PMID: 24748663; PubMed Central PMCID: PMC4066747.
50. Kraus A, Grampp S, Goppelt-Struebe M, Schreiber R, Kunzelmann K, Peters DJM, et al. P2Y2R is a direct target of HIF-1alpha and mediates secretion-dependent cyst growth of renal cyst-forming epithelial cells. *Purinergic Signal*. 2016; 12(4):687–95. PMID: 27565965. <https://doi.org/10.1007/s11302-016-9532-5>
51. Flockhart I, Booker M, Kiger A, Boutros M, Armknecht S, Ramadan N, et al. FlyRNAi: the *Drosophila* RNAi screening center database. *Nucleic Acids Res*. 2006; 34(Database issue):D489–94. PMID: 16381918. <https://doi.org/10.1093/nar/gkj114>
52. Elmer K, Helfer S, Mirkovic-Hosle M, Forstemann K. Analysis of endo-siRNAs in *Drosophila*. *Methods in molecular biology*. 2014; 1173:33–49. Epub 2014/06/13. [https://doi.org/10.1007/978-1-4939-0931-5\\_4](https://doi.org/10.1007/978-1-4939-0931-5_4) PMID: 24920358.
53. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009; 10(3):R25. PMID: 19261174. <https://doi.org/10.1186/gb-2009-10-3-r25>
54. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>( $\Delta\Delta C_T$ ) Method. *Methods*. 2001; 25(4):402–8. PMID: 11846609. <https://doi.org/10.1006/meth.2001.1262>
55. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*. 2009; 10:48. PMID: 19192299. <https://doi.org/10.1186/1471-2105-10-48>