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

Proteomic insights into spliceosome components involved in antisense transcription at DNA breaks

Petar Đurić

aus

Kragujevac, Serbien

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.
Frankfurt am Main, 07. 05. 2023

.....

Petar Đurić

Dissertation eingereicht am 27. 05. 2023

1. Gutachter: Prof. Dr. Klaus Förstemann
2. Gutachter: Prof. Dr. Julian Stingele

Mündliche Prüfung am 17. 07. 2023

Abstract

Small interfering RNAs (siRNAs) are expressed directly at DNA breaks in multiple eukaryotic species. This opens questions on their expression mechanism in *Drosophila*, where they naturally arise via double-stranded precursors generated by sense and antisense transcription. This thesis aims to examine the missing link between the DNA break and Polymerase II-mediated antisense transcription at DNA breaks.

Previously shown requirement of splicing components for the efficient generation of siRNAs at the DNA break was taken as a starting point: multiple splicing proteins were proteomically examined unraveling potential new proteins in the signaling cascade at the DNA break. Genome-wide DNA damage was induced before the proteomic measurements to examine the function of spliceosome outside of the splicing context. To follow splicing proteins of interest, epitope tags were introduced at C-termini using CRISPR/Cas9 in *Drosophila* cells.

Proteomic results suggested spliceosome might be binding Putzig more during DNA damage. Putzig is a zinc-finger transcription factor known for its role in transcription and an associate of multiple transcriptional complexes in *Drosophila*. Using luciferase-based reporter assay and deep-sequencing, it could be confirmed that Putzig has a role in generating small RNAs at DNA breaks. Its role in the genomic context was shown for the first time in this work.

1. Introduction

1.1. Sense and antisense transcription

The process of transcription is the synthesis of RNA molecules based on information in DNA. It is an essential intermediate step of converting the genetic program into effector molecules, which is not only specific to the cell type but also the current cellular context. This makes it a highly dynamic process [1] that is dependent on and regulated according to internal and external conditions. The RNA molecules that result from the process can directly exert effector roles in the cell [2] [3] [4] or serve as a template for protein synthesis [5].

Central to transcription is the RNA polymerase [6], an enzyme that catalyzes the formation of a phosphodiester bond between two ribonucleotides. The ribonucleotide chain corresponds to the DNA sequence being transcribed, which is ensured by the complementary binding of ribonucleoside triphosphates (rNTPs) to the deoxyribonucleotides in DNA via hydrogen bonds. After their binding, RNA polymerase catalyzes the phosphodiester bond between the rNTP and a growing ribonucleotide chain, which leads to the release of the pyrophosphate ions and extends the ribonucleotide chain. RNA is extended in a 5’ to 3’ direction, referring to the inherent directionality of the molecule. 5‘ (five-prime) end presents the free phosphate group attached to the fifth carbon atom of the pentose sugar, while 3‘ (three-prime) end presents a free hydroxyl group on the third carbon atom of the pentose sugar. RNA polymerase is only able to catalyze the addition of a ribonucleotide at the free 3‘ end of the growing ribonucleotide chain.

Next to RNA polymerase, various other proteins are involved in transcription, and they form distinct complexes that mark different transcription phases. This work focuses on eukaryotic transcription, to which the further text refers. The process of transcription continues from the initiation phase through the elongation phase ending in the termination phase. In the initiation phase, the initiation complex [7] is formed including

various proteins (transcription factors) next to RNA polymerase, which help its recruitment to the distinct DNA sequences relevant for the transcription start named the core promotor elements [8] [9].

The sense DNA strand has the same sequence as the mRNA sequence, which means that the opposite strand was used as a template for transcription, the so-called antisense DNA strand. Sense and antisense strand are terms relevant only in the case of a particular RNA transcript and cannot be used for DNA as a whole. Sense transcripts are considered to give rise to mRNA and later proteins, while antisense non-coding transcripts were initially considered noise. It is known today that they are important regulators of gene expression [10]. Two important mechanisms are to be considered: First, that the mere event of antisense transcription can affect another transcription event [11], and second, that the products resulting from the antisense transcription can themselves undertake various effector roles in transcription. Their role in transcriptional regulation [12] ranges from regulating chromatin state [13] [14] to regulating the expression of particular mRNAs and/or post-transcriptionally regulating their stability [15] [16] [17] [18].

Transcription in *Drosophila melanogaster* [19] includes numerous confirmed antisense transcripts [20]. Its genome also shows the classical landmarks of how the eukaryotic genome evolved into expanding its regulatory instead of protein-coding potential [241], wherein the approximately 170 million base pair genome there are only around 13000 genes [21] [22], with computational predictions that around 50% of the genome is intergenic [23]. This trend continued later, with *Homo sapiens* having a 2.9 billion base pair genome with an estimated around 20000 genes [24] [25].

Drosophila takes advantage of maximizing the genome potential through sex-specific and tissue-specific alternative splicing [26] [27] that leads to the same DNA sequence producing distinct transcripts that later contribute to fine-tuning protein functions. It is estimated that 60% of the genes are being alternatively spliced, comparing to 95% in *Homo sapiens* [28]. Alternative splicing has been confirmed to have a role in development [29] [30] [31], stress response [32] [33], metabolism [34] as well as DNA repair [35].

Drosophila melanogaster also has a very active content of transposable elements [36] [37], that can change position and potentially cause genome instability through insertions, deletions, and illegitimate recombination [38] [39], and played an important role in eukaryotic evolution [39]. DNA transposons, which include a transposase gene (coding for transposase, the enzyme responsible for excising the transposon and its subsequent insert into a new genomic location), can move inside the genome with the cut and paste mechanism. While they are not active in the genome of *Homo sapiens*, 16% of *Drosophila*'s DNA transposons are considered active [36] [37]. A second big group of transposons are retrotransposons, which require an RNA intermediate during their position moving. The RNA is reverse transcribed and subsequently inserted into a different genomic position [40]. These transposons are classified into two distinct groups, based on the presence of a long-terminal repeat (LTR) flanking genes with the mobilization role. Both groups, LTR and non-LTR transposons are present in *Drosophila* [41] [42] [43]. It is estimated that 45% of LTR and 21% of non-LTR transposons are potentially active in *Drosophila* [36]*,* among others copia [44], blood [45], gypsy [46], and roo [47]. This has led to a very delicate system of antisense transcripts regulating the transposon levels [48], which will be discussed further in the text.

Apart from the internal challenges to genome stability in the form of transposons, *Drosophila* is challenged by various viruses [49, 50], many of them discovered through metagenomic analysis [51] [52]. Results show *Drosophila* is host to numerous viruses with negative-strand RNA genomes, positive-strand RNA genomes, double-stranded RNA genomes, as well as a smaller number of viruses with DNA genomes [51]. *Drosophila* was thus an important experimental model [53] for understanding virus-host interactions [53] [54], especially since many of the signaling pathways triggered as a consequence of the viral infection in *Drosophila* are conserved in mammals [49] [55] [56]. Extended to that, some viruses have both *Drosophila* and *Homo sapiens* as a host, including the West Nile virus [57], Dengue virus [58], and other arboviruses [49]. Similar to the endogenous regulation of transposons, they use a system of non-coding RNAs to combat the viral infection. Important to mention is, their viral response extends beyond small RNAs, including phagocytosis of infected cells, transcriptional

pausing et cetera [50].

1.2. Classes of small non-coding RNAs in *Drosophila*

Different non-coding transcripts (known as small non-coding RNAs) participate in the sequence-specific degradation of nucleic acids. They assist in the fine regulation of endogenous gene expression by affecting the stability and thus the amount of mRNA [59], control transposon levels [60], and combat viral infections as well [61]. The latter two are performed by siRNAs, as part of a system commonly known as RNA interference (RNAi), which was confirmed in many eukaryotic species [62]. The RNAi signal is a form of nucleic acid-based immunity, as it is known that it propagates between cells, protecting not yet infected cells [63]. First used in *C. Elegans* [64], RNAi triggered by synthetic double-stranded RNAs (dsRNAs) found meaningful applications. These include numerous medical and industrial applications including among others partial disease control in animal models [65], partial parasite reduction [66], crop improvement [67], and its contribution to basic research by providing a way to cost-effectively perform gene knockdowns [68] [69].

The system, although present in many species, has a common principle of function: (i) processing of the transcript by cellular RNases, (ii) complexing the processed transcript together with a nuclease from the Argonaute family [70], (iii) sequence-specific pairing of the RNA-nuclease complex with a target nucleic acid and (iv) cleavage of the target nucleic acid by the nuclease. Three classes of small non-coding RNAs in *Drosophila melanogaster* are micro RNAs (miRNAs), PIWI-interacting RNAs (piRNAs), and small-interfering RNAs (siRNAs) [71].

MicroRNAs (miRNAs) present a class of conserved, 21-24 nucleotide (nt) long class of small non-coding RNAs that post-transcriptionally regulate the protein-coding transcripts (mRNAs), mostly via their 3’ UTR (untranslated region). They are linked to many physiological (germline development [72], tissue growth [73], cell differentiation [74], and among others behavior [75]) as well as pathological conditions (including

neurodegenerative disorders and more [76] [59] [77]). Their genes are mostly transcribed by RNA polymerase II in the nucleus resulting in primary miRNAs (pri-miRNAs) which are subsequently cleaved by the microprocessor complex [78]. The member of the microprocessor complex that is central to processing is an RNase III nuclease Drosha (together with the dsRNA-binding protein Pasha) [79], resulting in around 70 nt long precursor miRNA (pre-miRNA) with a hairpin structure and a 5' phosphate end, and a 2 nt single-stranded overhang at the 3' end. At this stage, pre-miRNAs are exported to the cytoplasm via Exportin-5 [80] [81], where they are processed by the RNase III nuclease Dicer-1 [82] together with Loquacious [83] [84] isoform B into a 22 nt long miRNA/miRNA* duplex. Particularly Dcr-1 from the Dicer family is involved in the processing of the miRNAs in *Drosophila*. One of the miRNA strands [85] is being incorporated into the RNA-induced Silencing Complex (RISC complex) which contains the Argonaute (Ago) nuclease [86], and where Ago1 is preferentially involved in the miRNA pathway [87]. The strand that is incorporated in the RISC complex is called the guide strand, whereas the strand not being incorporated in the RISC complex is called the passenger strand and is eventually degraded (miRNA*) [88]. The RISC complex is with the incorporation of the miRNA activated and can now cleave the mRNA complementary to the miRNA incorporated in the complex. Like the silencing miRNA system seen in other animals, perfect matching of miRNA to the transcript is necessary for efficient degradation, while a few mismatches can lead to translation attenuation. Separately, miRNA can arise also from introns, named mirtrons [89].

Small interfering RNAs (siRNAs) present a class of 21 nt long small non-coding RNAs and are considered a part of the RNAi system. The pathway can be activated either due to exogenous (of viral origin or in-vitro synthesized) or endogenous (originating from the transcription of the host's genome) dsRNA precursors [90]. Endogenous precursors can be long-structured loci, natural sense-antisense transcripts (NATs), as well as transposable elements. The siRNAs are present in both somatic and gonad cells.

In the siRNA pathway, the dsRNA precursor is bound and processed by the heterodimer consisting of Dcr-2 and R2D2. Additional proteins join to form the RISC

loading complex (RLC). Next, Ago2 joins which marks the forming of the pre-RISC complex. As in the case of miRNAs, the passenger strand is cleaved and subsequently ejected. siRNA-directed cleaving by Ago2 is a single cut of the target sequence. The cut is precise, and it is exactly the siRNA residues 10 and 11 if counted from the 5' end. The cut is followed by the degradation of the fragments by the cellular exonucleases.

siRNAs are at least in worms amplified to create secondary siRNAs, and in these organisms, this is possible through the presence of the RNA-dependent RNA polymerase (RdRP). In worms [91] or plants [92], this can lead to a systemic RNAi response through the organism.

PIWI-interacting RNAs (piRNAs) are a 24-31 nt long class of small RNAs with a role in regulating transposons in the gonads [93] [94] [95] where their expression is spatially restricted [96]. They protect the germline's genome from transposable elements and take part in germline differentiation [97] [98] [99] [100]. Just as miRNAs and siRNAs, they form a complex with a nuclease from the Argonaute family [101]. In *Drosophila*, while Piwi nuclease is present in complexes with piRNAs in both ovarian follicle- as well as germ cells, Aub and Ago3 are present only in the latter. Piwi silences transposons co-transcriptionally in the nucleus, while Aub and Ago3 exhibit slicer activity by cleaving transposon transcripts [102].

piRNAs can arise in three different ways: the de-novo biogenesis [102], the ping-pong cycle [103], as well through phasing [104]. The de-novo pathway is not understood well, but it gives rise to piRNAs that enter the ping-pong cycle. In the first step of the cycle, the Aub-loaded piRNA, which is usually antisense to transposon mRNA, drives the sequence-specific slicing of transposons by Aub. After the single cleavage, the transposon transcript is now divided into two fragments. The RNA helicase Vasa helps the transfer of the 3‘ fragment onto Ago3, which is further processed by a 3‘-5‘ exoribonuclease. This leads to a mature piRNA, which aids sequence-specific slicing of complementary transposon by Ago3. To my knowledge not yet identified protein drives the transfer of the new transposon fragment back to another Aub nuclease, where it matures and aids the Aub slicing activity. This way, the ping-pong activity of

piRNA between Aub and Ago3 leads to the maturation of piRNAs. Ago3 drives separately the process of phasing. In this pathway, the cleavage products by Ago3 are processed by Zucchini, and after that loaded into the piRISC complex together with Piwi. Piwi can separately bind de-novo piRNAs, but their abundance is lower than those produced via the phasing [102].

1.3. Small interfering RNAs expressed at the DNA break

Small interfering RNAs are also known to be expressed directly at the DNA break [105]. They were first discovered in the experiments carried out in *Arabidopsis thaliana* [106] and were quickly confirmed in human cells as well [107], providing proof of evolutionary conservation of the phenomenon. The expression of damage-induced RNAs (diRNAs) was independently found in *Drosophila* [108]. This chapter will cover the comparative analysis of diRNA synthesis in these species: *Arabidopsis thaliana*, *Homo sapiens*, *Mus musculus*, and *Drosophila melanogaster*.

In *Arabidopsis thaliana*, damage-induced RNAs (diRNAs) arise from the double-stranded precursors that are made from the sense or antisense transcript by one of the few RNA-dependent RNA polymerases. Results show that the generation of diRNAs happens at a similar frequency from sense and the antisense strands [106]. Mutations in RNA-dependent RNA polymerases RDR6 and RDR2 lead to a massive decline in diRNA production (82 and 87%). The same RNA-dependent RNA polymerases have a role in other small RNAs processed from double-stranded precursors [109] [110]. The transcription of diRNAs also depends on the presence of RNA polymerase IV, and ATR kinase, and they are complexed with AGO2 as mature siRNAs [106].

In *Homo sapiens*, diRNAs are detected at the repetitive rDNA regions, whilst it was not shown they could be detected at the unique gene or intergenic loci [107]. In the experiments with I-PpoI nuclease, diRNAs at the repetitive loci were detected dominantly upstream, and not downstream of the cut. Inhibition of RNA Pol II by α-

Amanitin (binds the RNA polymerase II active site [111]) or DRB (blocks carboxyl-domain terminal kinases and thus transcription elongation [112]) leads to a reduced number of the reads upstream of the cut with the I-PpoI nuclease, proving its direct role in the process. Inhibition of RNA Pol I as a control did not lead to a decrease in the reads upstream of the cut [107].

Processed independently of the microprocessor complex members Drosha and DGCR8, they are however processed by Dicer and subsequently loaded to AGO-family protein. However, a Dicer knock-out in the mouse embryonic stem cells (mESCs) did not lead to the complete abolishment of diRNA expression, but only those of 21 nt in length, typical for the Dicer-processed small RNA. The longer ones were still detected by deep sequencing, implying at least two different classes of diRNAs expressed at the DNA break at least in mice [107].

In *Drosophila*, double-strand breaks but not nicked DNA lead to the production of the diRNAs. Experiments with the expression vectors showed that (i) their efficient production depends on the RNAi factors and in particular Dcr2 and Ago2 and that (ii) diRNAs produced can silence the transcripts *in trans*, as seen with the luciferase reporters [108] [113]. Furthermore, their expression is dependent on the active transcription [114]. This argues that the double-strand precursor in *Drosophila* arises from the sense strand of the currently produced transcript, and the potential recruitment of a polymerase to initiate transcription at the DNA break leads to antisense transcription.

Drosophila lacks RNA Polymerase IV and RNA-dependent RNA polymerase. This opens the question of which RNA Polymerase is recruited to the DNA double-strand break to perform the antisense transcription. The double-stranded precursor for diRNA is in *Drosophila* different than in other species, made directly on the DNA. Under the assumption that the currently transcribed region in the sense direction by the RNA polymerase II serves as one-half of the double-strand template for siRNA, mechanistic difficulty exists in inhibiting RNA Polymerase II to test for its antisense transcription activity. However, nascent sequencing followed by an RNA Polymerase II immunoprecipitation (NET-Seq) points to the activity of the enzyme in the antisense

transcription at the DNA break. In other organisms, the hypothesis on RNA Polymerase choice is partially challenged by the reports of RNA Polymerase III being recruited to the break event [115]. This however could not be confirmed in Drosophila with the same NET-Seq experiments [116] [113].

This opens questions on mechanistic aspects of the antisense transcription at the DNA break by RNA Polymerase II and whether a new transcriptional complex is being recruited or the already transcribing polymerase makes a U-turn-like movement.

1.4. Proposed role of damage-induced siRNAs in DNA repair

DNA repair is crucial for genome stability and cell survival, as it ensures intactness of DNA. Multiple DNA repair pathways exist in *Drosophila* [117] [118] and the pathway of choice depends on the type of break and current cellular context. Those include base excision repair [119] [120], nucleotide excision repair [121] [122], mismatch repair [123] [124] [125], interstrand crosslink repair [126], repair by homologous recombination and repair by end-joining, where the canonical and theta-joining [117] (also known as microhomology-mediated end joining [127]) have been described. DNA double-strand breaks are repaired with either repair by homologous recombination [128] [129] [130] (referred to as HR) or non-homologous end-joining repair [131] (from now referred to as NHEJ). Outcomes of these two pathways are different: while HR restores the genetic information, NHEJ can result in mutagenic outcomes, for example, insertions or deletions [132]. Classically seen the choice between HR and NHEJ depends on the cell cycle, where NHEJ is predominantly active in cells currently in the G1 phase, while HR is predominantly active in the S and G2 phases since a sister chromatid is required. However, evidence shows that in *Drosophila* the repair pathway choice also depends on tissue type and development stage [117]. Germline cells predominantly repair DNA double-strand breaks by HR, while somatic cells are thought to rely more on NHEJ.

Small interfering RNAs resulting in the place of DNA damage, damage-induced RNAs (diRNAs), have been postulated to affect DNA repair by homologous recombination. This was shown in experiments on *Arabidopsis thaliana*, where the efficient HR is perturbed upon DCL2, DCL3, and DCL4 knockdowns, which also affects diRNA expression. Effects were particularly strong for DCR3 (98% reduction in diRNA expression) [106]. Given that they are expressed either adjacent to the break site (in *Drosophila melanogaster* and *Arabidopsis thaliana*) or a few nucleotides away (in *Homo sapiens*), it is postulated that they are acting in cis. However, a clear role of diRNAs remains elusive.

However, a few mechanisms of action have been proposed. There is no evidence that diRNAs are involved in the formation of the H2AX foci in *Arabidopsis* or healthy human cells, characteristic for the initial recruitment of DNA repair factors to the break site, and their role is postulated to be downstream of these events, leaving the initial DNA damage response independent of their presence [106]. Later it was proposed that their role is limited to homologous recombination via AGO2, important for the RAD51 recruitment. According to this model, diRNAs guide AGO2 to the site of the DNA break [133]. It was independently found that they have an importance in the recruitment of RAD51 and BRCA1, aside from engaging in histone modifications [134].

In *Drosophila*, reporter assays show no evidence for the requirement of diRNAs in the DNA repair by homologous recombination [113], which is unperturbed upon Dcr-2 knockdowns, as well as Dcr-2 and Ago-2 mutant flies. Particularly in the case of *Drosophila,* the collateral effects on the miRNA pathway are avoided, as the pathway is separated from the miRNA pathway with its own Dicer and Ago protein.

1.5. The role of the spliceosome in the expression of diRNAs and DNA repair

The spliceosome [135], a multi-component structure constituted of around 100 proteins and small nuclear RNAs (snRNAs) is an essential part of the process of gene expression. The eukaryotic genes evolved into containing not only coding sequences, named exons, but non-coding sequences as well, named introns. Even though cellular roles have been described for some introns [136], they are usually excised from the transcript for mRNA to get to a mature state. The process of excising introns from the newly forming transcript is called splicing and it occurs co-transcriptionally, meaning it happens while the remaining transcript is still synthesized by the RNA polymerase [137].

The spliceosome is a dynamic structure with multiple complexes forming and resolving along the splicing cycle. This cycle begins with the formation of the early spliceosome E complex via the complementary binding of the U1 snRNA to the 5' splice site [138] [139] [140]. This event is followed by the formation of the A complex [141], where U2 snRNA binds the branching point (BP). Upon joining of the three snRNPs together U4/U6.U5 [142] [143] (the so-called tri-snurp), the pre-B complex is formed. The U4/U6 association is subsequently loosened for U6 to replace U1 at the 5' splice site, and the binding of additional accessory proteins marks the formation of the B complex [144]. The interaction of U4/U6 is then completely lost and U6 can bind U2 to form the active site for the first catalytic reaction, making the activated B complex [145]. When the adenosine from the branching point becomes docked in the active site, the first catalytic reaction happens which results in an intron-lariat [146] intermediate. This transitions into the C complex [147], which represents also the branching position of the spliceosome, with the position of the intron lariat differing. In the activated B complex the phosphate group of the first intron nucleotide is bound to the 3'OH group of the first exon, while in the C complex, it is bound to the 2'OH group of the branching point. For the second catalytic step, the same active site is being used, but the spliceosome moves from its branching conformation to its exon-ligating conformation. The 3' splice site migrates to the active site where the activated C complex makes the second catalytic reaction. The spliceosome then moves to the P complex, marking the

post-catalytic part of the splicing cycle, after which it is disassembled and the individual components recycled for the new round of splicing.

The splicing machinery can extend its function beyond the canonical role of excising the introns from the newly forming mRNA [148]. It is thought that the spliceosome is involved in the early phases of DNA damage events [149] [150] [151]. However, the non-transcribed breaks are repaired efficiently which implies that splicing should not be the essential part of all the repair pathways. Due to its close relationship with the RNA Polymerase II, whose speed and progression in the elongation phase influences the splicing kinetics [152] and vice versa, it is very possible spliceosome senses first-hand the DNA break events, that presumably cause the stalling of the RNA Polymerase II and its displacement from the DNA break position, as shown in yeast [153].

The spliceosome is affected by DNA damage on multiple levels. Various post-translational modifications of splicing proteins have been reported after DNA damage, including PARylation [154], acetylation [155], phosphorylation [156], and ubiquitylation [157]. Various splicing proteins are recruited to the sites of DNA damage [158] [159] [160]. Furthermore, DNA damage affects the expression of splicing factors [161] [162] [163]. On the other side, the depletion of the splicing factors causes DNA damage, as reported in the genome-wide siRNA screen [164].

The spliceosome plays a surprising role in the expression of the damage-induced small RNAs (diRNAs), apart from the classical RNAi components. A genome-wide RNAi screen in *Drosophila* not only revealed components necessary for the efficient generation of the diRNAs, but further analysis also showed that a DNA break in a gene that is spliced triggers more damage-induced siRNAs than those that are not spliced, demonstrating the importance of active transcription and splicing upstream of the break [114].

In particular, the Prp19 complex and its associated proteins (Prp19-related complex) were predominantly identified in the RNAi screen in *Drosophila*. The complex shows certain versatility in the roles it plays in the cell, with the best-described role in splicing. During the formation of the pre-B complex, the three snRNA complexes U4/U6.U5 join and become extensively re-modeled towards the activation of the spliceosome through the forming of the B and the activated B-complex. The Prp19 complex is responsible for the stable association of U6 and U5 to mRNA. Apart from its role in splicing, it mediates DNA damage response, transcriptional elongation, and proteasome degradation [165]. By being included in both the elongation of the RNA Polymerase II and the positioning of particular snRNAs in the precatalytic phase of the splicing, Prp19 has a good sensing opportunity for stalling of the transcriptional machinery due to a DNA break.

A big portion of this work is on Tango4 (PLRG1 in *Homo sapiens*), a *Drosophila* Prp19-related complex protein that was shown to influence the generation of the diRNAs. It is a 52 kDa protein with several WD-40 repeats. The second part of this work was done on Sans fille (*snf*), a protein from the U1/U2 interacting spliceosome E-complex. It is a 24.5 kDa protein with two RNA-recognition motifs.

Under the current model, stalled spliceosomes as a result of the DNA break serve as a signal for the recruitment of the new RNA polymerase II or a turn of the polymerase currently transcribing the sense transcript to perform the antisense transcription, thus forming the double-stranded precursor that will eventually become a substrate for Dcr-2 and lead to mature diRNAs (Figure 1.5.1.). There are initial indications that these events also depend on the canonical homologous recombination factors Mre11-Rad50-Nbs1, shown by the inhibitor Mirin effect in the luciferase reporter assay [116].

It is an open question through which signaling mechanism the spliceosome recruit sor changes the polymerase activity – there is a missing link between the stalled spliceosome and the antisense transcription.

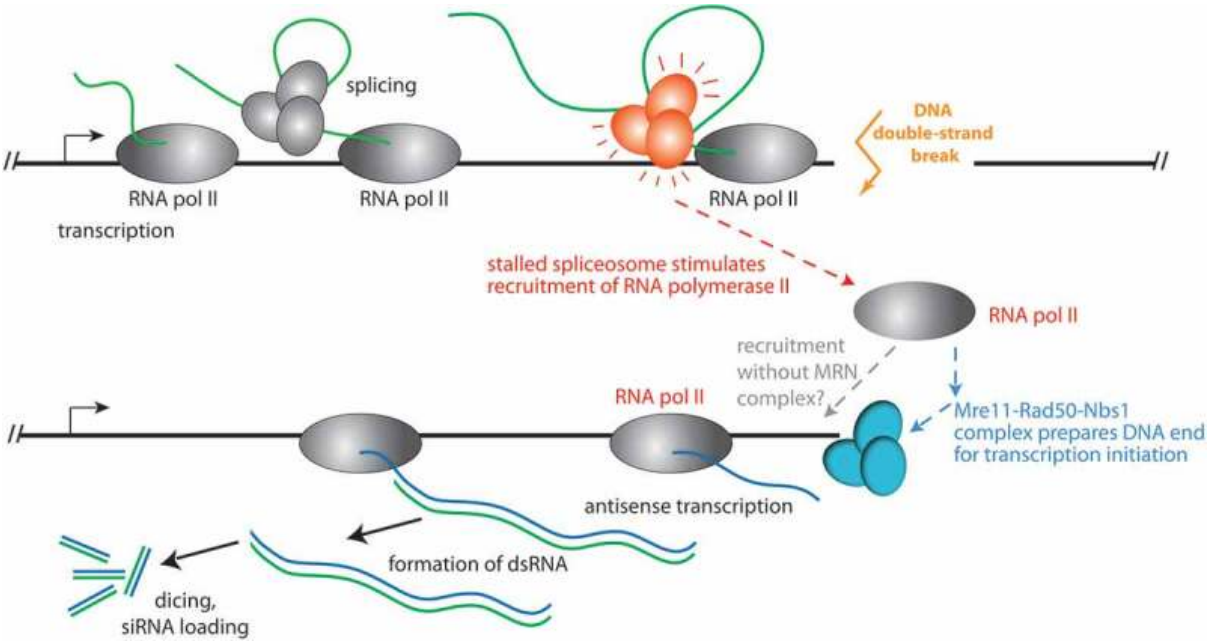


Figure 1.5.1 The model of stalled spliceosomes leading the antisense transcription via RNA polymerase II [116].

2. Materials and methods

2.1. Materials

2.1.1. Primers

Primer name	**Sequence**
CRISPR_Ct_snf	CCTATTTTCAATTTAACGTCGGGAGTTGCGTTCACTTCTGTTTAAGAGCTATGCTG
Ctag_Snf_s	GCGCTGCAGGGCTTCAAGATTACGCCGACGCACGCCATGAAGATAACGTTCGCCAAGAAGGGATCTTCCGGATGGCTCGAG
Ctag_Snf_as	TTATTCTATAGTTTAAATTATTATTGGAGTTGCTGGAACGCCCGGCGACTGGAGTTGCGTGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCCATATG
CRISPR_Ct_tango4	CCTATTTTCAATTTAACGTCGCTAGGAGTCTAGAATTTGGTTTAAGAGCTATGCTG
Ctag_tango4_s	GCCAGCGAGGAGTCGCATCCGATTAACTGGCGACCGGATCTACTGAAGCGCCGCAAATTCGGATCTTCCGGATGGCTCGAG
Ctag_tango4_as	TTTTAATGAAATCGTTTGGAAAATCGTATTCTATTATGGATAGGCTTACAAACTAGGAGTGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCCATATG
CRISPR_Ct_sf3b1	CCTATTTTCAATTTAACGTCGAGTACGAGCGGTACGAGTGTTTAAGAGCTATGCTG
Ctag_sf3b1_s	ACGCTCTGATCGCCGGCTATCCGCGGATTACTAATGATCCCAAGAACCAGTACGAGCGGTACGAGTTGGACTACACGCTAGGATCTTCCGGATGGCTCGAG
Ctag_sf3b1_as	TTTGGATTTATATATTGGTATCTTAGAGAATCCTACCACTAGCAGATAAGCTATTACAATGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCCATATG
CRISPR_Ct_cactin	CCTATTTTCAATTTAACGTCGTATCGCTACAGGCGGTGAGTTTAAGAGCTATGCTG
Ctag_cactin_s	CAGTTCCACAACAACATCTTCCAGCTGTGGTTCCACTTCAAGCGCTATCGCTACAGGCGGGGATCTTCCGGATGGCTCGAG

Ctag_cactin_as	ATAACATAGGACTGCACAAAAGGATGGTTTGGGTAATCTAGGCCAGACGGTGACAATCCTGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCCATATG
pzg_Crispr_cterm	CCTATTTTCAATTTAACGTCGCCCTAGGGCCAATCCTAGTGTTTAAGAGCTAT GCTG
pzg_sense_cterm	GGTGGCCCCTCCGGCGGCAGCTTTTTTGCCCTGGAGGAGTTTACGGAGAC AAAGACCGACGGATCTT CCGGATGGCTCGAG
pzg_as_cterm	CCAGTCCTCTTCGCATGCCAGTATGGTATAGTGGTCATGACCACGCCCCTAGGGCCAATCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCCATATG
Pzg1_t7_s	TAATACGACTCACTATAGGGAACAATCTGCTCAGTAGCGGA
Pzg1_t7_as	TAATACGACTCACTATAGGGCCTGCAGGAGTTTACGTGGAC
Pzg2_t7_s	TAATACGACTCACTATAGGGCGGCTGTTCAACACGCATT
Pzg2_t7_as	TAATACGACTCACTATAGGGGCACTTGAAGAAGTTGGTCTGA
Pzg3_t7_s	TAATACGACTCACTATAGGGGAGGCTACAGATACGCCCAC
Pzg3_t7_as	TAATACGACTCACTATAGGGCACTAGATGCCTCCGACACT
ATPsynB_exon2_qPCR_left	CGTTGCAGAGTACTTCCGTG
ATPsynB_exon2_qPCR_right	CATGTCGGTAGCCAAGGTTG
ATPsynB_exon2_qPCR_right	CATGTCGGTAGCCAAGGTTG
pzg_exon1_q PCR_le	TTGTTCGATGACCTTCC CGA

pzg_exon1_qPCR_right	AAGGAGCTTCTGGGTGT TCA
woc_t7_s	TAATACGACTCACTATAGGGGGGAGACCATGGACTTCT
woc_t7_as	TAATACGACTCACTATAGGGCTCACAAACTTAAAGGCGG
woc_t7_s_v2	TAATACGACTCACTATAGGGCGACTATGGATTTCGGG
woc_t7_as_v2	TAATACGACTCACTATAGGGATCTCGGCCATCATAAGCAA
hp1c_t7_s	TAATACGACTCACTATAGGGCTCATCCAGAAATTCGAGG
hp1c_t7_as	TAATACGACTCACTATAGGGACTCAGCGGACTGATCCAC
hp1c_t7_s_v2	TAATACGACTCACTATAGGGCTATTTCCAGAAGATGG
hp1c_t7_as_v2	TAATACGACTCACTATAGGGGATTTTCCGCCATGGGCA
Row_t7_sense	TAATACGACTCACTATAGGGTTAACGGCCAGGTGTTCCA
Row_t7_antisense	TAATACGACTCACTATAGGGAGCACCAGCTTAACCTTC
Row_t7_sense_vers2	TAATACGACTCACTATAGGGCGGGCTTATGCAATAGAGA
Row_t7_antisense_vers2	TAATACGACTCACTATAGGGCGTTCTTGTCCTGCAAA
rp49 A2	ATCGGTTACGGATCGAACA
rp49 B2	ACAATCTCCTTGCGCTTCTT
CG2493_CRISPR_	cctattttcaatttaacgtcgATAGTGGAGCGATTCAGTAgtttaagagctatgctg

sf3b1_Cterm_tagcheck	CCGTACGGCATTGAATCCT

2.1.2. Plasmids

Plasmid name and description	**Made in this work**
TT01, homology donor for C-term genomic tagging with TurboID	Yes
PD01, CG2493 gRNA	Yes
RB1, Renilla luciferase expression plasmid	No
RB2, Firefly luciferase expression plasmid	No
RB4. Renilla luciferase expression plasmid, truncated variant	No
RB10, plasmid used as a backbone for cloning	No
MH4, homology donor for CRISPR/Cas9 with Flag tag und Blasticidin resistance	No
KF297, homology donor for CRISPR/Cas9 with V5 tag und Puromycin resistance	No

2.1.3. Enzymes

Enzyme	**Producer**
Benzonase	Merck, Germany
DNAse I	ThermoFischer Scientific, USA
Proteinase K	Promega, USA
Taq polymerase	Home-made
Pfu polymerase	Home-made
EcoRI HF	New England Biolabs, USA

NotI HF	New England Biolabs, USA
BamHI HF	New England Biolabs, USA
T7 endonuclease	New England Biolabs, USA

2.1.4. Primary antibodies

Antibody	**Producer**
α*Flag, HRP-coupled	Merck, Germany
αFlag M2	Merck, Germany
αV5, HRP-coupled	Bio-Rad Laboratories GmbH, USA
αV5	Bio-Rad Laboratories GmbH, USA
αHA, HRP-coupled	Roche, Switzerland
αHA, HRP-coupled	Company unknown
αStreptavidin, HRP-coupled	ThermoFisher Scientific, USA
αMyc	Santa-Cruz Biotechnology, USA
αT7 RNAPol	Novagen, Merck, Germany

*anti

2.1.5. Secondary antibodies

Antibiotic	**Producer**
Anti-mouse-HRP	Jackson Immunoresearch, United Kingdom

2.1.6. Antibiotics

Antibiotic	**Producer**
Blasticidin S HCL	Invitrogen, USA
Puromycin	Invitrogen, USA
Zeocin	Invitrogen, USA
Ampicillin in powder	Carl Roth GmbH + Co. KG, Germany
Kanamycin in powder	Carl Roth GmbH + Co. KG, Germany

2.1.7. Kits

Kit	**Producer**
CloneJET PCR Cloning Kit	ThermoFisher Scientific, USA
DyNAmo Flash SYBR Green qPCR Kit	ThermoFisher Scientific, USA
Pierce™ Silver Stain Kit	ThermoFisher Scientific, USA
PureYield™ Plasmid Miniprep System	Promega, USA
QIAquick Gel Extraction Kit	Qiagen, Germany
ReliaPrep™ gDNA Tissue Miniprep System	Promega, USA
RNA Clean & Concentrator-5	Zymo Research, USA
SuperScript™ III Reverse Transcriptase	ThermoFisher Scientific, USA
Wizard® SV Gel and PCR Clean-Up System	Promega, USA
ZR small-RNA PAGE Recovery Kit	Zymo Research, USA

2.1.8. Chemicals

Chemical	**Producer**
Acetic acid	VWR Chemicals, USA
Acetonitrile	Carl Roth GmbH + Co. KG, Germany
Agarose	Biozym Scientific GmbH, USA
Agarose Low Melt	Carl Roth GmbH + Co. KG, Germany
Aluminum sulfate	Carl Roth GmbH + Co. KG, Germany
Ammonium bicarbonate	Carl Roth GmbH + Co. KG, Germany
Ammonium peroxodisulfate	Carl Roth GmbH + Co. KG, Germany
Bradford solution	Bio-Rad, Germany
Chloroform	Merk, Germany
cOmplete™, EDTA-free Protease Inhibitor Cocktail	Roche, Switzerland
Dimethylsulfoxide	Carl Roth GmbH + Co. KG, Germany
Disuccinimidyl suberate	Thermo Fisher Scientific, USA
DTT	Carl Roth GmbH + Co. KG, Germany
EDTA	AppliChem, Germany
Ethanol 99%	Merck, Germany
Ethanol 99%	Carl Roth GmbH + Co. KG, Germany
Formaldehyde 30%	Carl Roth GmbH + Co. KG, Germany
Formic acid	Merck, Germany
FUGENE	Promega, USA
Glycerol	Carl Roth GmbH + Co. KG, Germany
Glycine	Carl Roth GmbH + Co. KG, Germany
Iodoacetamide	Merck, Germany
Isopropanol	VWR Chemicals, USA
Magnesium chloride hexahydrate	Merck, Germany

Magnesium chloride hexahydrate	Merck, Germany
Methanol	VWR Chemicals, USA
N-laurylsarcosine	Merck, Germany
Potassium hydroxide pellets	Merck, Germany
ProtoGel	National Diagnostics, USA
Sodium chloride	Merck, Germany
SybrGold	Carl Roth GmbH + Co. KG, Germany
SybrSafe	Carl Roth GmbH + Co. KG, Germany
TEMED	Carl Roth GmbH + Co. KG, Germany
Tergitol	Thermo Fisher Scientific, USA
Trifluoroacetic acid	Carl Roth GmbH + Co. KG, Germany
TRIS	Carl Roth GmbH + Co. KG, Germany
Triton X-100	Merck, Germany
Trizol	Invitrogen, USA
Tryptone	Carl Roth GmbH + Co. KG, Germany
Tween 20	Carl Roth GmbH + Co. KG, Germany
Urea	Carl Roth GmbH + Co. KG, Germany
UreaGel System Buffer	National Diagnostics, USA
UreaGel System Concentrator	National Diagnostics, USA
UreaGel System Diluent	National Diagnostics, USA

2.1.9. Buffers

Buffer	**Ingredients**
10x TBS	1.5 M NaCl 500 mM Tris pH=7.4
5x SDS Running Buffer	125 mM Tris

	1.25 M Glycine 1% SDS (w/v)
Lysis buffer for the IP	10 mM HEPES 1.5 mM MgCl2 1% Tergitol NP-40 Type 1 mM DTT 10% Glycerol 1x cOmplete™, EDTA-free Protease Inhibitor Cocktail Roche
Wash buffer for the IP	25 mM HEPES 150 mM NaCl 12.5 mM MgCl2 1% Tergitol NP-40 Type 1 mM DTT 10% Glycerol 1x cOmplete™, EDTA-free Protease Inhibitor Cocktail Roche
10x Western blotting solution	250 mM Tris 1.92 M Glycine
Lysis buffer for the Comet Assay	2.5 M NaCl 0.1 M EDTA 10 mM Trizma base of pH 10 1% N-laurylsarcosine 0.5% Triton X-100 10% DMSO
Electrophoresis buffer for the Comet Assay	300 mM sodium acetate 100 mM Tris-HCl

10x PBS	2.7 mM KCl 135 mM NaCl 10 mM Na_2HPO_4 2 mM KH_2PO_4
TBS-T	1x TBS with 0.05% Tween
50x TAE	2 M Tris-base 100 mM EDTA 5.8 % acetic acid
10x TBE	0.9 M Boric acid 0.9 M Tris-base 0.5 M EDTA
DNA loading buffer	30% glycerol 0.025% xylene cyanole 0.025% bromophenol blue
Formamide loading dye	80% formamide 1 mg/ml xylene cyanole 1 mg/ml bromophenol blue 10 mM EDTA
4x SDS loading buffer	200 mM TRIS HCl of pH 6.8 8% SDS 40% Glycerol 0.04% Bromophenolblue
Coomassie G250 stock solution	0.5 g/l stock solution in 100% EtOH
Colloidal Coomassie staining solution	0.5 % Coomassie G250 stock solution 10% EtOH 50 g/l aluminium sulphate 2% H_3PO_4
Coomassie staining solution	45% methanol 0.25% Coomassie Brilliant Blue 10% acetic acid

Coomassie destaining solution	45% methanol 10% acetic acid

2.1.10. *D. Melanogaster* cell lines

Cell line	**Made in this work**
S2-cells with stable Cas9 expression, clone 5-3	No
Tango4-Flag, C term	Yes
Snf-Flag, C term	Yes
Tango-Flag, Snf-V5, C term	Yes
Putzig-Flag, C term	Yes
Putzig-Flag, Tango4-V5, C term	Yes

2.1.11. *E. co*li strains

Bacterial strain	**Made in this work**
Top10	No

2.1.12. Media

Media	**Ingredients**
SOB medium for making the SOC medium	0.5% yeast extract 10 mM NaCl 2% Tryptone mM KCl 10 Mm $MgCl_2$ 10 mM $MgSO_4$
SOC medium	20 mM glucose in SOB medium
LB medium	0.5% yeast extract 1% Tryptone 1% NaCl Supplemented with 100 µg/ml Ampicillin
Schneider's Drosophila medium with L-Glutamin	Produced by Bio & Sell, Germany Supplemented with 10% % fetal bovine serum (Sigma-Aldrich, USA) for cultivation of S2 *Drosophila* cells or without it for transfection purposes

2.1.13. Hardware

Device	**Producer**
Amersham Imager 600	General Electrics, USA
EASY-nLC 1000 Liquid Chromatograph	ThermoFisher Scientific, USA

FACScalibur Flow Cytometer	Becton Dickinson, USA
Gel imager with a transilluminator	Intas Science Imaging Instruments GmbH, Germany
Guava® easyCyte™ Flow Cytometer	Luminex Corporation
K5250 basic shaker	IKA-Werke, Germany
Leica TC-SP2 fluorescent microscope coupled to a Canon Eos 77D camera	Leica, Germany Canon, Japan
LTQ Orbitrap XL Hybrid Ion Trap Mass Spectrometer	ThermoFisher Scientific, USA
Fluorescent microscope equipped with a photo camera	Leica Camera, Germany Canon, Japan
PCR thermocycler Lab Cycler 1119270116	SensoQuest GmbH, Germany
PCR thermocycler Life Touch	Bioer Technology, Germany
Polymax 1040 Shaker	Heidolph Instruments GmbH & Co. KG, Germany
PowerPac Basic Power Supply	Bio-Rad Laboratories GmbH, USA
PowerPac HC Power Supply	Bio-Rad Laboratories GmbH, USA
Q Exactive HF-X Hybrid Quadrupole-Orbitrap Mass Spectrometer	ThermoFisher Scientific, USA
qPCR thermocycler BioMetra	Analytik Jena, Germany
Reax 2 Shaker	Heidolph Instruments GmbH & Co. KG, Germany
Table centrifuge Rotanta 460 R	Hettich, Germany
Tabletop centrifuge 5417R	Eppendorf, Germany
Tabletop centrifuge 5415R	Eppendorf, Germany
Tecan-M1000	Tecan Group, Switzerland
Thermo mixer compact	Eppendorf, Germany
UltiMate 3000 HPLC System	ThermoFisher Scientific, USA

2.1.14. Software

Program	**Producer**
ImageJ	National Institute of Health, USA
Serial Cloner	http://serialbasics.free.fr/Serial_Cloner.html
Ape	M. Wayne Davis, USA
R	The R Foundation, The Comprehensive R Archive Network
Microsoft Office	Microsoft, USA
Perseus	Max Planck Society, Germany
MaxQuant	Max Planck Society, Germany
Guava Soft 2.7	Merck, Germany
Flowing software	Turku Bioscience Center, Finland
qPCR soft 3.4	Analytik Jena, Germany
Tecan i-control	Tecan Group, Switzerland
Integrative Genomics Viewer (IGV)	Broad Institute, USA

2.1.15. Databases

Database	**Owner**
InterPro	EMBL-EBI, United Kingdom
FlyBase	Harvard University, Indiana University, University of Cambridge

2.2. Methods

2.2.1. Culturing *Drosophila* S2 cells

The cells were cultured in Schneider's *Drosophila* medium with L-Glutamine (Bio & Cell, Germany), supplemented with 10% fetal bovine serum (Sigma-Aldrich, USA) and 1% Gibco™ Penicillin-Streptomycin (Thermo-Fisher Scientific, USA). They were split in TC plates (Sarstedt, Germany) at the concentration of $1x10^6$ cells/ml and kept at 25° C. S2 cells were in suspension and they were split once to twice weekly by diluting them 1:10 into fresh medium.

2.2.2. Transfection of *Drosophil*a S2 cells

Drosophila S2 cells were transfected for different purposes by diluting the nucleic acid of interest in Schneider's Drosophila medium (Bio & Cell, Germany) without the fetal bovine serum and 4% FuGENE® HD Transfection Reagent (Promega, USA). The transfection mix was incubated at room temperature and after 30 minutes added to the cells (at a concentration of $1x10^6$ cells/ml).

2.2.3. Cloning of *Drosophila* S2 cells

Drosophila S2 cells were diluted to either $8x10^3$ cells/ml or $16x10^3$ cells/ml and plated in TC 96-well plates (Sarstedt, Germany) in 100 µl of the adapted medium. The medium was composed of 80% selection media with either Blasticidin S HCl (Thermo-Fisher Scientific, USA) in the final concentration of 10 µg/ml or Puromycin-Dihydrochloride (Thermo-Fisher Scientific, USA) in the final concentration of 0.5 µg/ml and 20% of conditioned media (filtered cell culture medium in which the S2 cells have

been present). Clones were picked after two or three weeks.

2.2.4. Freezing and thawing of the *Drosophila* S2 cells

Drosophila S2 cells were frozen in Schneider's *Drosophila* medium (Bio & Sell, Germany) with 10% Dimethylsulfoxide (Roth, Germany) and kept at -170 °C in cryovials (Biozym, Germany). The cells were thawed on ice, centrifuged at 5000 rpm for 5', and resuspended in fresh media.

2.2.5. Isolation of genomic DNA

For the isolation of genomic DNA, 500 µl of $1x10^6$ cells/ml cells were harvested and centrifuged at 6000 rpm for 5' and then resuspended in 160 µl of phosphate-buffered saline (PBS). The DNA was subsequently isolated using the ReliaPrep™ gDNA Tissue Miniprep System (Promega, USA) according to the manufacturer's instructions. Genomic DNA was stored at -20 °C.

2.2.6. DNA clean-up from gel or PCR reaction

DNA was cleaned up using the Wizard® SV Gel and PCR Clean-Up System (Promega, USA). If from a gel, the band was cut and incubated with the same volume of Membrane Binding Solution and incubated at 50-65 °C until dissolved. If from a PCR reaction, the whole reaction was mixed with the same volume of Membrane Binding Solution. DNA was subsequently cleaned up according to the manufacturer's instructions.

2.2.7. DNA Sanger sequencing

The Sanger sequencing was performed by Eurofins Genomics (Germany).

2.2.8. DNA digest and end-modification for cloning

1-2 µg of DNA was incubated with the respective restriction enzyme either overnight (non-high-fidelity enzymes) or 1 hour (high-fidelity enzymes) from New England Biolabs (USA) at 37 °C. The reaction for enzymes with compatible buffer was assembled as follows:

Component	50 µl reaction
DNA	1-2 µg
10X CutSmart Buffer	5 µl
Restriction enzyme 1	1.0 µl (20 units)
Restriction enzyme 2	1.0 µl (20 units)
Water	up to 50 µl

The end modification for the backbone was performed with 1 µl FastAP Thermosensitive Alkaline Phosphatase (Thermo-Fisher Scientific, USA) and incubated at 37 °C for 30'.

2.2.9. DNA ligation

The DNA was ligated overnight at 24 °C in a 10 µl reaction with 1 µl T4 DNA Ligase (NEB, USA) and the 1 µl 10x T4 DNA Ligase Reaction Buffer (NEB, USA).

2.2.10. Bacterial transformation after cloning

Top 10 Competent cells (lab stock) were thawed on ice and 1 µl of plasmid mini prep was added (concentration varied from 100 to 300 ng/µl). Cells were incubated on ice for 30', heat-shocked in the water bath for 1' at 42 °C, and then put on ice for 2'. 900 ml of SOC medium was added, and cells were shaken at 37 °C on 300 rpm for 30'. For cloning, the whole mix was plated on agar plates with either Ampicillin or Kanamycin to select the transformed cells.

2.2.11. DNA gel electrophoresis

1% agarose (Biozym Scientific GmbH, USA) in 1x TAE buffer (3.1.9.) was boiled. 5 µl SYBRSafe (ThermoFischer Scientific, USA) was added. GeneRuler DNA Ladder Mix (ThermoFischer Scientific, USA) was used as a marker. Gel was run on 55 V for approximately 45'.

2.2.12. DNA editing with the CRISPR/Cas9 system

CRISPR/Cas9 system based on the one from *Streptococcus pyogenes* has been adapted for genome editing in *Drosophila melanogaster* cells [166]. Sequence specificity is ensured via linking the U6 snRNA promoter from the Polymerase III and specific sequence of interest via an extension PCR. Next to sgRNA, a homology donor was also made that included an epitope tag as well as a selection marker. After a 72-hour knockdown of NHEJ factors with dsLig4 and dsMus308, the sgRNA and homology donors were transfected using FUGENE, and selected first for the four days, then in the second round seven days with either Blasticidin (10 µg/ml) or Puromycin (0.5 µg/ml).

2.2.13. RNA isolation

Cells in the concentration of $1x10^6$ cells/ml were pelleted by centrifugation at 5000 rpm for 5' at 4 °C. The pellet was mixed with 500 µl Trizol by repetitive pipetting. The homogenized sample was incubated for 5' at 15-30 °C to dissociate the nucleoprotein complexes. 100 µl of chloroform was added, shaken by hand for 15'', and incubated for 3' at 15-30 °C. The samples were centrifuged for 15' at 10000 rpm in a pre-cooled centrifuge (4 °C). The upper aqueous phase was transferred to a fresh tube. RNA was precipitated with an isopropyl-alcohol content of 50% and the sample was incubated for 10' at room temperature. Samples were centrifuged for 15' at 10000 rpm in a pre-cooled centrifuge (4 °C). The precipitate formed was washed with 1 ml of 75% EtOH. The sample was mixed by vortexing and then centrifuged for 5' at 7500 rpm in a pre-cooled centrifuge (4 °C). RNA was air dried for 10' and re-dissolved by adding 20 µl nuclease-free water.

2.2.14. RNA clean-up and digest

DNAse digest was done with DNase I (ThermoFisher Scientific, USA) for 30' at 37 °C.

Component	Amount
RNA	5-10 µg
DNase Buffer (DNase I with magnesium)	5 µl
1 µl RiboLock	1 µl
1 µl DNase I	1 µl
Water	up to 50 µl

Clean-up and concentration of RNA were done with an RNA Clean & Concentrator kit (Zymo Research, USA) according to the manufacturer's instructions.

2.2.15. Synthesis of cDNA

cDNA was synthesized with either oligoDT or random hexamers using the SuperScript™ III Reverse Transcriptase (ThermoFisher Scientific, USA) according to manufacturer's instructions. The ribonuclease inhibitor of choice was RiboLock RNase Inhibitor (ThermoFischer Scientific, USA).

2.2.16 Synthesis of dsRNA

The PCR reaction was assembled with primers adjusted for T7 polymerase.

Component for PCR
5 µl Taq Puffer +$(NH4)_2SO_4$-$MgCl_2$
1µl dNTPs (10mM each)
5 µl $MgCl_2$ (25mM)
1 µl T7prom_primerspecific_fw (10µM)
1 µl T7prom_primerspecific_rv (10µM)
0.5 µl Taq
35,5 µl H2O
1:10 cDNA

Component for in-vitro transcription
10 µl 10x T7-buffer
25 µl PCR product

0.5 µl 1 M DTT
3.5 µl ATP, 4.25 mM
3.9 µl CTP, 4.25 mM
4.3 µl UTP, 4.25 mM
6.8 µl GTP, 6.5 mM
44 µl Water
2 µl T-7 Polymerase (home-made)

The reaction was incubated overnight at 37 °C. DNase treatment was done with 1 µl of DNase I for 30' at 37 °C. The sample was then centrifuged at maximum speed in a pre-cooled centrifuge for 15'. Strands were annealed for 5' at 95 °C.

2.2.17. T7 endonuclease assay

Genomic DNA was isolated according to 3.2.5. and the DNA was used as a template for the PCR.

Component	**Amount**
10X buffer KCl	2,5 µl
dNTPs	0,5 µl
MgCl2 (25mM)	2,5 µl
Primer 1	0,5 µl
Primer 2	0,5 µl
Taq (bought Taq)	0,5µl
Genomic DNA	2 µl

PCR program:

Initial denaturation	95 °C 5’
Denaturation	95 °C 30’’
Annealing	58 °C 30’’
Extension	72 °C 30’’
Final extension	72 °C 5’’
Cooldown	4 °C hold

Denaturation and reannealing of the PCR product was done by the following program:

95 °C 5’
95 to 85 °C at -2 °C/s
85° to 25 °C at -0.1°C/s
4 °C hold

T7 endonuclease digestion was by adding 10 µl of the PCR product to 10 µl master mix out of 2 µl NEB 2 Buffer, 0.5 µl T7 endonuclease and 7.5 µl H_2O. The mix was digested for 16’ at 37 °C and analysed on a 2.5% agarose gel.

2.2.18. Quantitative PCR (qPCR)

Quantitative PCR was performed using the DyNAmo Flash SYBR Green qPCR kit (ThermoFisher Scientific, USA).

Component	**Amount**
Mastermix	5 µl
Forward primer	0.5 µl
Reverse primer	0.5 µl
ROX	0.5 µl
Water	2.5 µl
cDNA template 1:10	1 µl

2.2.19. Deep sequencing of small RNAs

Small RNAs were size selected by denaturing gel electrophoresis in the 20% urea gel in 1xTBE for 90' at 250 V. As the marker, miRNA marker was used after boiling 5 µl at 95 °C. The samples (~20 µg of RNA) were pre-mixed with an equal amount of formamide loading buffer (homemade) and boiled for 5' at 70 °C. The gel was afterwards stained with 0.1% SybrGold (ThermoFisher Scientific, USA) in a petri dish for 5'. Using the transilluminator the region corresponding to small RNAs was excised and the recovery was made using the ZR Small RNA PAGE Recovery Kit (ZymoResearch, USA) using the manufacturer's instructions.

Linker ligation to the 3'-end of the small RNAs was done for 1.5 hours at room temperature.

Component	Amount
Size selected RNAs in water	6 µl
ATP-free T4 RNA ligase buffer (10x)	1 µl
Solexa 3’ linker	1 µl
PEG-8000 (50%)	1 µl
Truncated T4 RNA ligase 2	1 µl

10 µl of formamide loading buffer was added to the sample and the reaction was heat-inactivated for 5’ at 95 °C. The sample was then loaded to a 15% urea gel and run in 1xTBE at 250 V for 75’. Gel was stained with 10% SybrGold for 5’ in a petri dish and the corresponding region of 25-50 nt was excised using a scalpel (B Brown, Germany). The RNA from the gel slice was recovered using the ZR Small RNA PAGE Recovery Kit (Zymo Research, Germany).

Adapter ligation to the 5’-end of the small RNAs was done for 1.5 hours at room temperature.

Component	Amount
Purified RNA after 3’ ligation in water	5 µl
ATP-free T4 RNA ligase buffer (10x)	1 µl
ATP	1 µl
Solexa 5’ linker	1 µl
DMSO	1 µl
Normal T4 RNA ligase 1	1 µl

The inactivation for done for 5’ at 95 °C. The reaction was mixed with 2 µl of thewith oligo RT primer index and incubated for 2’ at 95 °C. The following mix was prepared:

Component	Amount
5’ adapter ligation reaction	9 µl
Oligo RT primer index (5 µM)	2 µl
5x First Strand Buffer	4 µl
DTT (100 mM)	2 µl
RiboLock RNAse inhibitor	1 µl
dNTP mix (10 mM each)	1 µl

The mix was incubated for 2’ at 50 °C, after which 1 µl Superscript III enzyme (Invitrogen, USA) or 1 µl water as a control was added. The reaction was incubated for 30’ at 50 °C and deactivated for 5’ at 95 °C. PCR amplification of cDNA was done in 23 cycles (98 °C for 1’ initial denaturation, 98 °C for 15’’ denaturation, 58 °C for 30’’ annealing, 72 °C for 30’’ extension, 72 °C final extension).

Component	Amount
First Strand cDNA	5 µl
5x PCR buffer (with 2.5 mM Mg^{2+})	20 µl
dNTP mix (10 mM each)	2 µl
Solexa 5’ ext primer	1 µl
3’ PCR index 1-9 primer	1 µl
Water	70 µl
Hot Start Phusion polymerase	1 µl

The PCR product was gel-purified in a 3% agarose gel for 50’ at 55 V. Bands around 160 nt were excised and purified using the Qiagen gel extraction kit (Qiagen, Germany) according to the manufacturer’s instructions. The product was then quantified in a 3% agarose gel for 50’ at 55 V. The product was then diluted two times separately for the final PCR reaction.

Component Mix 1	Amount	Component Mix 2	Amount	Component Mix 3	Amount
Mixed sample	18 µl	Mixed sample 1:3	18 µl	Mixed sample 1:10	18 µl
Illumina qP 1.1	0.5 µl	Illumina qP 1.1	0.5 µl	Illumina qP 1.1	0.5 µl
Illumina qP 2.1	0.5 µl	Illumina qP 2.1	0.5 µl	Illumina qP 2.1	0.5 µl
5x HF buffer	5 µl	5x HF buffer	5 µl	5x HF buffer	5 µl
dNTPs	0.5 µl	dNTPs	0.5 µl	dNTPs	0.5 µl
Phusion pol.	0.5 µl	Phusion pol.	0.5 µl	Phusion pol.	0.5 µl

PCR amplification was done in 1 cycle (98 °C for 1’ initial denaturation, 98 °C for 15’’ denaturation, 59 °C for 30’’ annealing, 72 °C for 30’’ extension, 72 °C final extension). Samples were run on a 3% agarose gel for 50’ at 55 V. Purification was done with Beckman Coulter™ Agencourt AMPure XP (Beckman Coulter, USA) according to the manufacturer’s instructions. The samples were then sequenced using an Illumina instrument.

2.2.20. Western Blot

Harvested cells were centrifugated at 5000 rpm for 5’ and the medium was removed. Cells were then resuspended in PBS and mixed with SDS loading buffer (200 mM Tris HCl with pH 6.8, 8% SDS, 40% Glycerol) and boiled at 95 °C for 5 minutes before being loaded to an 8-10% SDS gel. The gel was run at 150 V for 1 hour in 1% Running Buffer (125 mM Tris, 1.25 M Glycine, 10% SDS, pH=8.3). The membrane for transfer was activated for 20’’ in 100% EtOH. The transfer was done at 100 V for 1.5 hours in 1x Western Solution (3.1.9.). The membrane was washed two times in water for 10 minutes, followed by one wash for 10 minutes in 1x TBS-T (1.5 M NaCl, 500 mM Tris, pH=7.4, 0.05% Tween). Blocking was done for 1 hour in 5% milk in TBS-T and incubated with a primary antibody (diluted 1:1000-1:5000 times) overnight at 4 °C.

Before exposure, the membrane was washed 3 times for 10’ in TBS-T, after which the Thermo Scientific™ SuperSignal™ West Dura Extended Duration Substrate (ThermoFisher Scientific, USA) solution was used if the primary antibody with HRP was used. If not, the membrane was incubated with the secondary antibody (1:5000-1:10000) for 2 hours at room temperature in TBS-T, and the membrane was washed three times for 10 minutes in TBS-T, before the Western Dura solution was used.

2.2.21. Immunoprecipitation

For immunoprecipitation 10 ml of cells ($1x10^6$ cells/ml) were harvested and washed with 10 ml ice-cold PBS. The pellet was resuspended in 300 µl lysis buffer without urea and left for 10’ on ice. The sample was then centrifuged for 5’ at 5000 rpm in a pre-cooled centrifuge at 4 °C. The supernatant was removed, and the pellet was resuspended with 100 µl lysis buffer with urea and left on ice for 5’. The sample was then centrifuged for 5’ at 5000 rpm. The supernatant was removed, and the pellet was resuspended with 110 µl lysis buffer without urea. The chromatin-bound components were released in this step by adding 2 µl of the nuclease benzonase and incubation at 37 ◦C for 2’. The reaction was stopped with EDTA-NaCl solution and put on ice. After that the sample was centrifuged for 5’ at maximum speed and the supernatant was collected to be put on beads. The beads were washed 3 times with the IP buffer (25 mM HEPES, 150 mM NaCl, 12.5 mM MgCl2, 1% Tergitol NP-40 Type, 1 mM DTT, 10% Glycerol, 1x cOmplete™, EDTA-free Protease Inhibitor Cocktail), after which they were diluted in the lysate and agitated for 1 hour at 4 °C. After that, supernatants were collected, and the beads were washed 4 times with the IP buffer and then boiled with 1x SDS loading dye for 10’ at 95 °C if the sample proceeded to western blotting.

2.2.22. Protein crosslinking

Protein crosslink was done with 1% Formaldehyde for 10’ at room temperature. The reaction was quenched with 1.25 M glycine.

2.2.23. Protein staining with silver stain

Gels were stained with the PierceTm Silver Stain Kit (ThermoFisher Scientific, USA) according to the manufacturer’s instructions.

2.2.24. Mass spectrometry

To get rid of the detergent whose presence can interfere with the measurement, beads from the IP were washed three times with 50 mM NH_4HCO_3. The beads were then centrifuged at maximum speed for 15’ and the supernatant was removed. Trypsin digestion was done with 5ng/µl trypsin in 1M Urea/50 mM NH_4HCO_3 overnight at 37 °C. The sample was centrifuged for 10’ at maximum speed and beads were washed 2 times with 50 mM NH_4HCO_3. To reduce the disulfides, 1 mM DTT in the final concentration was added and the sample was incubated for 10’ in the dark on 25 °C. Subsequently, iodoacetamide was added in the final concentration of 1.35 mM and the sample was incubated for 30’ at 25 °C in the dark. Trifluoroacetic acid (TFA) was then added to a final concentration of 1.25% and was proceeded with Solid Phase Extraction to clean up the sample before measurement, a chromatography which was done in EmporeTM C18-SD 7mm/3ml Extraction Disk Cartridges (Supelco, USA). The columns were conditioned with 1 ml of methanol (1.500 rpm for 1’), followed by 0.1% trifluoracetic acid in 70% acetonitrile (1.500 rpm for 1’), 0.1% trifluoracetic in water (1.500 rpm for 1’). Eluate was loaded and centrifuged at 1500 rpm for 5’. 1% TFA in water was added and the sample was centrifuged at 1500 rpm for 3’. The sample was

eluted with 250 µl of 70% acetonitrile in water (150 rpm for 3'). The sample was lyophilized using the speed vacuum at 42-55 °C for approximately 2 hours. Peptides were then diluted in 15 µl CH_2O_2 for measurement. The samples were then measured on the mass spectrometer.

2.2.25. Luciferase assay

Luciferase experiments were done either after a knockdown (20 µg/ml dsRNA for 72 hours) or a chemical treatment (different concentrations for 2 hours). After that the luciferase plasmids were transfected for 72 hours in the following quantities per well of a 96-well plate (using 1 µl Fugene-HD): circular Renilla luciferase (5 ng total, plasmid pRB1), circular Firefly luciferase (25 ng total, plasmid pRB2) and linear Renilla luciferase (40 ng total, plasmid pRB4 linearized with BamHI). Samples were prepared for measurement with Dual-Glo Luciferase Assay (Promega, USA) according to the manufacturer's instructions. Measurement was done with the machine Tecan Infinite M1000 in the Greiner 96 Flat Bottom Black Polystyrene plates (Greiner, Austria).

2.2.26. Fluorescence-activated cell sorting

The cells were not labeled for the FACS measurement in the Guava® easyCyte™ Flow Cytometer. 10000 cells were selected, and the different cell populations were followed. The quantification was done in the Flowing Software (Turku Bioscience Centre, Finland).

2.2.27. Neutral comet assay

Microscopic slides were coated at least 24 hours before electrophoresis with 0.8% normal melting agarose and kept on room temperature. Low melting agarose of 0.7% was prepared 2 hours before cells were harvested and kept at 37 °C. Around $150x10^3$ cells in total were mixed for each sample with 65 µl of low melting agarose and spread on the microscopic slide. Cell lysis was done directly on the microscopic slide in the dark using the lysis buffer (2.5 M NaCl, 0.1 M EDTA, 10 mM Trizma base of pH 10, 1% N-lauryl sarcosine, 0.5% Triton X-100 and 10% DMSO) at 4 °C for an hour. Electrophoresis was done on 10V using the electrophoresis buffer (300 mM sodium acetate, 100 mM TRIS-HCl). Cells were fixed 2 times 10' using 100% EtOH and air-dried for two hours at least. Cells were stained with 1:10000 Hoechst dye and subjected to microscopy. Comet tails were analyzed by ImageJ.

3. Results

3.1. Tagging spliceosome components using CRISPR/Cas9

The goal of this work was to follow particular splicing proteins after massive induction of DNA damage, to understand better their protein interactions outside of the splicing context. I started with introducing epitope tags to splicing proteins first.

Illustration 3.1.1. Experimental approach in this work. Massive chemical damage was followed by proteomic analysis of particular splicing components.

The DNA sequences coding for the splicing proteins Tango4 and Snf (Figure 3.1.1., A), both belonging to the pre-catalytic spliceosome, were edited using CRISPR/Cas9 to express a Flag tag at the C-terminus of the respective protein in *Drosophila* S2 cells [166]. Tagging of Act5C served as a positive control.

Based on western blotting results using the Flag antibody, it was confirmed that Act5C, Tango4, and Snf were successfully tagged, while other spliceosome candidates, like Sf3b1 and Cactin, were not. Using the PCR check for the tag insertion it was later confirmed that the gene coding for Sf3b1 was edited too but failed to produce a functional protein (Figure 3.13.1, A). Further work was continued with these two cell lines, further referred to as Tango4-Flag and Snf-Flag. Both of these cell lines were further used for proteomic studies where Tango4 and Snf served as bait, and mass spectrometry signals confirmed finally the successful tagging of both proteins.

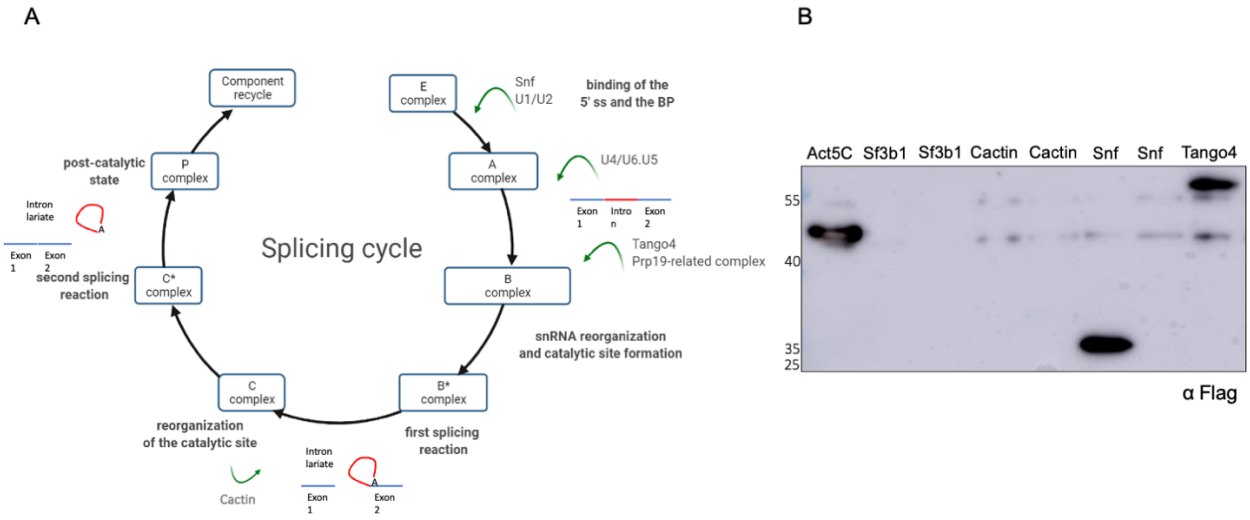


Figure 3.1.1. The splicing cycle. In A, a graphical representation of where proteins of interest have a known function in the splicing cycle. In B, western blot confirmation of successfully expressed spliceosome proteins with an epitope tag. Act5C was the tagging control. Snf and Tango4 were successfully tagged except in one case for Snf, while Sf3b1 although genomically edited (Figure 3.13.1, A), failed to produce a functional protein detected on the western blot. Expected molecular weight: Act5C=41.8 kDa, Sf3b1=149.6 kDa, Cactin=86.6 kDa, Snf=24.5 kDa, and Tango4=52.7 kDa.

3.2. Optimization of the extraction procedure for immunoprecipitation

To test the optimal co-immunoprecipitation conditions, a double-tagged cell line with both Tango4 and Snf having an epitope tag was made, and clonal selection was performed. The standard IP procedure described in the methods was extended with testing the nuclease Benzonase to test for improved extraction of chromatin-bound complexes. An extract for the co-immunoprecipitation experiment with the double-tagged cell line Tango4-Flag, Snf-V5 was then made with and without the nuclease Benzonase. The results show the Benzonase treatment improves the extraction efficiency (Figure 3.2.1). As a secondary conclusion which was not the aim of the

experiment, results indicate a potential interaction of Tango4 and Snf. However, to evaluate the extent of the background, an IP with a non-specific antibody control would be needed.

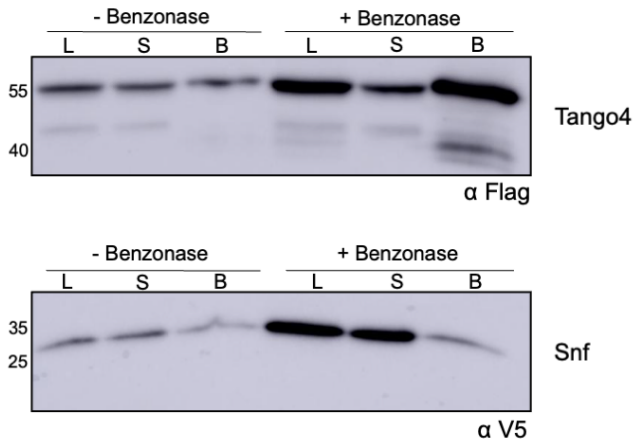


Figure 3.2.1. A co-immunoprecipitation experiment on the double-tagged Tango4-Flag, Snf-V5 cell line in the presence and absence of Benzonase with the αFlag M2. Top panel: Tango4. Bottom panel: Snf. L=Lysate, S=Supernatant, B=Bound fraction.

3.3. Crosslinking followed by Mass Spectrometry in the Tango4-Flag cell line

I used the crosslinking with formaldehyde (FA) to stabilize the transient interactions and perform a test mass spectrometry measurement.

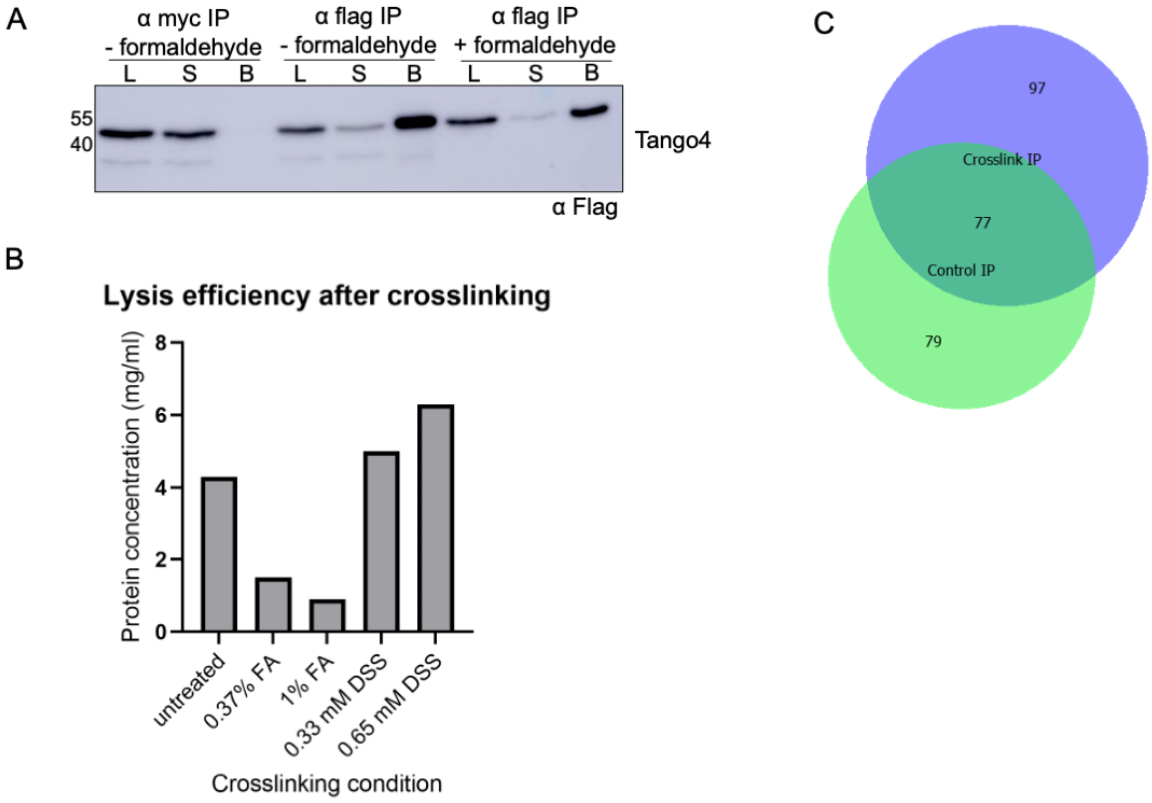


Figure 3.3.1. Crosslinking followed by mass spectrometry in the Tango4-Flag cell line. The immunoprecipitation was done with Flag and control Myc antibody. L=Lysate, S=Supernatant, B=Bound fraction (A). Lysis efficiency was measured in the sample using the Bradford assay (B). Venn diagram shows number of proteins detected in the FA-crosslinked (Flag-IP) versus non-crosslinked sample (Flag-IP) (C).

The immunoprecipitation experiment was performed on the Tango4-Flag cell line, using the Myc antibody as an antibody control. Tango4 was pulled with a Flag antibody using the magnetic beads and two conditions were tested: with and without the 0.37% Formaldehyde crosslinking. The immunoprecipitation was successful and there was no binding in the Myc sample (Figure 3.3.1., A), while there was binding in both samples with the Flag-coated beads.

Further, a Bradford assay was done to test the protein concentration of lysates in all the crosslinking samples versus controls. Formaldehyde decreased the protein concentration in both 0.37% and 1% FA samples (Figure 3.3.1., B). Crosslinker disuccinimidyl suberate (DSS) did not affect the extraction efficiency but required lengthy sample preparation.

Comparison of the formaldehyde crosslinked versus non-crosslinked sample resulting from the mass spectrometry measurement revealed a portion of proteins detected in either of the conditions (Figure 3.3.1., C), suggesting that the formaldehyde crosslinking changes meaningfully the Tango-4 bound fraction towards unspecific interactions. But primarily, it affected extraction efficiency and was dismissed from further experiments.

3.4. Testing the S2 cell line sensitivity to Zeocin

I chose two approaches to test for the *Drosophila* S2 cell line sensitivity to Zeocin treatment, a drug that chemically induces DNA double-strand breaks. The first was to perform fluorescence-activated cell sorting (FACS) to test for general cell survival. The second approach was a neutral comet assay to confirm DNA damage at the single-cell level through imaging the DNA fragmentation and quantification of the comet tails created in the electric field due to DNA migration.

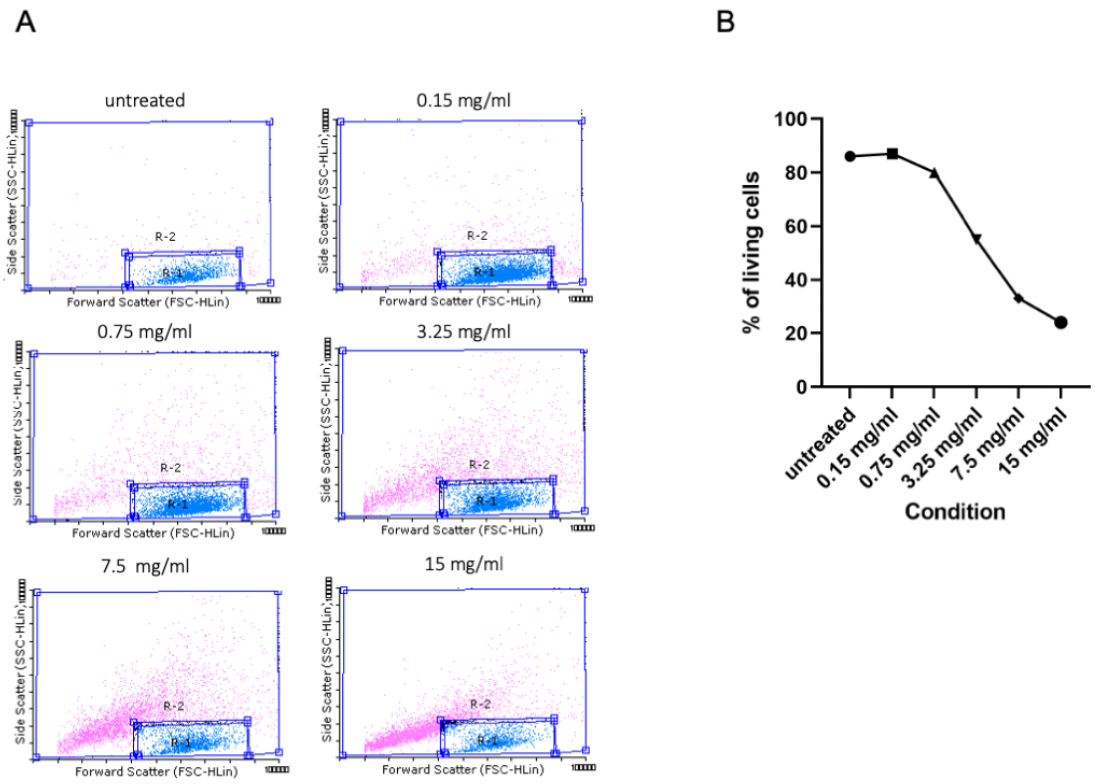


Figure 3.4.1. Fluorescence-activated cell sorting (FACS) after increasing Zeocin concentrations in *Drosophila* S2 cells for 4 hours. In A, the population of living cells (R-1) and dead cells (R-2) were counted, and this served to calculate the killing curve plotted in B.

The cells were not specially labeled for the fluorescence-activated cell sorting. However, two populations of cells were visible with no labeling as well due to their light-scattering properties (Figure 3.4.1, A), going in line with the presence of dead cells in every cell sample. I used this to select the two populations of cells and follow the development upon Zeocin treatment.

With increasing concentrations of Zeocin, the population of dead cells increased (R-2, Figure 3.4.1., A). The cells in both populations were counted, and the percentage of R-2 in overall cells was calculated, based on which the kill-curve was plotted (3.4.1., B), starting from the initial 80% percent of living cells, and IC50 concentration was

estimated as close to 3.25 mg/ml.

The selected concentrations of Zeocin treatment for 4 hours were used for the neutral comet assay to test the DNA fragmentation visible in the electric field. I was using the protocol for the comet assay that detects specifically the double-strand DNA breaks (neutral comet assay).

Minimal DNA fragmentation was visible in the untreated samples. The tails were longer with increasing Zeocin concentrations, being maximal in the 3.25 mg/ml Zeocin treatment. Interestingly, the 0.35 mg/ml and 0.75 mg/ml concentrations produced similar comets, while the big shift was achieved with the 3. 25 mg/ml Zeocin concentration (Figure 3.4.2, A). The comet tails were then quantified manually using ImageJ (Figure 3.4.2., B).

Since the IC50 Zeocin concentration based on the fluorescence-activated cell sorting also produced the longest comet tail lengths, I decided to proceed with this concentration further in the proteomic measurements.

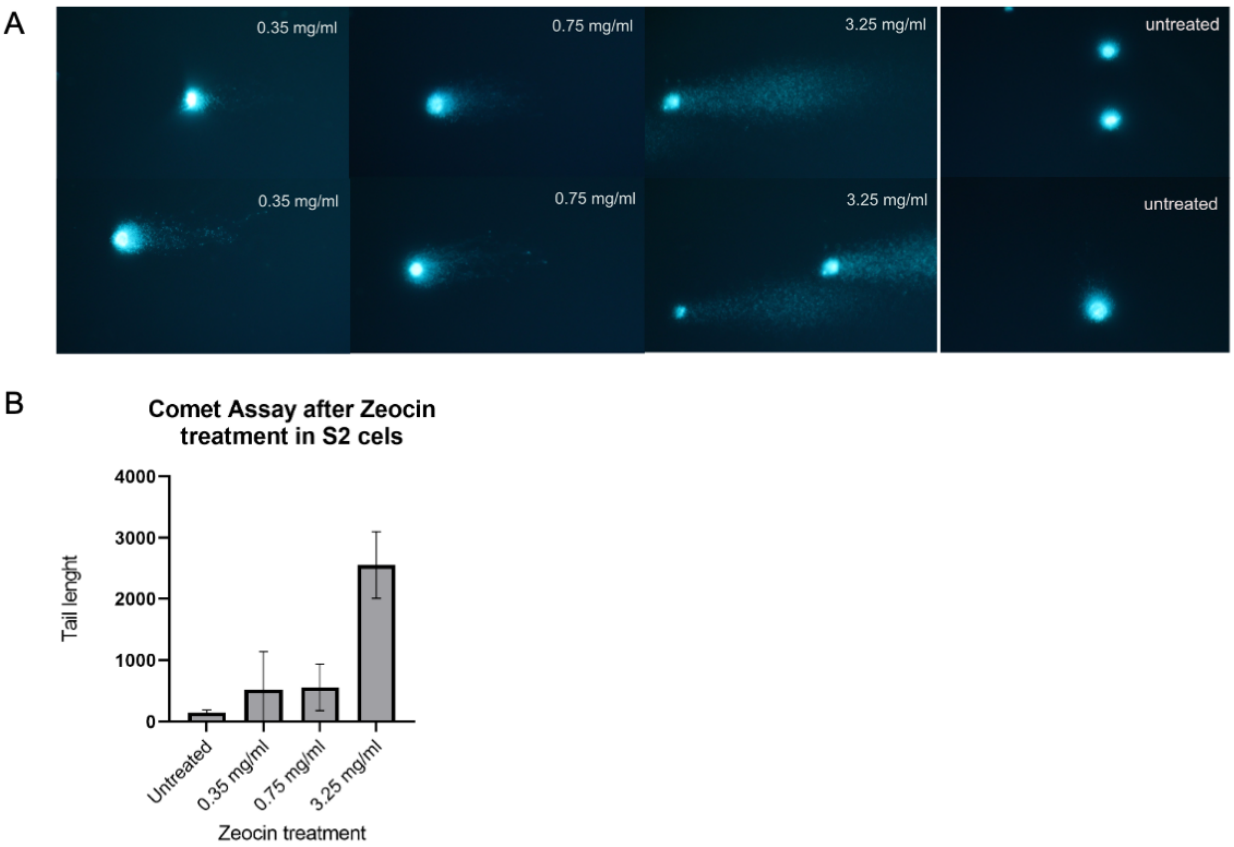


Figure 3.4.2. Neutral comet assay after increasing concentrations of Zeocin in *Drosophila* S2 cells. In A, micrographs of single Drosophila S2 cells producing comets after the treatment with Zeocin. In B, quantification of comet tail lengths in different concentrations of Zeocin (n=15).

3.5. Interactome changes of Tango4 upon induction of DNA double-strand breaks

The immunoprecipitation experiment after the 4-hour 3.25 mg/ml Zeocin treatment was analyzed in three biological replicates. The Myc antibody was used as the non-specific antibody control, while the Flag M2 antibody was used to pull down Tango4 using magnetic beads and the extraction protocol shown in the methods section. It included lysis buffers with and without urea, Benzonase the nuclease for releasing the nuclear fraction, and no crosslinking to stabilize the transient spliceosome interactions. The procedure was successful in showing no binding in the non-specific antibody, while it showed specific binding in the case of the specific antibody, in this case, Flag (Figure 3.5.1., A), with similar levels of protein in each sample (always the same amount taken from the sample, one-tenth of the lysate).

After the measurement, principal component analysis was performed to group samples based on the protein abundancies, later described in detail in the discussion. This statistical method can quickly show in 2D or 3D the correlation between all samples or all proteins across all samples. In the analysis considering the samples themselves, two distinct groups can be noticed. One is the Myc pull-down control samples, and the other group presents the Flag pull-down samples. This argues for the specificity of the immunoprecipitation samples. The principal component analysis across all detected proteins shows quickly Prp19 complex as a distinct category, implying the specificity of the bait (3.5.1., C).

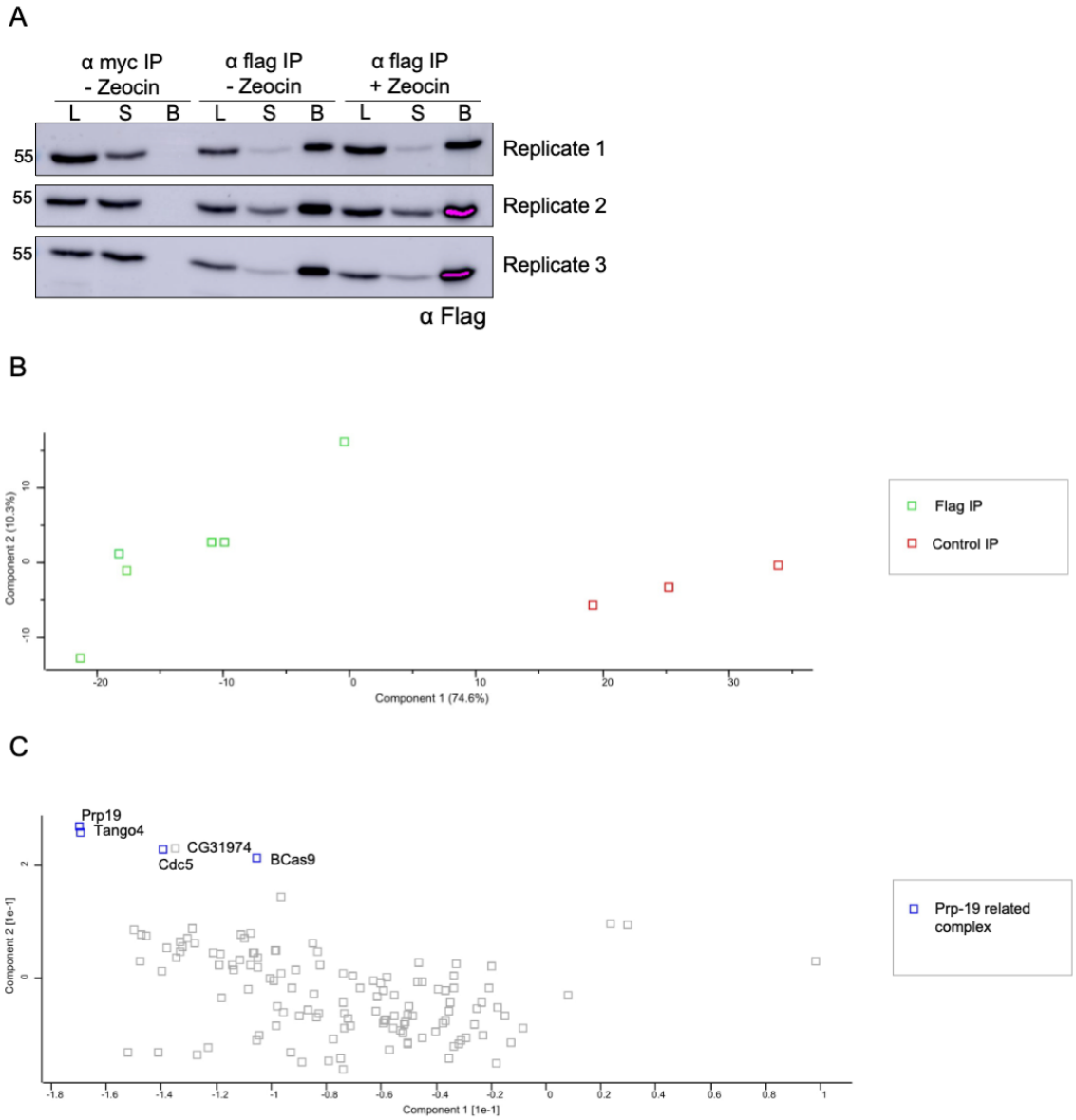


Figure 3.5.1. Mass spectrometry after Zeocin treatment in the Tango4-Flag cell line. In A, immunoprecipitation experiment in three biological replicates. L=Lysate, S=Supernatant, B=Bound fraction. In B, principal component analysis across all individual samples. In C, principal component analysis for individual proteins across all samples.

Next, hierarchical clustering of identified proteins was done, where proteins are clustered based on their expression profiles. The first cluster at the top of the tree was the Prp19 complex, with almost all of its primary components (Figure 3.5.2., A). An interesting uncharacterized protein with kinase-like domains, CG31974, was identified in the clustering as well as in the principal component analysis of all proteins (Figure 3.5.1, C). However, in the mass spectrometry measurement in the 5-3 cell line without any DNA sequence tagged, CG21974 was significantly associated with the Flag IP. It is a specific non-specific antibody interaction of the Flag antibody (data not shown).

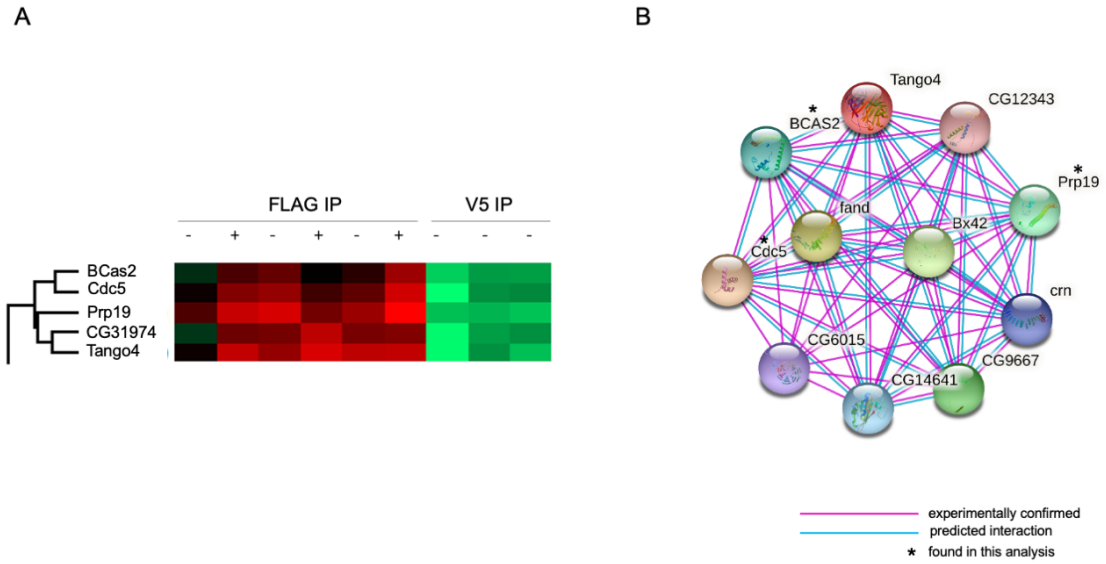


Figure 3.5.2. Mass spectrometry after Zeocin treatment in the Tango4-Flag cell line. In A, hierarchical clustering of associated proteins around Tango4. In B, STRING analysis of experimentally confirmed (pink) and predicted (blue) protein interactors of Tango4, and proteins found in this analysis are labelled with an asterisk.

I compared these results with the STRING database analysis of protein interactions and found known interaction partners, confirming that the core of the Prp19 complex has been pulled down via Tango4 using the optimized immunoprecipitation approach. I included both the experimentally confirmed and predicted interactors (Figure 3.5.2, B).

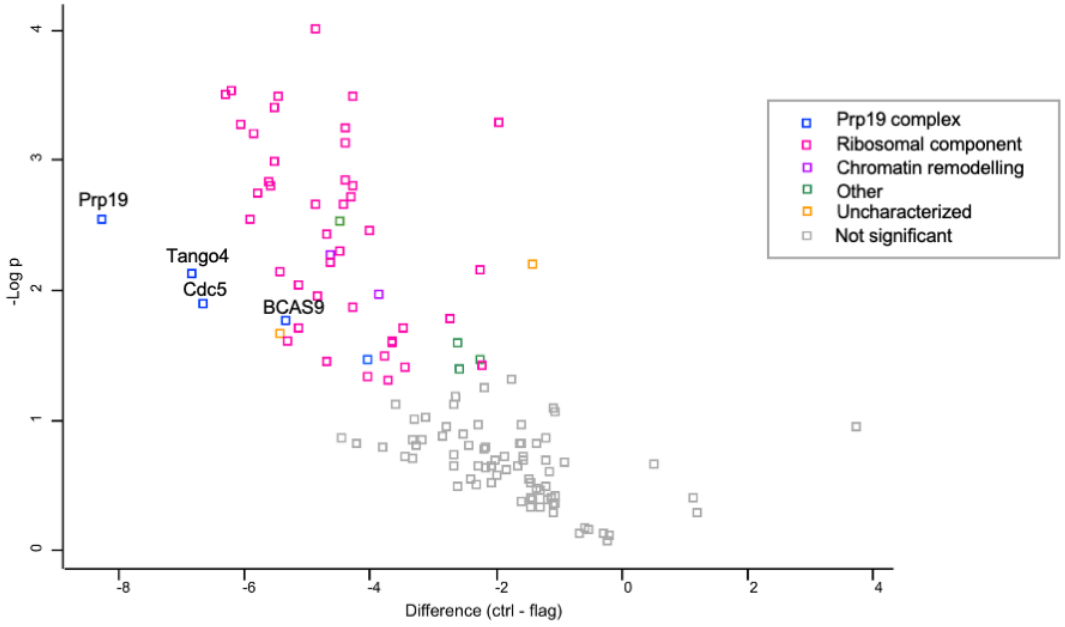


Figure 3.5.3. Mass spectrometry in the Tango4-Flag cell line. Volcano plot based on Welch's T-test shows the label-free quantification intensity log2 differences between control antibody Myc- and Flag pulled-down samples. Color-coded protein groups are presented in the legend on the left.

Label-free mass spectrometry was done instead of i.e. SILAC sample labeling before the measurement. The qualitative difference between labeled and non-labelled approaches has been discussed in [167] and [168]. I have opted for the non-labeled approach which saves time in sample preparation, but where samples are measured in individual runs.

During the analysis, I took into account the LFQ intensities, which are calculated for each protein taking into account all pair-wise peptide comparisons by the MaxQuant software [169]. However, I found the absolute intensities useful in seeing where the bait positions compared to the rest identified peptides. I then removed potential contaminants identified by MaxQuant, as well as peptides that are detected in the reverse database. I performed imputation to replace missing values for statistical purposes. The rest of the proteins were then stringently selected, requiring identification in 3 out of 3 replicates.

The volcano plot comparing the proteins detected in Myc or Flag IP reveals that the immunoprecipitation experiments are specific. The analysis identified major Tango4 interactors, including Prp19, Cdc5, and Bcas9 as strongly enriched in comparison with the control immunoprecipitation (Figure 3.5.3., A). Above the significance score are many ribosomal proteins as well, often seen in mass spectrometry measurements, which could in part explain it. Apart from those, a small group of chromatin remodeling factors was also identified in the search.

When I compared the Zeocin treated and untreated samples, there was a single statistically significant hit, a seven zinc-finger transcription factor Putzig (Figure 3.5.4). Apart from it, the analysis revealed a few more both DNA and RNA binding zinc-finger proteins, as well as chromatin remodeling factors as potential interactome differences due to DNA double-strand breaks.

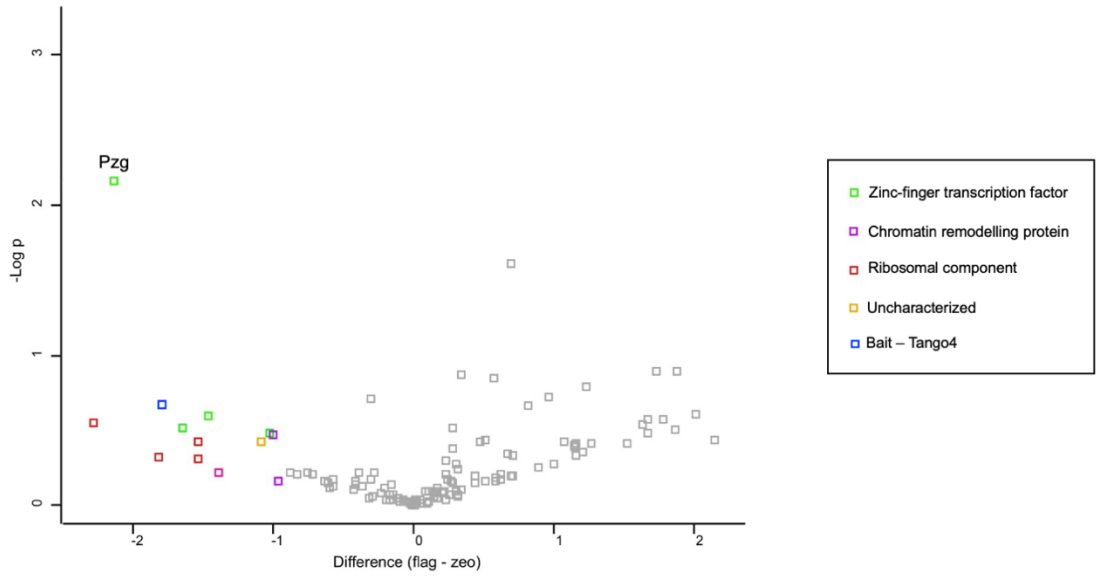


Figure 3.5.4. Mass spectrometry in the Tango4-Flag cell line. A volcano plot based on Welch's T-test shows the label-free quantification intensity differences between Zeocin-treated and untreated samples. Color-coded protein groups are presented in the legend on the left.

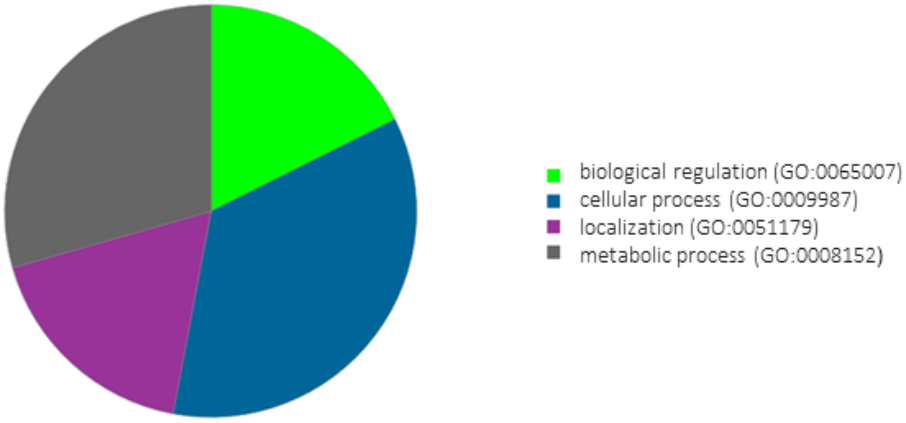
Given the fact that Putzig had already been identified in a separate RNAi screen as a factor important for the efficient generation of small RNAs at the DNA break, I decided to pursue further by tagging Putzig and doing non-proteomic experiments with the protein.

Also given that mass spectrometry revealed transcription factors binding Tango4 upon Zeocin treatment, I performed a domain search to see in detail what type of transcription factors they are (Figure 3.5.5.). Domain search was done using InterPro [170] which classifies the proteins based on their sequence into superfamilies and predicts protein domains. It combines the information from 13 different databases into one single output (CDD, Prosite, CATH, Panther, HAMAP, Pfam, MobiDB Lite, PIRSF,

PRINTS, SFLD, SUPERFAMILY, SMART, and TIGRfams). Three proteins showing a trend for increased association with Tango4 upon Zeocin treatment have the zinc-finger domains.

Figure 3.5.5. InterPro domain search for three RNA-polymerase interacting proteins revealed in the Tango4 proteomic analysis upon Zeocin treatment: CG8487, CG4360 and Pzg. Protein length indicated at the top; domain super-family from the InterPro base annotated with Znf_CH2H_type in light green, different colors below present all Zinc finger C2H2 type domains identified through different databases: dark blue=Pfam, purple=SMART, pink=PROSITE profiles, dark green=PROSITE patterns.

A



B

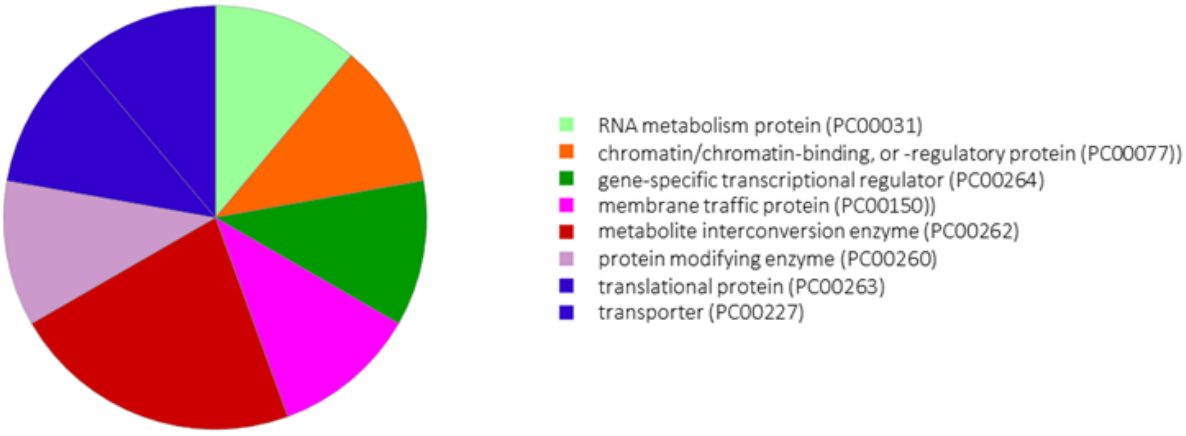


Figure 3.5.6. Panther GO-Term analysis of peptides with detected phospho-ST sites in mass spectrometry measurement after the Tango4-Flag pull down. GO Term analysis classified 16 proteins. From those 16 proteins, further classified for biological process (A) protein class (B).

As DNA damage triggers a series of phosphorylation events, I performed a first analysis in MaxQuant of the phosphorylated peptides upon DNA damage and further did a PANTHER GO-Term analysis to classify them to understand which functions they have in the cell (Figure 3.5.6.). I used the PANTHER classification system (Protein ANalysis THrough Evolutionary Relationships) to classify the retrieved hits [171]. PANTHER classifies according to different categories, such as molecular function, biological pathway, etc [172] [173].

3.6. Relative expression of Putzig after DNA damage with Zeocin

Since the interactome analysis suggested that Putzig may play a role in the response to DNA damage, I wanted to test whether its expression increases upon a 4-hour Zeocin treatment. For this purpose, I designed primers for quantitative PCR (qPCR) which I used to amplify cDNA derived from cells that had been treated with Zeocin at the standard 3.25 mg/ml concentration for four hours. I wanted to test if the increased association with the spliceosome we detected may simply reflect an increased overall abundance of Putzig after DNA damage. A gene used for normalization was RP49. Putzig expression did not increase after the Zeocin treatment (Figure 3.6.1.).

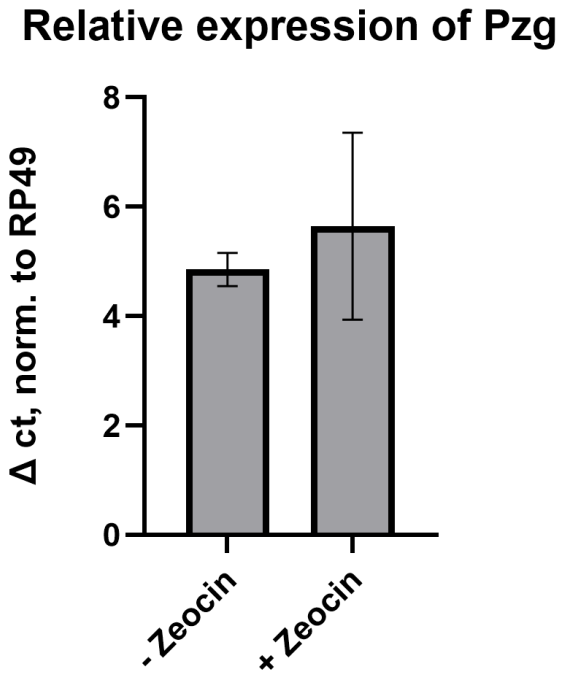


Figure 3.6.1. Quantitative PCR (qPCR) showing relative Putzig expression normalized to RP49 with and without 3.25 mg/ml Zeocin treatment for four hours.

3.7. Detection of Putzig interactome changes upon DNA damage

To understand the Putzig proteome after the DNA damage and confirm the spliceosome interaction in the reverse orientation, I have used CRISPR/Cas9 once again to genomically edit the *Pzg* gene to express an epitope tag at the C-terminus in *Drosophila* cultured cells. I selected the cells with Blasticidin in two rounds and seven days after the second round of selection I confirmed the tag in our *Drosophila* S2 cells with a western blot. Tagged Putzig shows a big degradation pattern that one could notice in a publication detecting Putzig [174] on the western blot level (Figure 3.7.1, A). In the publication, an antibody against the endogenous untagged protein was used.

I have prepared mass spectrometry samples according to the standard procedure described in the methods section and performed a MASCOT [175] analysis in addition to checking for proteins with the highest intensity and confirming the bait and thus the cell line I have made. The MASCOT analysis was added to the workflow to add an extra database check whether Putzig was a bait, as there were doubts about the Western Blot regarding its size, which did not fully match the estimation according to its sequence on FlyBase. The analysis revealed Putzig as the first hit with the highest absolute intensity. I proceeded with mass spectrometry measurement and initially did a hierarchical clustering analysis, revealing the first cluster of Putzig and its well-known interacting factor Chromator (the relationship and complexes of Putzig are in detail described in the discussion section). The principal component analysis did not differentiate strongly between control and Flag IP, implying lower IP sample quality (Figure 3.7.1).

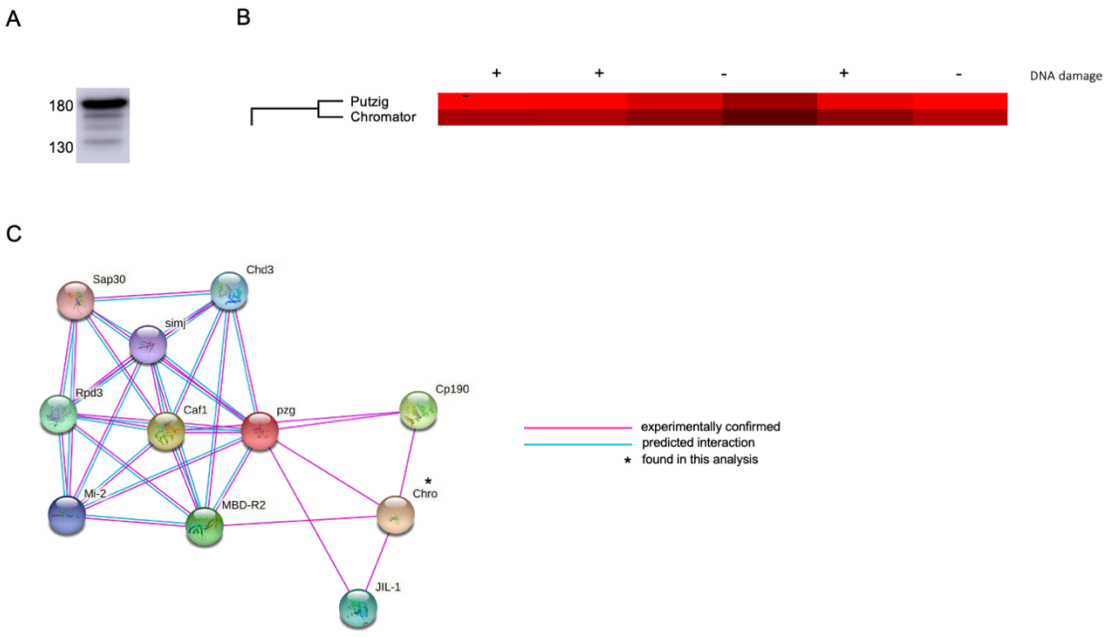


Figure 3.7.1. Mass spectrometry in Putzig-Flag cell line. In A, western blotting against the Flag antibody confirms the typical Putzig band pattern. In B, a hierarchical clustering analysis reveals the Putzig clustering with its interactor protein Chromator. In C, STRING analysis of experimentally confirmed (pink) and predicted Putzig interactors (blue). Identifiers: pzg=Putzig, chro=Chromator.

STRING analysis revealed the experimentally confirmed and revealed Putzig interactors such as Chromator, but interestingly did not reveal many of the interaction partners such as HP1c complex or TRF2/DREF [176].

Mass spectrometry analysis revealed the binding partners of Putzig after DNA damage with Zeocin. Putzig is shown to bind Histone H4 more in the case of massive DNA damage, which could go in line with its chromatin remodeling functions, which are the best described Putzig factors (Figure 3.7.2.). Putzig interacts with several partners to ensure the chromatin accessibility and transcription.

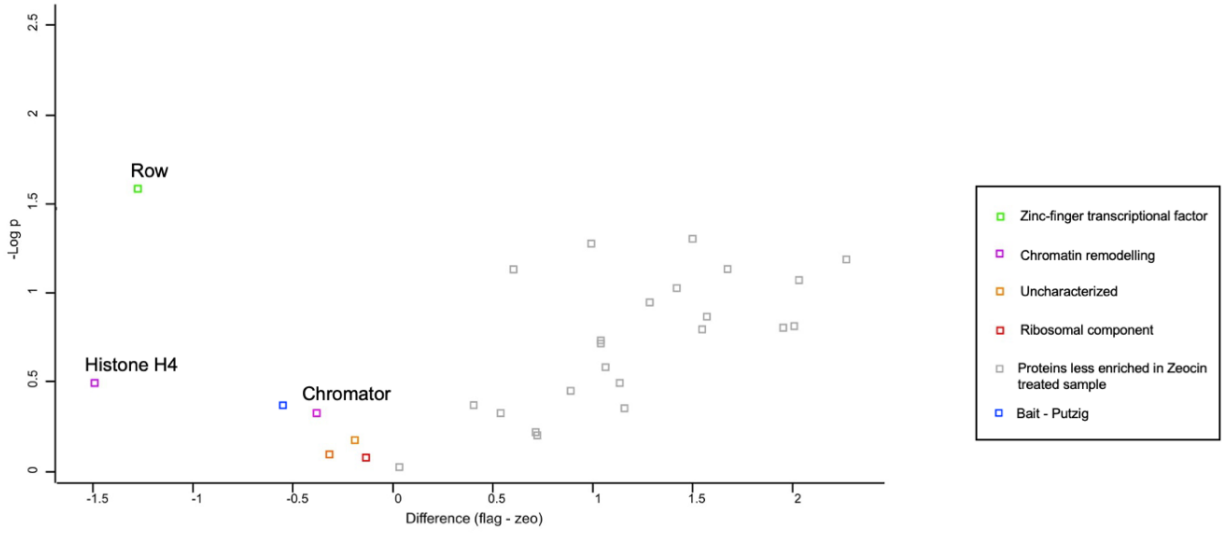


Figure 3.7.2. Mass spectrometry in the Putzig-Flag cell line. A volcano plot based on Welch's T-test reveals Putzig interactors that bind it with bigger LFQ intensity upon DNA damage with 3.25 mg/ml Zeocin after four hours. The color code on the side depicts the protein category.

Another candidate binding Putzig after DNA damage revealed by the same experiment (3.7.2.) was Relative of Woc (Row), a member of the HP1c complex together with HP1c and Woc. The interaction of Row and Putzig was shown before [177]. However, the change in LFQ intensities between the treated versus non-treated samples was small. I nonetheless wanted to further test this complex' role in the antisense transcription that happens as a result of DNA break.

Given the very stringent pre-requirements I used for the analysis in the statistical program Perseus (as described in 3.5.), in an already very limited proteomic analysis after the immunoprecipitation compared to analyses in the i.e. whole-cell lysates, important interactors may be lost. This was especially important to understand in looking for Tango4 in the Putzig proteomic data. Tango4 does not pass the Perseus filter as it is not detected in 3 out of 3 replicates.

However, it is important to mention that Putzig does bind Tango4 according to mass spectrometry results, although inconsistently across replicates. Interestingly, the Tango4 hit does not show up as one single identifier, but rather several small peptide fragments that belong to Tango4, according to my BLAST analysis (3.7.3.).

When comparing absolute intensities merged for all Tango4 peptide fragments that MaxQuant calculated across all samples, Tango4 is detected in the list of proteins, in the 54th position of 188 identified proteins. Putzig has the first position in the absolute intensities list, followed directly by Chromator.

Query: A9YGS7_DROME
Subject: Q9VYQ9_DROME (Tango4)

```
Query:   1          RPAEKTAPLVTGTHTSLVRASLANSNGDKSANGAIASHLQLIPKKAPSIPKPKWHAPWKL   60
                    RPAEKTAPLVTGTHTSLVRASLANSN D SANGAIASHLQLIPKKAPSIPKPKWHAPWKL
Subject: 105        RPAEKTAPLVTGTHTSLVRASLANSNADMSANGAIASHLQLIPKKAPSIPKPKWHAPWKL   164

Query:   61         SRVISGHLGWVRCIAVEPGNEWFATGAGDRVIKIWDLASGKLKLSLTGHVSTVRGVAVST   120
                    SRVISGHLGWVRCIAVEPGNEWFATGAGDRVIKIWDLASGKLKLSLTGHVSTVRGVAVST
Subject: 165        SRVISGHLGWVRCIAVEPGNEWFATGAGDRVIKIWDLASGKLKLSLTGHVSTVRGVAVST   224

Query:   121        KHPYLFSCGEDRQVKCWDLEYNKVIRHYHGHLSAVYSLALHPTIDVLATSGRDSTARIWD   180
                    KHPYLFSCGEDRQVKCWDLEYNKVIRHYHGHLSAVYSLALHPTIDVLATSGRDSTARIWD
Subject: 225        KHPYLFSCGEDRQVKCWDLEYNKVIRHYHGHLSAVYSLALHPTIDVLATSGRDSTARIWD   284

Query:   181        MRTKANVHTLTGHTNTVASVVAQATNPQIITGSHDSTVRLWDLAAGKSVCTLTNHKKSVR   240
                    MRTKANVHTLTGHTNTVASVVAQATNPQIITGSHDSTVRLWDLAAGKSVCTLTNHKKSVR
Subject: 285        MRTKANVHTLTGHTNTVASVVAQATNPQIITGSHDSTVRLWDLAAGKSVCTLTNHKKSVR   344
```

Query: F6J9X0_DROME
Subject: Q9VYQ9_DROME (Tango4)

```
Query:   1          GARPAEKTAPLVTGTHTSLVRASLANSNGDMSANGAIASHLQLIPKKAPSIPKPKWHAPW   60
                    GARPAEKTAPLVTGTHTSLVRASLANSN DMSANGAIASHLQLIPKKAPSIPKPKWHAPW
Subject: 103        GARPAEKTAPLVTGTHTSLVRASLANSNADMSANGAIASHLQLIPKKAPSIPKPKWHAPW   162

Query:   61         KLSRVISGHLGWVRCIAVEPGNEWFATGAGDRVIKIWDLASGKLKLSLTGHVSTVRGVAV   120
                    KLSRVISGHLGWVRCIAVEPGNEWFATGAGDRVIKIWDLASGKLKLSLTGHVSTVRGVAV
Subject: 163        KLSRVISGHLGWVRCIAVEPGNEWFATGAGDRVIKIWDLASGKLKLSLTGHVSTVRGVAV   222

Query:   121        STKHPYLFSCGEDRQVKCWDLEYNKVIRHYHGHLSAVYSLALHPTIDVLATSGRDSTARI   180
                    STKHPYLFSCGEDRQVKCWDLEYNKVIRHYHGHLSAVYSLALHPTIDVLATSGRDSTARI
Subject: 223        STKHPYLFSCGEDRQVKCWDLEYNKVIRHYHGHLSAVYSLALHPTIDVLATSGRDSTARI   282

Query:   181        WDMRTKANVHTLTGHTNTVASVVAQATNPQIITGSHDSTVRLWDLAAGKSVCTLTNHKKS   240
                    WDMRTKANVHTLTGHTNTVASVVAQATNPQIITGSHDSTVRLWDLAAGKSVCTLTNHKKS
Subject: 283        WDMRTKANVHTLTGHTNTVASVVAQATNPQIITGSHDSTVRLWDLAAGKSVCTLTNHKKS   342

Query:   241        VR   242
                    VR
Subject: 343        VR   344
```

Query: B4YWZG_DROME
Subject: Q9VYQ9_DROME (Tango4)

```
Query:   1          VTGTHTSLVRASLANSNGDMSANGAIASHLQLIPKKAPSIPKPKWHAPWKLSRVISGHLG   60
                    VTGTHTSLVRASLANSN DMSANGAIASHLQLIPKKAPSIPKPKWHAPWKLSRVISGHLG
Subject: 114        VTGTHTSLVRASLANSNADMSANGAIASHLQLIPKKAPSIPKPKWHAPWKLSRVISGHLG   173

Query:   61         WVRCIAVEPGNEWFATGAGDRVIKIWDLASGKLKLSLTGHVSTVRGVAVSTKHPYLFSCG   120
                    WVRCIAVEPGNEWFATGAGDRVIKIWDLASGKLKLSLTGHVSTVRGVAVSTKHPYLFSCG
Subject: 174        WVRCIAVEPGNEWFATGAGDRVIKIWDLASGKLKLSLTGHVSTVRGVAVSTKHPYLFSCG   233

Query:   121        EDRQVKCWDLEYNKVIRHYHGHLSAVYSLALHPTIDVLATSGRDSTARIWDMRTKANVHT   180
                    EDRQVKCWDLEYNKVIRHYHGHLSAVYSLALHPTIDVLATSGRDSTARIWDMRTKANVHT
Subject: 234        EDRQVKCWDLEYNKVIRHYHGHLSAVYSLALHPTIDVLATSGRDSTARIWDMRTKANVHT   293

Query:   181        LTGHTNTVASVVAQATNPQIITGSHDSTVRLWDLAAGKSVCTLT   224
                    LTGHTNTVASVVAQATNPQIITGSHDSTVRLWDLAAGKSVCTLT
Subject: 294        LTGHTNTVASVVAQATNPQIITGSHDSTVRLWDLAAGKSVCTLT   337
```

Figure 3.7.3. BLAST analysis of various hits identified in proteomic analysis where Putzig served as a bait. The selected peptides show >99% sequence complementarity to Tango4.

3.8. Mass spectrometry independent validation of Tango4-Putzig interaction

To try to confirm the interaction of Tango4 and Putzig, I have created a double-tagged cell line using CRISPR/Cas9, which has a V5 tag on Tango4 and a Flag tag on Putzig (Figure 3.8.1). Results could not confirm consistently the interaction of these two proteins.

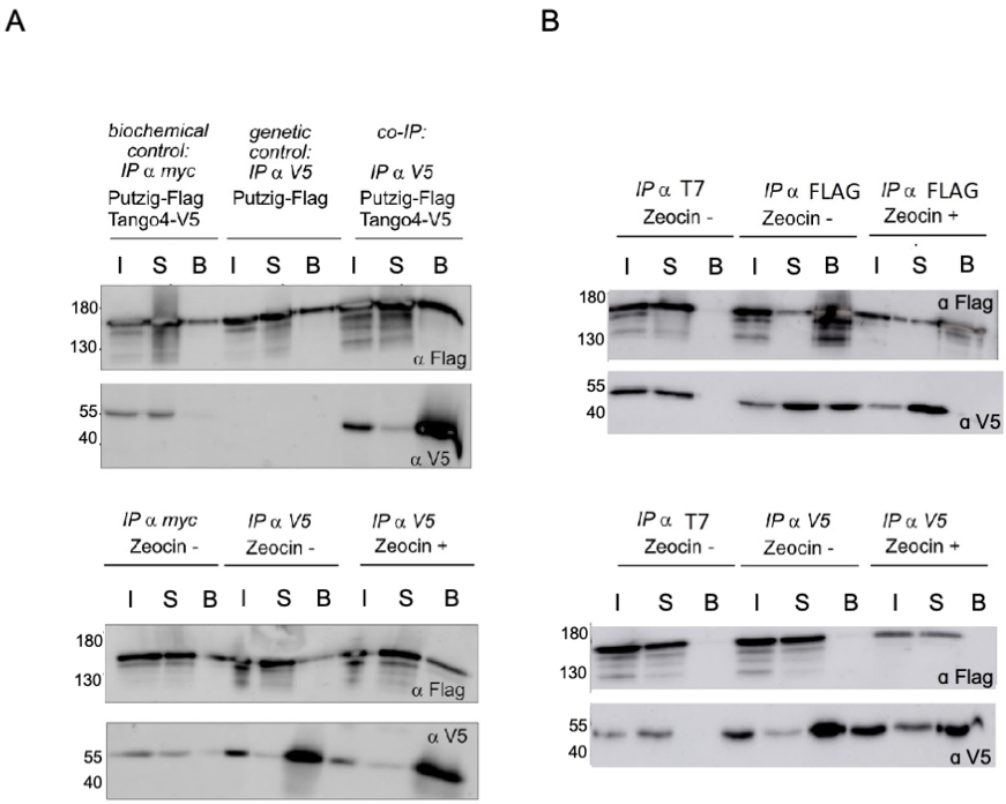


Figure 3.8.1. Co-immunoprecipitation using double-tagged cell line Putzig-Flag, Tango4-V5. In A, the Myc antibody served as a control antibody. In B, the T7 pol antibody served as a control antibody. L=Lysate, S=Supernatant, B=bound fraction.

3.9. Reporter-based validation of Putzig's implication in damage-induced siRNA biogenesis

I wanted to indirectly test the effect of Putzig knockdown with three independent siRNA constructs on the silencing machinery of the cells and their ability to produce siRNAs. For this, I have used a luciferase-based reporter system, where I transfected cells with either linear or circular variants of the Renilla luciferase. The ends of the linearized plasmid trigger the production of the siRNAs in the cell, which can then silence in-trans the circular Renilla luciferase. For transfection control, Firefly luciferase was transfected as well. The expectation was to reproduce the published result where Putzig showed an effect on the expression of diRNAs [114].

I have designed and made three different dsRNAs against Putzig exons that I used before the luciferase assay. Results show the increase of signal in all three constructs compared to the GFP dsRNA knockdown. The increased signal essentially means that the silencing machinery is perturbed, therefore the silencing of the circular Renilla luciferase is not as efficient as in the controls. The experiment was done in three biological replicates, the measurement was done after a single 72-hour knockdown with respective dsRNAs, after which the luciferase plasmids were incubated for another 72 hours before the luciferase activity assay. The detailed protocol is described in the methods section (2.2.24.). Results show a clear effect of Pzg on the luciferase signals (Figure 3.9.1.)

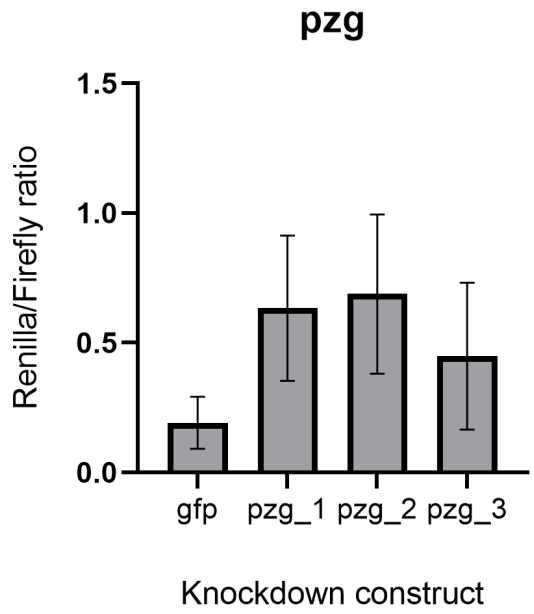


Figure 3.9.1. Luciferase-based reporter system for evaluating the functionality of the silencing machinery in the cell. Circular and linearized *Renilla* luciferase were transfected together with *Firefly* luciferase which served as a transfection control.

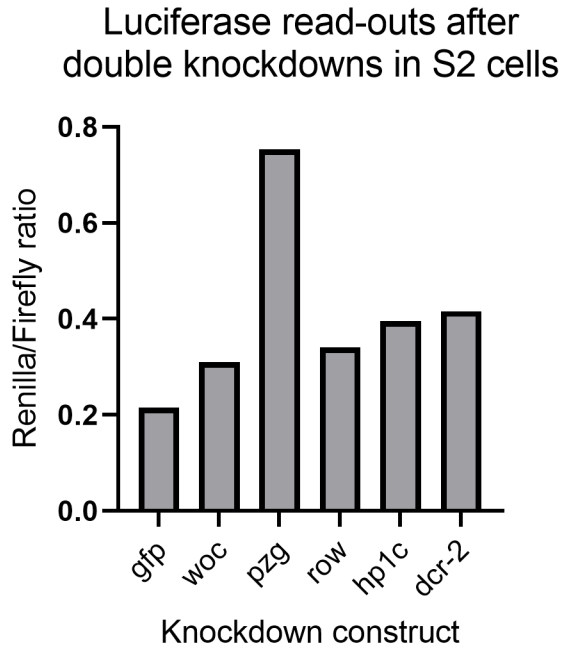


Figure 3.9.2. A single experiment of the luciferase-based reporter system for evaluating the functionality of the silencing machinery in the cell after a double knockdown with respective dsRNA constructs. Circular and linearized *Renilla* luciferase-encoding plasmids were

transfected together with *Firefly* luciferase-encoding plasmid that served as a transfection control.

In order to maximize the knockdown efficiency, I have tried a double-knockdown with cells that were made for deep sequencing and performed the same luciferase-based reporter assay to asses the silencing machinery in the S2 cells (3.9.2.).

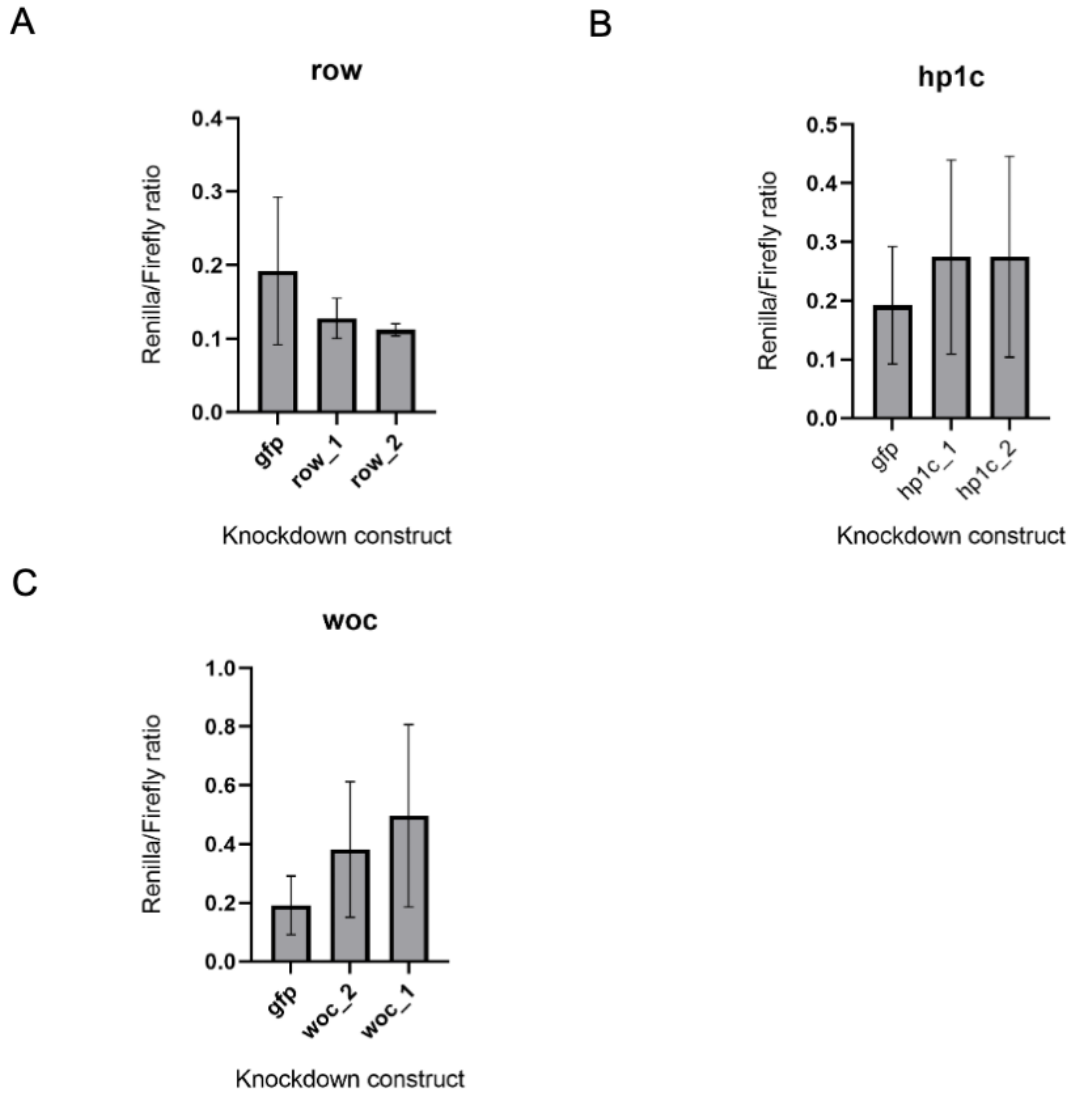


Figure 3.9.3. Luciferase-based reporter system for evaluating the functionality of the silencing machinery in the cell after the knockdown with respective dsRNA constructs of the HP1c complex. Circular and linearized *Renilla* luciferase-encoding plasmid were transfected together with *Firefly* luciferase-encoding plasmid that served as a transfection control.

In the double knockdown experiment, the absence of Putzig was more successful in impairing the silencing machinery of the cell. The absence of the rest of the factors, including the HP1c complex components did not show any effect. I have then also tested two different constructs for each HP1c complex factor and performed the same luciferase assay. None of the factors has shown an effect on the silencing machinery of the cell, with Woc having potentially the strongest effect. Taken together, Putzig consistently shows an effect on the efficient expression of DNA break-induced siRNAs, while the HP1c complex does not show a reproducible effect.

3.10. Deep sequencing of small RNAs at the DNA break in CG2493

This experiment aimed to evaluate the effect of Putzig and the HP1c complex on the generation of small RNAs at an induced DNA break at the CG2493 gene, which is an intronless gene whose transcription unit partially overlaps with an intron-containing gene CDC23. The guide RNA that leads Cas9 to the desired site in CG2493 was expressed from a plasmid (pPD02), rather than a PCR amplicon, to facilitate transfections at a somewhat larger scale. Before the CRISPR/Cas9 cut, two consecutive 72-hour knockdowns were performed with dsRNAs targeting *Pzg, Woc, Row, HP1c,* and *Dcr-2*. To verify efficient cleavage at CG2493, a T7 endonuclease assay was performed, where mismatches which arise due to the repair of the cut DNA are recognized and cleaved by the T7 endonuclease. As shown in Fig. 3.10.1., the T7 endonuclease cut every sample except for the uncut control.

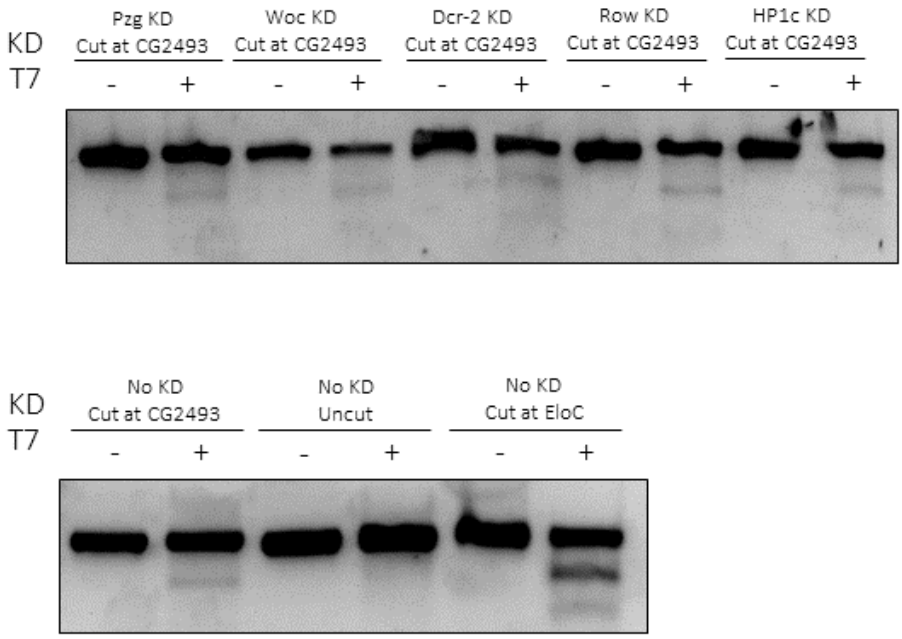


Figure 3.10.1. T7 endonuclease assay to assess the DNA cut induced by CRISPR/Cas9 in the CG2493 gene. The indicated knockdowns were carried out before the genomic cut at CG2493. A negative control for the experiment was the sample without the cut at CG2493, while the positive control for the assay was the already performed and confirmed T7 cut in the EloC gene from another experiment. KD=knockdown, T7=T7 endonuclease.

I then proceeded to make the small RNA libraries for short-read Illumina deep sequencing. Sample multiplexing was done for all the individual libraries to be pooled together and later sequenced at the same time in the Illumina instrument. This is achieved by introducing individual barcode sequences [178] to each ligated library sample, which later serves for sample sorting in the demultiplexing process where the samples can be separated for the analysis that will follow. The log file from the demultiplexing suggests a roughly similar amount of reads in each library sample, suggesting equal mixing when libraries were pooled together. The reads per individual library range from 21 to 37 million reads (Table 3.10.1.).

Index no.	Knockdown construct	All sorted reads	Genome matching after size-selection
i1	Pzg	23367813	5612633
i3	Woc	27829321	8867988
i4	Dcr-2	36611179	2494043
i5	HP1c	25884582	4856873
i6	no knockdown	21933449	8052427
i8	Row	29586481	5831423

Table 3.10.1. The total number of reads together with all genome-matching reads after size-selection for each sample used for deep sequencing.

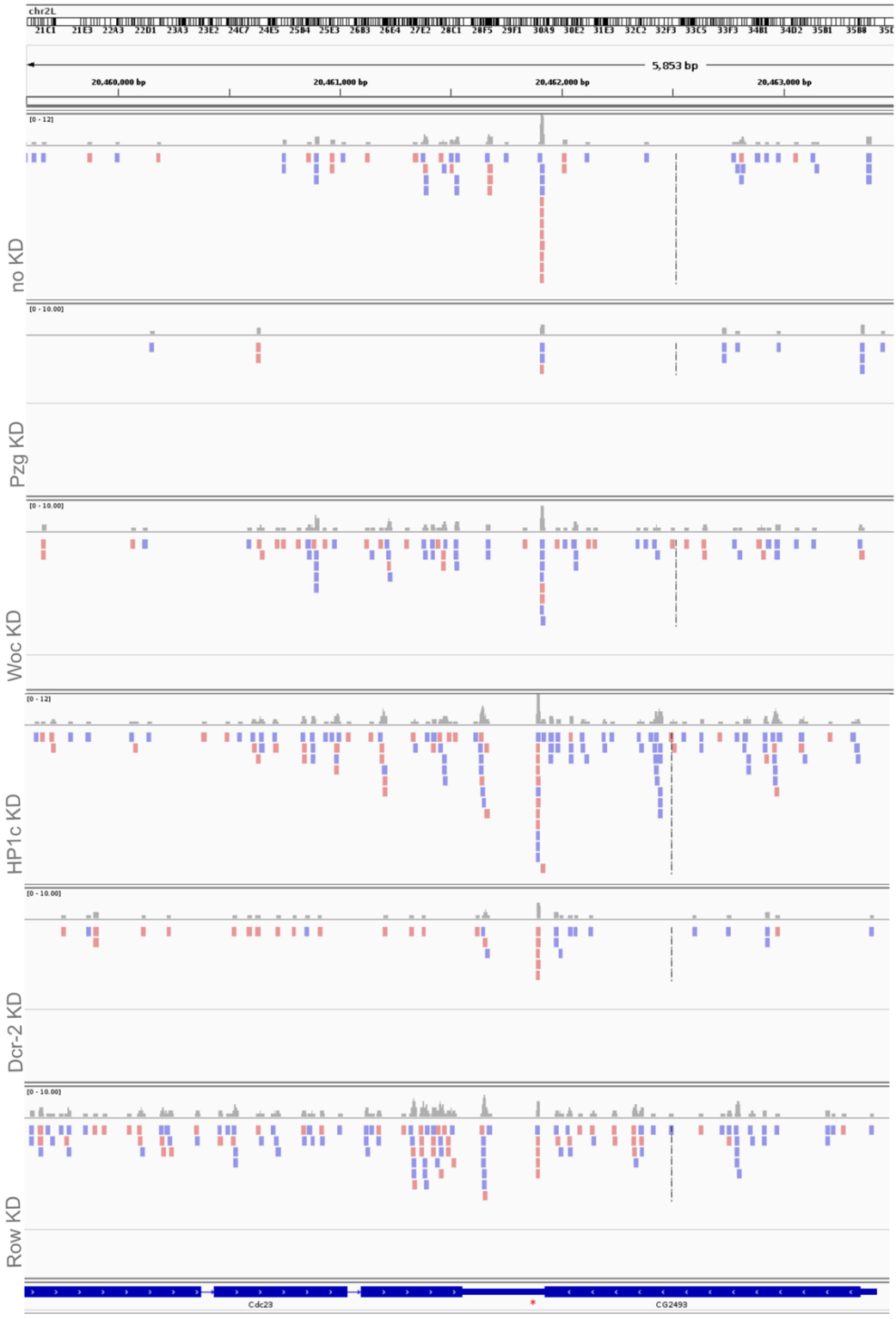


Figure 3.10.2. IGV (Integrated Genome Viewer) was used to align the sense and antisense reads to the CG2493 gene where the genomic cut using CRISPR/Cas9 was made (marked

with a red asterisk). The alignment was made with *Drosophila* (dm3) as a reference genome without corrections for the sequencing depths. The corresponding knockdown is indicated on the left side, with gene position and orientation at the bottom.

The reads were aligned using the BOWTIE [179], which uses the Burrows-Wheeler transform (BWT) technique. A very conservative BOWTIE analysis where no mismatch was allowed was performed to assure the specificity. Biochemically added adapters are also bioinformatically removed and the size range in nucleotides selected was 19-24 nt, to allow for slightly longer and slightly shorter RNAs than the typical 21 nt Dcr2 product. The reads in the library were normalized to the total number of the genome-matching reads.

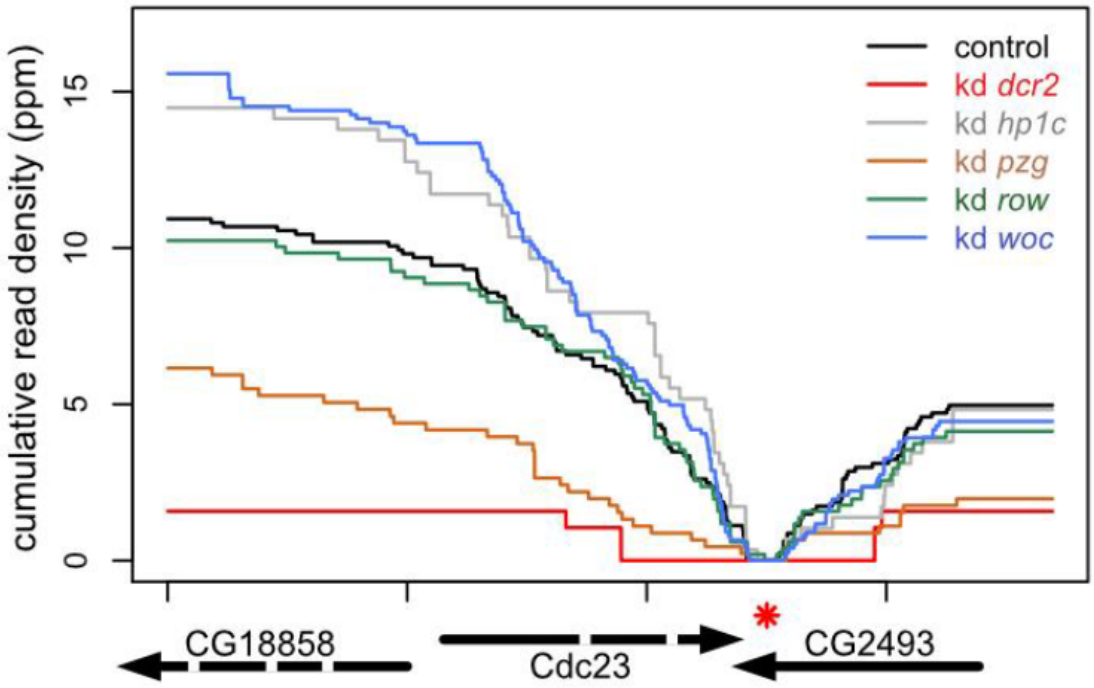


Figure 3.10.3. Cumulative antisense reads at the CG2493 gene resulted at the DNA break after various double knockdowns. Different knockdowns are color-coded in the legend on the right. Dcr-2 and Pzg knockdowns show a clear reduction in small RNA reads at *CG2493* and *Cdc23*.

Putzig once again shows an effect on the expression of small RNAs, now directly on the expression of small RNAs at the single DNA break, therefore confirming its role in the process (Figure 3.10. 2, 3.10.3.). This is an addition to results shown in the luciferase-based assays in this work, as well as the RNAi screen that also showed Putzig as a potential candidate in the process [114]. This argues that Putzig is not only required for the small RNAs resulting from a linearized plasmid but also for siRNAs arising from the chromosomal DNA in the natural chromatin context.

3.11. Proteomic analysis of the Snf interactome after DNA damage with Zeocin

The U1/U2 spliceosome complex protein Snf (sans fille) is another interesting candidate for its role in DNA damage expression of small RNAs, as found in the genome-wide RNAi screen in *Drosophila*. I edited its DNA sequence initially with CRISPR/Cas9 to express a Flag tag at the C-terminus of the protein, but later also made a double-tagged cell line with Tango4, where I tagged it with a V5 tag at the C-terminus.

I treated the Snf-Flag cell line with Zeocin in the concentration of 3.25 mg/ml for 4 hours, after which the immunoprecipitation was done under the conditions described in the methods section.

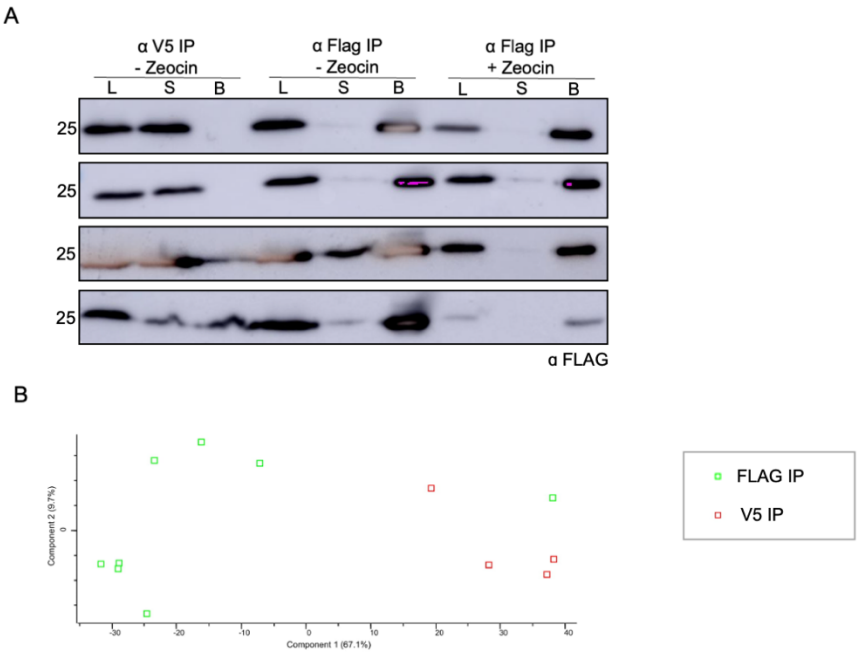


Figure 3.11.1. Mass spectrometry analysis of Snf-Flag cell line. In A, immunoprecipitation experiment in four biological replicates. V5 was used as an antibody control, while the Zeocin-treated versus non-treated condition was made together with the specific Flag antibody. L=lysate, S=supernatant, B=bound fraction. In B, principal component analysis across all samples. Flag IP in green, V5 IP in red.

The principal component analysis clearly showed the specificity of the IP, with almost all the Flag IP samples making a single group, leaving the V5 IP samples (where V5 was used as an antibody in immunoprecipitation in a cell line without a V5 tag as a control) as a separate entity too. Only one Flag sample co-localized with the V5 sample group (Figure 3.11.1).

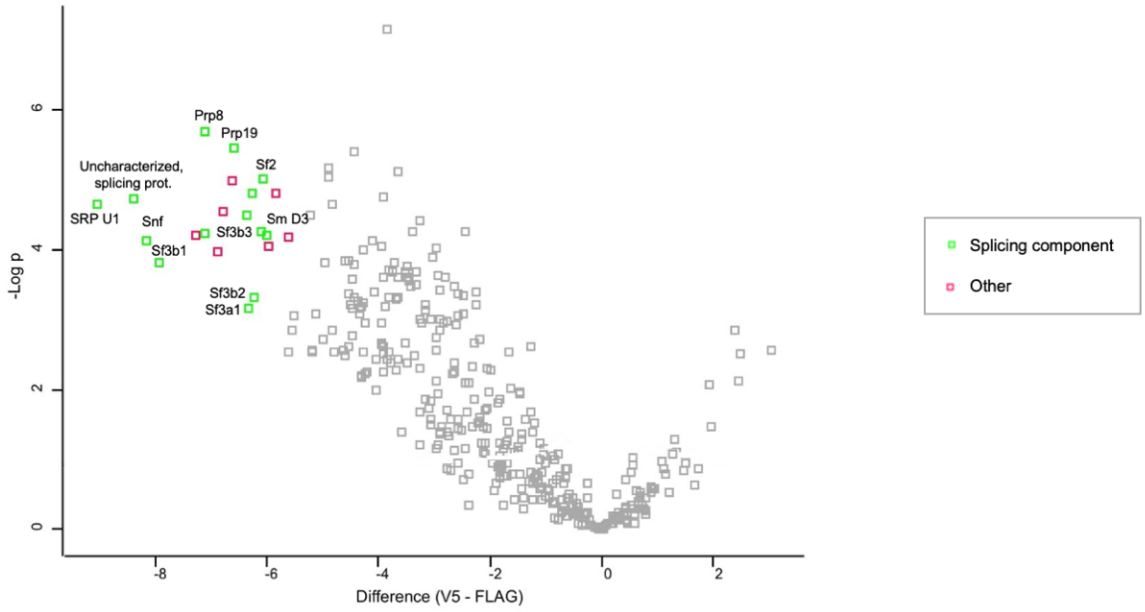


Figure 3.11.2. Mass spectrometry analysis of Snf-Flag cell line. Volcano plot based on Welch's T-test showing log2 difference in LFQ intensities between the control V5 antibody IP and Flag antibody IP sample. Snf as a bait reveals interactions with predicted partners from the U1/U2 complex as well as interactions with the Prp19 complex.

Mass spectrometry on the Snf-Flag cell line confirmed well-known interactors of Snf, such as Sf3b1, SRP U1, and Sm D3 (Figure 3.11.2). Both absolute intensities (data not shown) and differences in LFQ intensities argue for the specificity of the immunoprecipitation and confirm the successful pull-out of the U1/U2 spliceosome complex. I have performed the STRING analysis, which confirmed one Snf interactor found in my data (Figure 3.11.3).

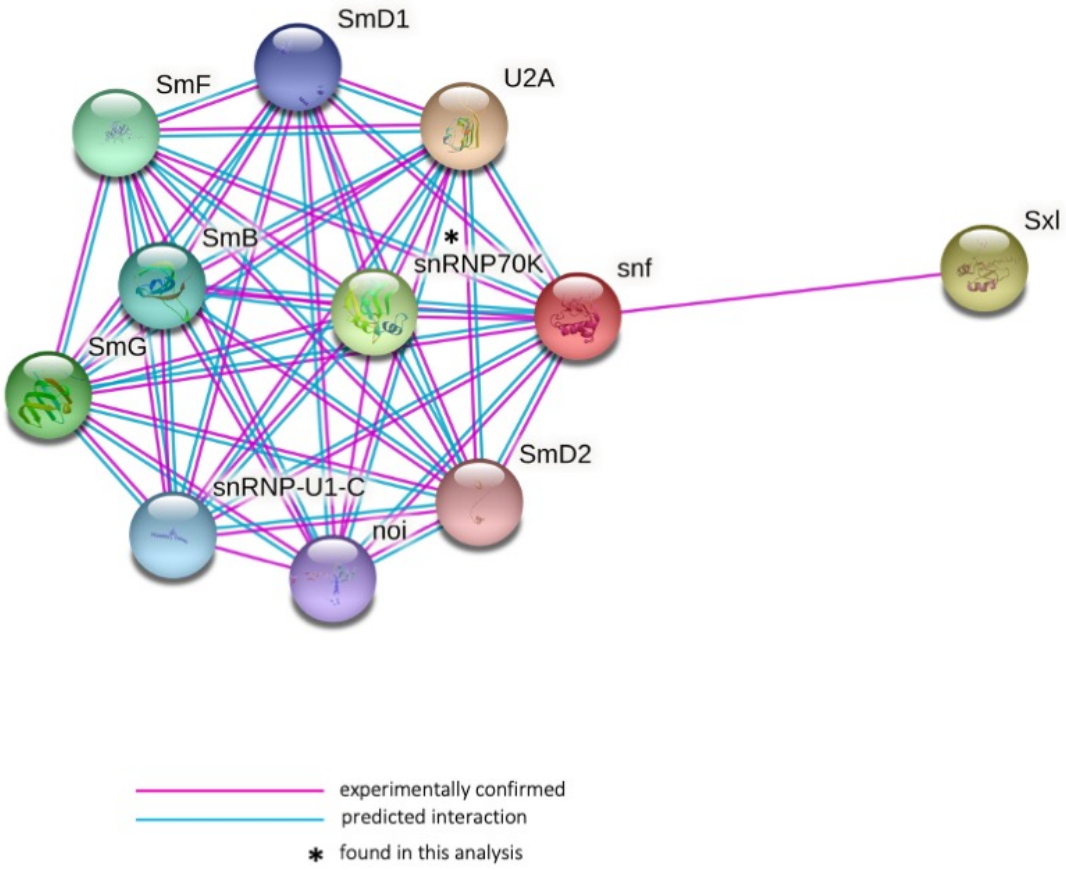


Figure 3.11.3. STRING analysis of experimentally confirmed (pink) and predicted (blue) Snf interactors. Found interactors in this analysis were marked with an asterisk.

I further compared the Zeocin treated versus untreated samples and the analysis surprisingly revealed a certain amount of mitochondrial factors and other metabolism-related proteins as interaction partners of Snf (Figure 3.11.4). These proteins were mostly not present in the control mass spectrometry measurement where a cell line without tag was used in the immunoprecipitation using Flag or Myc antibodies. Therefore, these proteins do not seem to be background noise or contamination in my

samples.

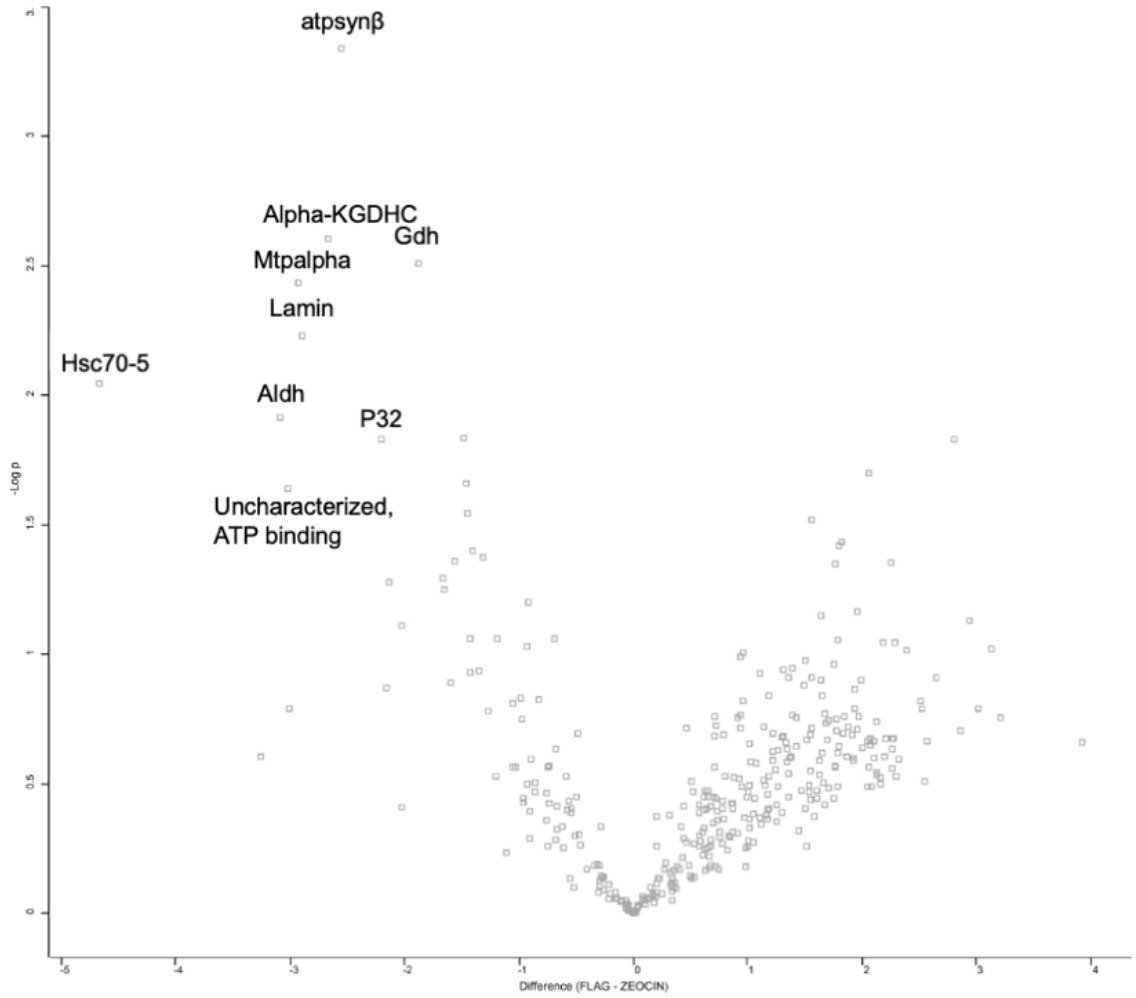


Figure 3.11.4. Mass spectrometry analysis of Snf-Flag cell line. Volcano plot based on Welch's T-test showing log2- difference in LFQ intensities between the Zeocin treated and untreated sample, revealing mitochondrial and other metabolic proteins as interactors of Snf.

However, if confirmed with downstream experiments, this would be to my knowledge one of the first evidence of the spliceosome relationship with mitochondria outside of the splicing canonical function [180].

I have performed a quantitative PCR with the primers for the beta subunit of the ATPsynthase (one of the hits in my proteomic measurement) to see if the expression changes upon Zeocin treatment. Results show no change in the mRNA expression.

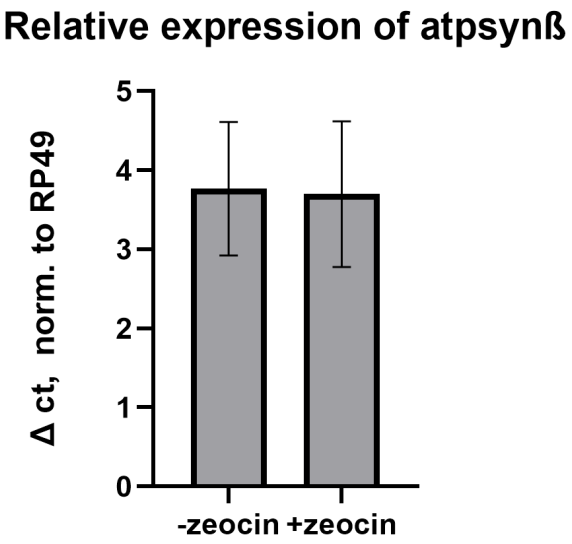


Figure 3. 11. 5. Quantitative PCR (qPCR) showing relative atpsynβ expression normalized to RP49 with and without 3.25 mg/ml Zeocin treatment for four hours.

3.12. Cloning, tagging, and testing of various biotin ligases to test for proximity labeling in *Drosophila* S2 cells

Although I have been successful in recovering the spliceosome complexes using my immunoprecipitation approach, the number of identified factors especially in Zeocin-treated samples was small. To find an alternative for stabilizing transient protein-protein interactions, while not damaging DNA as with formaldehyde crosslinking, the proximity labeling using biotin ligases seemed like a good approach. The idea was to use CRISPR/Cas9 to introduce the sequences of various biotin ligases at the sequences coding for proteins of interest. I took advantage of an already PCR-based

CRISPR/Cas9 system developed in the laboratory, where plasmids together with gene-specific primers are used to create homology donors. I have performed cloning of our standard plasmids to include the sequences of BASU, MiniTurbo, and Turbo biotin ligases sequences together with an HA tag I could use to check for successful tagging.

I first used CRISPR/Cas9 to edit the genomic sequence for Act5C in *Drosophila* S2 cells to code for the BASU ligase at the C terminus and confirmed the tag with the western blot (Figure 3.12.1., A). BASU ligase is a modification of BirA* biotin ligase from *Bacillus subtilis* with a faster biotinylation kinetics [181].

Next, I performed a streptavidin binding assay by detecting the biotinylated proteins with Streptavidin coupled to the HRP enzyme, which made them visible on the membrane. It was noticed some signal was detected in the sample where there was no biotin added (as it is present both endogenously and in the cell culture medium), but the signal increased with externally added biotin, showing the pattern of biotinylation (Figure 3.12.1, B).

I then cloned the plasmids for our CRISPR/Cas9 system where the homology donors carry the improved and more successful biotin ligases MiniTurbo and TurboID. These plasmids I have then used to tag Snf and Tango4, where only the Snf tagging was successful (Figure 3.12.2, C). I then performed the same streptavidin assay as before where I was able to see that compared to the BASU ligase, the TurboID ligase achieves stronger biotinylation activity (Figure 3.12.2, D) and at the same biotin concentration biotinylates more proteins. This would be a favorable approach for making the samples for mass spectrometry.

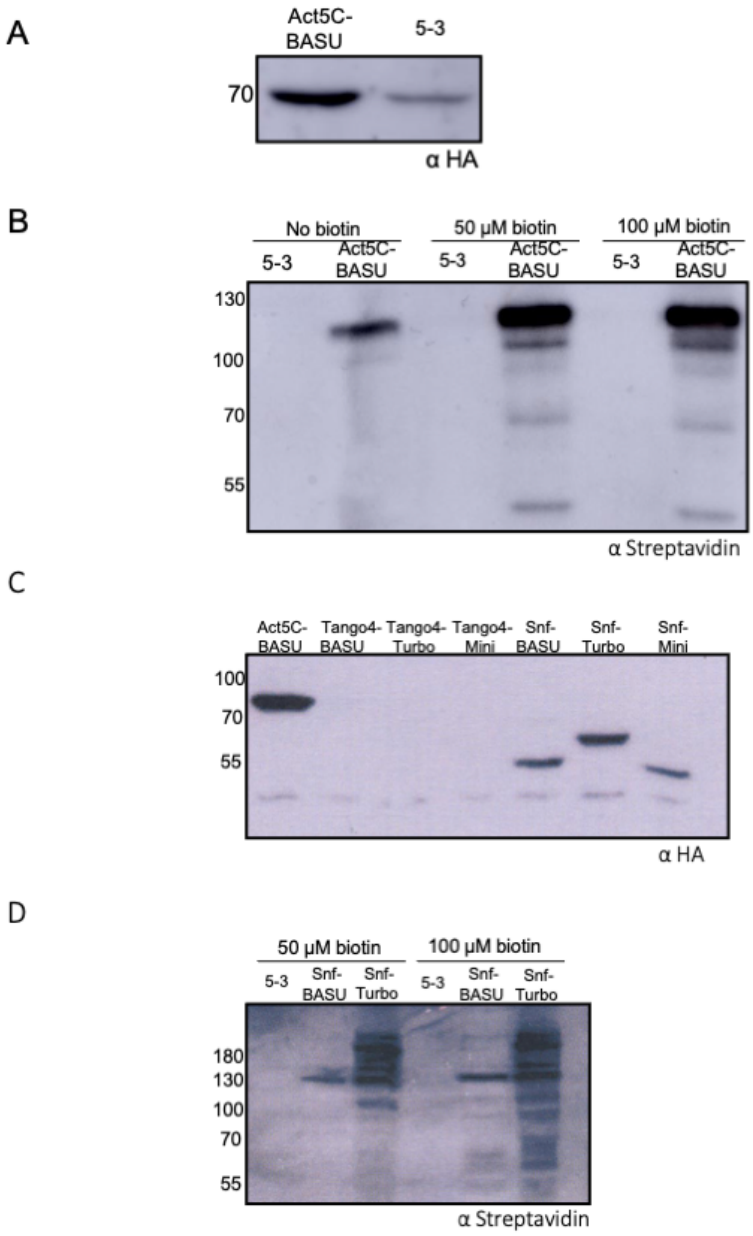


Figure 3. 12. 1. Proximity-labelling using biotin-ligases. In A, the Act5C was tagged with a BASU ligase carrying the HA tag. In B, a streptavidin assay was done where Streptavidin-HRP was used to detect biotinylated proteins on the membrane. In C, confirmation of successful tagging of Snf with various biotin-ligases. In D, the same streptavidin assay as in B was made where Streptavidin-HRP was used to detect the biotinylated proteins.

3.13. Luciferase-based reporter to assess the influence of spliceosome-inhibitors on DNA break-derived small RNA biogenesis

Using western blot as a technique, I could not confirm the production of the functional Sf3b1 protein with a Flag tag after I used CRISPR/Cas9 to edit its genomic sequence. However, I performed a PCR check where one primer was binding in the Flag tag, and the second was binding inside Sf3b1. The results showed that the genomic sequence of Sf3b1 was edited (3.13.1., A).

To approach a question from a different angle, I have used Sf3b1 chemical inhibitors. In this experiment, I have performed a 2-hour drug treatment with the concentration of two splicing drugs Pladienolide B and Herboxidiene, and used Mirin as a control. The concentrations used were based on IC50 kill curves made by Romy Böttcher. The killing curves were used to understand the lethality of the drugs on our S2 cells, as they affect an essential cellular function such as splicing. I have also used Mirin as a drug affecting the homology repair as a positive control [182]. In summary, I could not detect any effect of Herboxidiene or Pladienolide B on the generation of siRNAs at a DNA double-strand break. However, I have noticed a clear effect of Mirin which affected the silencing potential (3.13.1., B), which was later published as part of the study of the polymerase of choice for the expression of small RNAs at the DNA break [116].

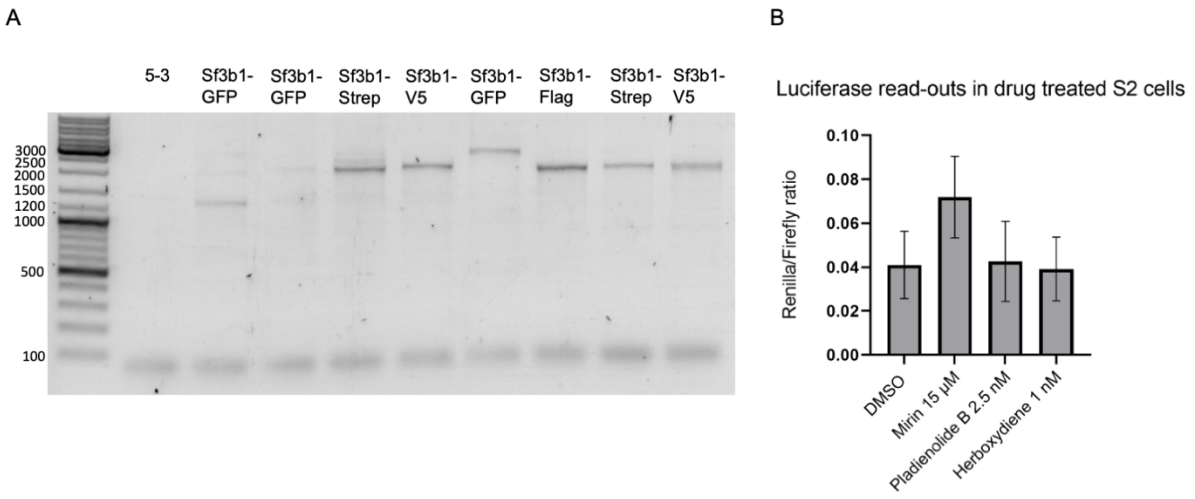


Figure 3. 13. 1. In A, PCR confirmations of tagged Sf3b1 with sequences of various epitope tags using CRISPR/Cas9 which did not yield functional proteins on western blot (Figure 3.1.1., B). Expected sizes of PCR products: GPF=2800 nt, Flag=2200 nt, Strep=2200 nt, and V5=2200 nt. In B, Luciferase-based reporter system for evaluating the functionality of the silencing machinery after chemical inhibition of Sf3b1. The S2 cells were treated for 2 hours with respective drug and then transfected with luciferase plasmids for 72 hours. The assay was done with three biological replicates. Mirin served as a positive control.

4. Discussion

4.1. Addressing transient spliceosome interactions: optimization of the methodological framework for proteomic analysis

This work aimed to recapitulate the *in vivo* interactions of the spliceosome in the context of DNA damage in *Drosophila* S2 cells, with a particular focus on proteins that participate in the efficient generation of small RNAs resulting at the DNA break (diRNAs). However, the rare events of spliceosomes engaging at or as a result of a DNA break compared to their canonical role in splicing are challenging to display with proteomic methods given the vastly more abundant usual splicing reactions that take place in the nucleus at any given moment.

Our strategy of inducing massive, genome-wide DNA damage using Zeocin, a phleomycin D1 formulation based on copper, was aimed to increase the proportion of DNA damage-associated events. I have tested various concentrations of Zeocin during a 4-hour incubation to identify a compromise between giving enough time for the cell to activate its DNA damage response, while still not expanding time over the frame where the transient interactions involving spliceosome would be finished. Literature reported DNA damage induction with Zeocin in yeast already with a concentration of 0.2 mg/ml [183]. It was also reported in the literature that Zeocin inhibited cell growth in *Drosophila* Kc1 and SL2 cell lines at 0.05 mg/ml and 0.075 mg/ml, respectively [184].

To evaluate the sensitivity of our cultured S2 cells to Zeocin, I have chosen two approaches: one was to perform fluoresce-activated cell sorting (FACS) to evaluate overall cell viability and the other was to perform a Neutral Comet Assay, single-cell electrophoresis to evaluate DNA double-strand break on a single cell level [185]. The initial fragmentation in the comet assay can already be seen with Zeocin concentrations starting at 0.35 mg/ml. It does not increase visibly at a dose of 0.75

mg/ml, but produces comets with very long tails at the concentration of 3.25 mg/ml (Figure 3.4.2.). The dose was higher than expected, when compared to the dosage reported in the literature, even though the manufacturer also recommends higher doses for some cell lines, even up to 1 mg/ml. However, each cell line could have a different tolerance to the drug, based on different intrinsic factors (i.e., genetic traits) and external factors (i.e., culturing conditions).

Optimization of the immunoprecipitation (IP) went in the direction of assuring that the chromatin-bound protein fraction was successfully extracted and that the transient spliceosome interactions were preserved. The first was addressed with a 2' incubation of the sample with the nuclease Benzonase, which came after a round of switching between the lysis buffers with and without urea, to finally extract the chromatin-bound fraction. Benzonase led to an increased efficiency in the final lysis fraction, shown as input in Figure 3.2.1.

To preserve transient spliceosome interactions, I have tried crosslinking the samples with either formaldehyde (FA) or disuccinimidyl suberate (DSS). FA is a potent crosslinker with a short spacer length and a known DNA-damaging agent [186] [187] [188], while DSS possesses a mid-size spacer length and is a less challenging agent for the cellular physiology [189]. FA crosslinking impaired the overall protein extraction measured by a Bradford Assay, more than two times in both 0.37% and 1% 10' incubation (Figure 3.3.1.). It is important to mention that the protein concentration was measured only in the final extraction fraction, which opens the possibility of extraction after FA treatment having different "dynamics" and proteins being extracted more and thus lost in the first extraction fractions. That seems plausible given the cell and membrane damaging effect that FA has [190]. Formaldehyde-treated samples were further tested on mass spectrometry and non-crosslinked samples served as a control, pulled down via the Flag-Tag in the Tango4-Flag cell line. Only half of the proteins identified were the same between the crosslinked and non-crosslinked samples, implying potentially big interactome changes upon crosslinking (data not shown). As it was affecting the extraction, further sample processing with FA before mass spectrometry was dismissed.

Figure 4.1.1 Graphic representation of the project workflow: inducing massive DNA damage after which the spliceosome proteins of interest will be pulled down via immunoprecipitation and later subjected to mass spectrometry measurement

4.2. Mass spectrometry confirms the specificity of the pull-down via Tango4-Flag

To qualitatively asses the mass spectrometry data, a principal component analysis (PCA) was done in Perseus [191]. The PCA takes into account the protein abundance levels and forms groups accordingly that are plotted in the two-dimensional space, revealing groups of samples that are more or less related to each other [192] [193]. This form of explorative data analysis is useful for summarizing the major characteristics of the samples and grouping them. The PCA plot separates the protein samples according to the antibody used (Figure 3. 5.1., B). In the PCA plot showing individual proteins that are associated, the Prp19 complex, in particularly Tango4 and its close partner Prp19, separate together. Other Prp19-related factors such as Cdc5 or Bcas9, also form closer mutual associations than with the rest of the group (Figure 3.5.2., C).

Hierarchical clustering recapitulates the specific nature of the performed immunoprecipitation experiments, where the Prp19 complex forms a separate cluster at the top of the cluster groups, with a clear difference between the V5 and Flag samples as seen in the colors of the heat map (Figure 3.5.2., A). A STRING analysis [194] suggests what are both the predicted and confirmed interactors of Tango4, which matches to a good extent my proteomic results.

Finally, a volcano plot comparing V5 versus Flag samples in all replicates indicates the biggest change of label-free intensity for the members of the Prp19 complex, such as Prp19 itself, Tango4, Cdc5, and Bcas9 (Figure 3.5.3.). All of these are well-described members of the Prp19 complex [165]. Other major groups of proteins resulting in the volcano plot are the chromatin remodeling factors and ribosomal proteins. A large amount of ribosomal proteins can be detected in control mass spectrometry experiments, potentially implying they represent the background in the sample.

4.3 Proteomic results suggest Prp19-related complex interacts with a few zinc-finger transcription factors including Putzig

When comparing Zeocin treated versus untreated samples, one can notice the increased binding of three different zinc-finger transcription factors upon Zeocin treatment, with the most prominent one being the seven zinc-finger transcription factors Putzig/Z4 (Figure 3.5.4.). Putzig has already been validated as a factor required for the efficient generation of small RNAs at a DNA break in *Drosophila* in the genome-wide RNAi screen [114]. Its association with the spliceosome has not been shown before. However, the presence of Putzig in a few nuclear complexes has been described, revealing its versatile nature in regulating genome stability.

Its role in telomere maintenance was described in *Drosophila*, whose mechanism of telomere extension differs from the other eukaryotes and includes three highly conserved non-LTR transposons HET-A, TAHRE and TART [195] [196]. The retrotransposons in the telomere regions are being duplicated and inserted, thus at the same time assuring their survival and increasing the length of the telomeres. Belonging to the class I of transposable elements, their duplication involves an RNA intermediate. Putzig regulates together with its partners Jil-1 and HP1a the chromatin state in the telomere region, thus actively influencing the transcription of the transposable elements. Putzig both assists Jil-1, which positively influences telomere extension through supporting euchromatin states and thus transcription, and HP1a which influences the heterochromatin spreading and thus acts negatively on telomere extension through transcription blockade. In Jil-1 mutants, heterochromatin marks are spreading, while in the Putzig mutants, the boundary is disturbed, and it leads to the expansion of heterochromatin markers. Its subtle interplay is important to assure that telomere extension through transposition stays inside the needed levels, as overexpression of transposable elements would eventually lead to genomic instability. Furthermore, another complex involving Putzig has a role in the expression of the transposons in the telomere regions in *Drosophila*. The DREF, TRF2, and KEN transcriptional complex has a direct role in the expression of the HeT-A and TAHRE Fields[176]. Together with them, Putzig is also involved in the regulation of gene expression and JAK/STAT pathway.

Putzig associates with the Jil-1 kinase in another complex as well, which regulates the genome architecture. Eukaryotic genomes are organized in topologically associated domains (TADs) [197], which was shown as a result of series of fluorescent in-situ hybridization (FISH) experiments and later the advancement of chromosome conformation capture methods that undisputedly showed that the eukaryotic chromatin divided in eu- and heterochromatin does not randomly take the position in the nucleus, but instead stays in a pre-defined niche, helped by multiple genome architecture proteins [198] [199] and regulatory small RNAs as well [200]. This complex includes Chromator and BEAF-32 and binds at the TAD borders.

However, *Drosophila* has a more complex chromatin organization in form of polytene chromosomes in certain cell types [201], where numerous copies of DNA are present forming a higher-form structure with band and interband regions, these representing the functional borders for regulating gene expression, similar to TADs [202]. Jil-1 regulates the phosphorylation of the histone H3S10 at the polytene chromosome interbands, in its absence the alignment of the fibrils is highly compromised. The proteins that co-localize with it, Chromator and Putzig [203], are responsible for the mitotic spindle formation and chromosome segregation [204], as well as the maintaining the band-interband structure, respectively [205].

Putzig controls the Notch pathway during the development via its association with the NURF complex. At first, it was shown that it binds the DNA at the Notch target genes vg and E(spl)m8 independently of its other close interactors with which it aids transcription and replication such as TRF2 and DREF [206]. Later it was shown that the binding of Putzig to the Notch target genes heavily depends on NURF complex [207] [208], and all subunits of this co-purified with Putzig. This implies another important role for Putzig in *Drosophila* development, whose loss leads to increased apoptosis [209]. This is an interesting insight into the potential mechanism of Putzig action: it does not bind the target genes directly, but rather via its chromatin remodelling partners, such as Chromator, HP1a, or HP1c.

The Putzig interaction with the spliceosome would be new in the literature if it can be further confirmed. My proteomics data suggest that this interaction exists but confirming it with immunoprecipitation experiments did not yield a clear answer after being tested in both orientations either via Tango4 or via Putzig. There could be a few explanations for that. First, the mass spectrometry measurement done on the Putzig-Flag cell line did not show statistically significant interaction with Tango4, although the protein was detected in many different fragments (peptides which by using BLAST uniquely mapped to Tango4, Figure 3.7.3.). It is imaginable that the complex stability is affected by the choice of the bait and that the interaction from the Putzig pull-down is lost in the mass spectrometry processing steps. Another possibility is the indirect binding of these two proteins. When comparing the control IP experiment (V5) with an experiment in which the specific antibody was used (Flag), Prp19 was binding more

Flag antibody than the bait Tango4 itself which carried the Flag tag (Table 5.1, Prp19=8.26, Tango4= 6.86 difference), so one could argue to which component exactly is the Putzig binding in the Prp19 complex. It would be useful in the future to test this possibility via an epitope tag on Prp19. Unfortunately, the previous efforts in our laboratory did not succeed in tagging the Prp19 in *Drosophila* S2 cells.

The mechanistic role of Putzig in the generation of the break-derived siRNAs remains an open question. Its zinc-finger nature speaks in favor of having the possibility to bind nucleic acids directly and/or recruit the transcription machinery, but in that case, the most fitting known partners would be the TRF2/DREF/KEN, which is the part of the core transcription initiation complex. However, I was not able to detect them in my proteomics data. The Putzig proteomic results that suggested its relationship to HP1c complex and in particular Row seemed like a plausible explanation of how Putzig can aid the transcription by engaging with an HP1 complex that has a positive rather than a negative effect on transcription, and a complex that aids the transcription initiation for genes lacking a canonical promoter. This is in the case of DNA damage inside a transcription unit something that is a potential problem if a new RNA polymerase is recruited as there is no available promoter to serve for the assembly of the initiation complex. If the polymerase transcribing the sense strand makes a “U-turn”, it is possible that the promoter is not the limiting factor, as the polymerase is already recruited. That could also explain the absence of TRF2/DREF in the measured interactome.

However, HP1c did not prove to be important for the process in the non-proteomic assays, leaving the open question of how Putzig aids antisense transcription at the DNA break.

Putzig could also be aiding the transcription indirectly through chromatin remodeling at the place of the DNA break. Its role in positively and negatively regulating the transcription has already been shown, and it might involve novel partners who reorganize the position where the RNA Polymerase stalled and make the transcription possible.

4.3. Non-proteomic assays support Putzig's role in the expression of small RNAs at the DNA break

Both the luciferase reporter and deep sequencing of small RNAs confirm Putzig's role in the expression of the small RNAs at the DNA break. However, Putzig's requirement in the process in the chromosomal setting has been shown for the first time in this work. Deep sequencing showed no meaningful effect of the Hp1c complex in the process (3.10.3.), which goes in line with luciferase experiments, where only Putzig affected the silencing machinery (3.9.3).

4.4. Proteomic data of Snf after Zeocin treatment shows interaction with various mitochondrial factors

There is an evolutionary connection between the spliceosome and the mitochondria, where introns of mitochondrial origin are found in the nuclear DNA and have a role in both constitutive and alternative splicing regulation [210].

Splicing is also a process dependent on ATP produced by the mitochondria in various parts of the splicing cycle, however the association of Snf with a subset of mitochondrial proteins remains mysterious. The hits include the β subunit of the ATPsynthase, a mitochondrial trifunctional protein with a role in β-oxidation of long-chain fatty acids that contributes to a few mitochondrial pathologies [211], as well as alpha-ketoglutarate-dehydrogenase, an enzyme mediating oxidizing stress, which would go in line with the very heavy Zeocin treatment [212]. It is important to note that these proteins have not been identified in other proteomic measurements with Zeocin (data not shown).

4.5. Introducing new biotin-ligases to our CRISPR/Cas9 system to aid the specificity of the mass spectrometry measurements in the future

The protein-protein interactions are cardinal for understanding cellular physiology and many advances were made in the direction of purification, enrichment, and proteomic

analysis of protein complexes in different physiological and pathological conditions. The immunoprecipitation followed by mass spectrometry approach I have taken in my work is efficient for identifying strong and relatively stable protein interactors, while many transient interactions could get lost in the process of purification or sample processing for the mass spectrometry. Because various detergents interfere with the mass spectrometry measurement, several additional washes need to be performed apart from the standard washes that are part of the immunoprecipitation experiments, thus increasing the possibility of losing transient or weak interactions.

The recently emerging proximity labeling techniques could overcome this by alternatively capturing the transient and weak protein-protein interactions when paired with mass spectrometry [213]. The concept behind this involves the process of the biotinylation of the neighboring proteins by a protein of interest fused with one of the biotin ligases. They convert the supplemented substrate into a reactive biotinylated intermediate that would then transfer the biotin to side-chain amino acids present in the close vicinity [213].

Proximity labeling labeling is being done with two groups of enzymes: biotin ligases and peroxidases. The first group uses biotin that is being transferred to lysine, while the second uses biotin-phenol together with hydrogen peroxide (H_2O_2) and has tyrosine as the target amino acid [214]. The first proximity labeling approach was named BioID after it was made based on the concept behind the DamID method and it involved a BirA, *E. coli* biotin ligase [215]. Later, a smaller in size biotin-ligase was developed in the BioID2 attempt to improve the proximal labeling using biotin [216]. This same biotin ligase was also produced using the *Bacillus subtilis*, with mutations that would increase the reaction kinetics and improve the signal-to-noise ratio, named BASU [217]. Overall, BioID, BioID2, and BASU are considered to show similar performance. Finally, TurboID and mini turbo variants of the biotin ligases were developed in 2012 [218], with higher activity compared to BioID and other previously developed biotin ligases [219].

APEX represents another promising way to perform proximity labeling by oxidising the biotin phenol in the presence of H_2O_2 into a biotin-phenoxyl radical that then reacts with tyrosines in the neighboring proteins [220]. Two challenges arise from this approach: low sensitivity and the necessity for H_2O_2, which causes DNA damage, and other physiological challenges for the cell [221]. On the other hand, since no biotinylation occurs before the reagents are added, the method allows for excellent temporal resolution.

I first used BASU to tag *in vivo* the *Drosophila* Act5C (Figure 3.12.1., A) and performed a streptavidin assay to detect the biotinylated proteins directly on the membrane (Figure 3.12.1, B). The concentration of extra biotin suggested in the literature of 50 µM was enough to visibly intensify the biotinylation compared to the control. Some biotinylation can be noticed in the control as well, probably due to intracellular biotin and biotin inside the cell culture medium that was used as a substrate by the biotin ligase. The assay could be a useful quality control before proceeding to mass spectrometry, where one could then even test if the biotinylated proteins detected match the protein sizes of proteins detected on the membrane.

In this work, plasmids for introducing the TurboID, Miniturbo, and Apex via CRISPR/Cas9 into the genome have been made. Snf was tagged (Figure 3.12.1, C) and the same streptavidin assay was performed, where it was shown that the biotinylation is stronger with TurboID and miniturbo than with BASU, consistent with the literature results (Figure 3.12.1, D). This could be a future way to address better the weak and transient spliceosome interactions.

4.6. Splicing drugs targeting Sf3b1 do not have an effect on the expression of small RNAs at the DNA break

Splicing is an essential process in gene expression and therefore stands very centrally in many pathological conditions [222] [223] [224]. One of the proteins involved is Sf3b1, which has also been one of the candidates hits in the genome-wide RNAi

screen that identified it as a factor important for the efficient generation of the small RNAs at the DNA break [114], but tagging it with CRISPR/Cas9 did not yield a functional protein although it has been tagged (Figure 3.13.1., A). Big drug discovery screens have revealed natural compounds that have an effect on splicing and are used to correct pathological splicing [225] [226] [227]. Two of the agents acting on splicing isolated from Streptomyces, Herboxidiene, and Pladienolide B, show anticancer activity [228] [229] [230] [231]. Both of them have a similar mechanism of action based on Sf3b1, a splicing protein that is part of the U2 snRNP [232] [233]. I incorporated a drug treatment by Pladienolide B and Herboxidiene before the luciferase reporter assay and thus checked whether DNA break-derived siRNA biogenesis is perturbed. The killing curves were made for both drugs (performed by Romy Böttcher) and a range of concentrations was chosen for the assay. The results did not show any effect on the silencing activity in *Drosophila*, in contrast to Mirin (Figure 3.12.1, B), a drug affecting the homologous recombination [234]. The effect of Mirin was published in our study related to the polymerase of choice at the DNA break that is responsible for the antisense transcription [116].

5. Appendix

5.1. List of proteins identified in proteomic analysis of Tango-Flag cell line

Protein identified	**Difference V5-Flag**
Prp19	-8.265488942
Tango4	-6.848833084
Cdc5	-6.652212143
RpL7A	-6.317071279
RpL7	-6.227045695
RpL27	-6.073028564
RpL3-RA	-5.931020101
RpL6	-5.865787506
RpLP1	-5.786828359
RpL15	-5.624596914
RpS4	-5.593062719
CG31664	-5.486794154
Dmel\CG31974	-5.431954066
RpL5	-5.428117752
BCAS2	-5.353844325
RpL18	-5.317047755
RpL23A	-5.157856623
RpLP0	-5.152421316
RpL12	-4.867942174
RpL10	-4.86382548
RpL13	-4.857906977
RpS13	-4.703018824

RpL30	-4.693223317
His1:CG33807	-4.64773496
RpL17	-4.641236623
RpS26	-4.499703089
Hsc70-4	-4.49489212
RpL27A	-4.430548986
RpL24	-4.403343201
RpLP2-RB	-4.400370916
RpL34b	-4.325454076
RpL10Ab	-4.281312943
RpL17A	-4.280557632
RpL37A	-4.272760391
RpS16	-4.058572769
B52	-4.042723338
RpL9	-4.004770915
Ssrp-RA	-3.855110168
RpS15Aa	-3.769371033
RpS8	-3.726329168
RpS3	-3.661627452
RpS23	-3.646792094
RpS24	-3.475048701
RpS7	-3.446961721
CG4038	-2.627561569
Hsc70-3	-2.268355687
RpL13A	-2.264189402
RpL36A	-2.216751734
RPS11	-1.96699206
SsRbeta	-1.745146434
Dmel\CG15784	-1.423137665

Protein identified (excluding the bait)	**Difference Flag-Flag Zeocin**
RpS28b	-2,275469462
pzg	-2,136934916
RpS9	-1,815758387
Dmel\CG31974	-1,6468455
RpL26	-1,537110647
RpL11	-1,532800039
Dmel\CG8478	-1,45778211
His2Av	-1,382656733
CG10373-RA	-1,085863749
Dmel\CG4360	-1,017359416
Ssrp-RA	-1,002853394
His2A:CG33859	-0,965216955
RpS27	-0,879981995
aru	-0,829570134
RpS16	-0,748252869
CG7993-RA	-0,723069509
RpS19a	-0,637430827
alt	-0,602233887
RpS2	-0,591943105
bel-RB	-0,575125376
Nop60B	-0,571428299
Ca-P60A-RC	-0,415452321
CG2064	-0,407372157
CG4038	-0,393218994
RpS3	-0,360421499
sta-RD	-0,317536036
Hsc70-3	-0,30372111
Hsc70-4	-0,298401515

ERp60	-0,294817607
sqd	-0,274960836
Tapdelta	-0,230542501
RpS26	-0,202377319
RpS29	-0,190823237
His4r	-0,170898438
RpL5	-0,165278753
RpS23	-0,160033544
RpS24	-0,146696091
Fmr1	-0,108023961
RpS18-RB	-0,095826467
RpL21	-0,06551679
RpL27	-0,02879715

5.2. List of proteins identified in proteomic analysis of Pzg-Flag cell line

Protein (excluding the bait)	**Difference Flag - Zeocin Flag**
His4r	-1.495682398
row	-1.291735967
Chro	-0.38111496
Hex-A	-0.320750554
Dmel\CG31974	-0.190646489
RpS14b	-0.138096491

5.3. List of proteins identified in proteomic analysis of Snf-Flag cell line

Protein	**Difference Flag - Zeocin Flag**
snRNP-U1-70K	-9.031178474
Dmel\CG31974	-8.396011353
snf	-8.149991989
Sf3b1	-7.942116261
B52	-7.270421982
Sf3b3	-7.124716282
Prp8	-7.101250648
Can	-6.794440746
Dmel\CG4849	-6.633405685
Prp19	-6.607271194
Sf3a1	-6.317574024
Sf3b2	-6.226937294
SF2	-6.064528465
Hfp	-6.009218693
Srp54	-5.831031799
CG14729	-5.628904343
RpS9	-5.600846767
RpL7	-5.541612625
RpL7A	-5.528006554
Cdc5	-5.219066143
RpL10Ab	-5.204394817

RpL5	-5.192184448
RpL3-RA	-4.999228954
Srrm234	-4.950532913
RnpS1	-4.905161858
Ciz1	-4.88724041
RpL6	-4.823321819
Doa	-4.817500114
Sf3a2	-4.805406094
RpL18	-4.619645119
CG31664	-4.596074104
Pnn	-4.584047318
Alsin2	-4.549799919
RpL36	-4.533802032
CDC40	-4.519691944
CkIIalpha	-4.500092506
Dmel\CG15784	-4.477025986
RpL30	-4.465763092
Bx42	-4.454398632
RpL11	-4.451289654
Nop60B	-4.44086647
Dmel\CG3436	-4.420285702
RpL9	-4.371193886
RpS24	-4.339687347
RpL15	-4.33466053
RpS13	-4.312465668
Tango4	-4.287638187
RpL13	-4.26299572
Caper	-4.25799942
RpL24	-4.245203495
Dhx15	-4.225458622

RpL27	-4.217125416
Pitslre	-4.114990711
Bin1	-4.080038071
Fand	-4.048753262

Protein (excluding the bait)	**Difference Flag - Zeocin Flag**
Hsc70-5	-4,668994904
Map205	-3,261417866
Aldh	-3,090513706
Mtor	-3,011713505
Mtpalpha	-2,932994366
Lam	-2,895811081
alpha-KGDHC	-2,668129444
ATPsynbeta	-2,55401659
P32	-2,208996773
Gp93	-2,157588482
Hsp83	-2,141209602
anon-WO0118547.80	-2,028298378
RAF2	-2,018818378
Gdh	-1,87869072
FK506-bp2-RA	-1,667090893
Akap200	-1,648581982
Tctp	-1,562399387
Sar1	-1,464134216
Mtpbeta	-1,442668915

Dmel\CG8507	-1,429447651
TER94-RC	-1,424752235
betaTub56D	-1,401036739
COX6B	-1,351112843
Nlp	-1,307950497
ERp60	-1,271244526
eEF1alpha2	-1,194045067
Hsc70-3	-1,187648773
Eps-15	-1,102864742
Pdi	-1,052904129
Arf79F	-1,052127361
PMCA	-1,032699585

6. References

1. Hager, G.L., J.G. McNally, and T. Misteli, *Transcription dynamics.* Mol Cell, 2009. **35**(6): p. 741-53.
2. Li, L. and Y. Liu, *Diverse small non-coding RNAs in RNA interference pathways.* Methods Mol Biol, 2011. **764**: p. 169-82.
3. Mattick, J.S. and I.V. Makunin, *Small regulatory RNAs in mammals.* Human Molecular Genetics, 2005. **14**(suppl_1): p. R121-R132.
4. Eddy, S.R., *Non-coding RNA genes and the modern RNA world.* Nat Rev Genet, 2001. **2**(12): p. 919-29.
5. Roux, P.P. and I. Topisirovic, *Regulation of mRNA translation by signaling pathways.* Cold Spring Harb Perspect Biol, 2012. **4**(11).
6. Cramer, P., *Organization and regulation of gene transcription.* Nature, 2019. **573**(7772): p. 45-54.
7. Plaschka, C., et al., *Transcription initiation complex structures elucidate DNA opening.* Nature, 2016. **533**(7603): p. 353-358.
8. Butler, J.E. and J.T. Kadonaga, *The RNA polymerase II core promoter: a key component in the regulation of gene expression.* Genes Dev, 2002. **16**(20): p. 2583-92.
9. Lee, D.H., et al., *Functional characterization of core promoter elements: the downstream core element is recognized by TAF1.* Mol Cell Biol, 2005. **25**(21): p. 9674-86.
10. Pelechano, V. and L.M. Steinmetz, *Gene regulation by antisense transcription.* Nat Rev Genet, 2013. **14**(12): p. 880-93.
11. Shearwin, K.E., B.P. Callen, and J.B. Egan, *Transcriptional interference--a crash course.* Trends Genet, 2005. **21**(6): p. 339-45.
12. Dupont, S. and S.A. Wickstrom, *Mechanical regulation of chromatin and transcription.* Nat Rev Genet, 2022.
13. Tufarelli, C., et al., *Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease.* Nat Genet, 2003. **34**(2): p. 157-65.
14. Guil, S. and M. Esteller, *Cis-acting noncoding RNAs: friends and foes.* Nat Struct Mol Biol, 2012. **19**(11): p. 1068-75.
15. Shimoni, Y., et al., *Regulation of gene expression by small non-coding RNAs: a quantitative view.* Mol Syst Biol, 2007. **3**: p. 138.
16. He, L. and G.J. Hannon, *MicroRNAs: small RNAs with a big role in gene regulation.* Nature Reviews Genetics, 2004. **5**(7): p. 522-531.
17. Ledda, B., et al., *Small RNAs in eucaryotes: new clues for amplifying microRNA benefits.* Cell & Bioscience, 2020. **10**(1): p. 1.
18. Ambros, V. and X. Chen, *The regulation of genes and genomes by small RNAs.* Development, 2007. **134**(9): p. 1635-1641.
19. Brown, J.B., et al., *Diversity and dynamics of the Drosophila transcriptome.* Nature, 2014. **512**(7515): p. 393-9.
20. Pérez-Lluch, S., et al., *Natural antisense transcription in Drosophila melanogaster.* 2020, PLOS Genetics.
21. Adams, M.D., et al., *The Genome Sequence of Drosophila melanogaster.* Science, 2000. **287**(5461): p. 2185-2195.
22. Lin, M.F., et al., *Revisiting the protein-coding gene catalog of Drosophila melanogaster using 12 fly genomes.* Genome Res, 2007. **17**(12): p. 1823-36.

23. Yandell, M., et al., *A computational and experimental approach to validating annotations and gene predictions in the Drosophila melanogaster genome.* Proceedings of the National Academy of Sciences, 2005. **102**(5): p. 1566-1571.
24. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.
25. Salzberg, S.L., *Open questions: How many genes do we have?* BMC Biology, 2018. **16**(1): p. 94.
26. Telonis-Scott, M., et al., *Sex-specific splicing in Drosophila: widespread occurrence, tissue specificity and evolutionary conservation.* Genetics, 2009. **181**(2): p. 421-34.
27. Venables, J.P., J. Tazi, and F. Juge, *Regulated functional alternative splicing in Drosophila.* Nucleic Acids Res, 2012. **40**(1): p. 1-10.
28. Mohr, C. and B. Hartmann, *Alternative splicing in Drosophila neuronal development.* J Neurogenet, 2014. **28**(3-4): p. 199-215.
29. Lees, J.G., J.A. Ranea, and C.A. Orengo, *Identifying and characterising key alternative splicing events in Drosophila development.* BMC Genomics, 2015. **16**(1): p. 608.
30. Srivastava, D., et al., *Modulation of Yorkie activity by alternative splicing is required for developmental stability.* Embo j, 2021. **40**(3): p. e104895.
31. Geyer, A., et al., *Impact of Ultrabithorax alternative splicing on Drosophila embryonic nervous system development.* Mech Dev, 2015. **138 Pt 2**: p. 177-189.
32. Ruan, K., et al., *Alternative splicing of Drosophila Nmnat functions as a switch to enhance neuroprotection under stress.* Nat Commun, 2015. **6**: p. 10057.
33. Martin Anduaga, A., et al., *Thermosensitive alternative splicing senses and mediates temperature adaptation in Drosophila.* Elife, 2019. **8**.
34. Kim, H.Y. and V.N. Gladyshev, *Alternative first exon splicing regulates subcellular distribution of methionine sulfoxide reductases.* BMC Mol Biol, 2006. **7**: p. 11.
35. Katzenberger, R.J., M.S. Marengo, and D.A. Wassarman, *ATM and ATR pathways signal alternative splicing of Drosophila TAF1 pre-mRNA in response to DNA damage.* Mol Cell Biol, 2006. **26**(24): p. 9256-67.
36. McCullers, T.J. and M. Steiniger, *Transposable elements in Drosophila.* Mob Genet Elements, 2017. **7**(3): p. 1-18.
37. Barrón, M.G., et al., *Population Genomics of Transposable Elements in Drosophila.* Annual Review of Genetics, 2014. **48**(1): p. 561-581.
38. Wells, J.N. and C. Feschotte, *A Field Guide to Eukaryotic Transposable Elements.* Annual Review of Genetics, 2020. **54**(1): p. 539-561.
39. Muñoz-López, M. and J.L. García-Pérez, *DNA transposons: nature and applications in genomics.* Curr Genomics, 2010. **11**(2): p. 115-28.
40. Goodier, J.L. and H.H. Kazazian, Jr., *Retrotransposons Revisited: The Restraint and Rehabilitation of Parasites.* Cell, 2008. **135**(1): p. 23-35.
41. Sheen, F.M. and R.W. Levis, *Transposition of the LINE-like retrotransposon TART to Drosophila chromosome termini.* Proc Natl Acad Sci U S A, 1994. **91**(26): p. 12510-4.
42. Bowen, N.J. and J.F. McDonald, *Drosophila euchromatic LTR retrotransposons are much younger than the host species in which they reside.* Genome Res, 2001. **11**(9): p. 1527-40.
43. Petrov, D.A., et al., *Size Matters: Non-LTR Retrotransposable Elements and Ectopic Recombination in Drosophila.* Molecular Biology and Evolution, 2003. **20**(6): p. 880-892.
44. Bryant, L.A., et al., *The retrotransposon copia regulates Drosophila gene expression both positively and negatively.* Nucleic Acids Res, 1991. **19**(20): p. 5533-6.
45. Shaposhnikov, M.V., et al., *Molecular mechanisms of exceptional lifespan increase of Drosophila melanogaster with different genotypes after combinations of pro-longevity interventions.* Communications Biology, 2022. **5**(1): p. 566.
46. Touret, F., et al., *In between: gypsy in Drosophila melanogaster reveals new insights into endogenous retrovirus evolution.* Viruses, 2014. **6**(12): p. 4914-25.

47. Papaceit, M., et al., *The dynamics of the roo transposable element in mutation-accumulation lines and segregating populations of Drosophila melanogaster.* Genetics, 2007. **177**(1): p. 511-22.
48. Russo, J., A.W. Harrington, and M. Steiniger, *Antisense Transcription of Retrotransposons in Drosophila: An Origin of Endogenous Small Interfering RNA Precursors.* Genetics, 2016. **202**(1): p. 107-21.
49. Xu, J. and S. Cherry, *Viruses and antiviral immunity in Drosophila.* Dev Comp Immunol, 2014. **42**(1): p. 67-84.
50. Swevers, L., J. Liu, and G. Smagghe, *Defense Mechanisms against Viral Infection in Drosophila: RNAi and Non-RNAi.* Viruses, 2018. **10**(5).
51. Webster, C.L., et al., *The Discovery, Distribution, and Evolution of Viruses Associated with Drosophila melanogaster.* PLOS Biology, 2015. **13**(7): p. e1002210.
52. Liu, S., D. Vijayendran, and B.C. Bonning, *Next generation sequencing technologies for insect virus discovery.* Viruses, 2011. **3**(10): p. 1849-69.
53. Bellen, H.J., C. Tong, and H. Tsuda, *100 years of Drosophila research and its impact on vertebrate neuroscience: a history lesson for the future.* Nat Rev Neurosci, 2010. **11**(7): p. 514-22.
54. Merkling, S.H. and R.P. van Rij, *Beyond RNAi: antiviral defense strategies in Drosophila and mosquito.* J Insect Physiol, 2013. **59**(2): p. 159-70.
55. Medzhitov, R., P. Preston-Hurlburt, and C.A. Janeway, Jr., *A human homologue of the Drosophila Toll protein signals activation of adaptive immunity.* Nature, 1997. **388**(6640): p. 394-7.
56. Poltorak, A., et al., *Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene.* Science, 1998. **282**(5396): p. 2085-8.
57. Chotkowski, H.L., et al., *West Nile virus infection of Drosophila melanogaster induces a protective RNAi response.* Virology, 2008. **377**(1): p. 197-206.
58. Mukherjee, S. and K.A. Hanley, *RNA interference modulates replication of dengue virus in Drosophila melanogaster cells.* BMC Microbiol, 2010. **10**: p. 127.
59. Carthew, R.W., P. Agbu, and R. Giri, *MicroRNA function in Drosophila melanogaster.* Semin Cell Dev Biol, 2017. **65**: p. 29-37.
60. Ghildiyal, M., et al., *Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells.* Science, 2008. **320**(5879): p. 1077-81.
61. Zambon, R.A., V.N. Vakharia, and L.P. Wu, *RNAi is an antiviral immune response against a dsRNA virus in Drosophila melanogaster.* Cell Microbiol, 2006. **8**(5): p. 880-9.
62. Shabalina, S.A. and E.V. Koonin, *Origins and evolution of eukaryotic RNA interference.* Trends Ecol Evol, 2008. **23**(10): p. 578-87.
63. Jose, A.M. and C.P. Hunter, *Transport of sequence-specific RNA interference information between cells.* Annu Rev Genet, 2007. **41**: p. 305-30.
64. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.* Nature, 1998. **391**(6669): p. 806-11.
65. Raoul, C., et al., *Lentiviral-mediated silencing of SOD1 through RNA interference retards disease onset and progression in a mouse model of ALS.* Nat Med, 2005. **11**(4): p. 423-8.
66. Pereira, T.C., et al., *Schistosoma mansoni: evaluation of an RNAi-based treatment targeting HGPRTase gene.* Exp Parasitol, 2008. **118**(4): p. 619-23.
67. Xiong, A.-S., et al., *Different effects on ACC oxidase gene silencing triggered by RNA interference in transgenic tomato.* Plant Cell Reports, 2005. **23**(9): p. 639-646.
68. Shan, G., *RNA interference as a gene knockdown technique.* The International Journal of Biochemistry & Cell Biology, 2010. **42**(8): p. 1243-1251.
69. Neumeier, J. and G. Meister, *siRNA Specificity: RNAi Mechanisms and Strategies to Reduce Off-Target Effects.* Frontiers in Plant Science, 2021. **11**.
70. Höck, J. and G. Meister, *The Argonaute protein family.* Genome Biology, 2008. **9**(2): p. 210.

71. Soleimani, S., et al., *Small regulatory noncoding RNAs in Drosophila melanogaster: biogenesis and biological functions.* Brief Funct Genomics, 2020. **19**(4): p. 309-323.
72. Yu, J.Y., et al., *Dicer-1-dependent Dacapo suppression acts downstream of Insulin receptor in regulating cell division of Drosophila germline stem cells.* Development, 2009. **136**(9): p. 1497-507.
73. Brennecke, J., et al., *bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila.* Cell, 2003. **113**(1): p. 25-36.
74. Wu, Y.C., et al., *Let-7-complex microRNAs regulate the temporal identity of Drosophila mushroom body neurons via chinmo.* Dev Cell, 2012. **23**(1): p. 202-9.
75. Luo, W. and A. Sehgal, *Regulation of circadian behavioral output via a MicroRNA-JAK/STAT circuit.* Cell, 2012. **148**(4): p. 765-79.
76. Zhao, J., et al., *MicroRNA-7: expression and function in brain physiological and pathological processes.* Cell & Bioscience, 2020. **10**(1): p. 77.
77. Zipper, L., et al., *The MicroRNA miR-277 Controls Physiology and Pathology of the Adult Drosophila Midgut by Regulating the Expression of Fatty Acid β-Oxidation-Related Genes in Intestinal Stem Cells.* Metabolites, 2022. **12**(4).
78. Denli, A.M., et al., *Processing of primary microRNAs by the Microprocessor complex.* Nature, 2004. **432**(7014): p. 231-5.
79. Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing.* Nature, 2003. **425**(6956): p. 415-419.
80. Yi, R., et al., *Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs.* Genes Dev, 2003. **17**(24): p. 3011-6.
81. Yi, R., et al., *Overexpression of exportin 5 enhances RNA interference mediated by short hairpin RNAs and microRNAs.* Rna, 2005. **11**(2): p. 220-6.
82. Hutvágner, G., et al., *A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA.* Science, 2001. **293**(5531): p. 834-8.
83. Forstemann, K., et al., *Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein.* PLoS Biol, 2005. **3**(7): p. e236.
84. Lim, Mandy Yu T., et al., *The Drosophila Dicer-1 Partner Loquacious Enhances miRNA Processing from Hairpins with Unstable Structures at the Dicing Site.* Cell Reports, 2016. **15**(8): p. 1795-1808.
85. Medley, J.C., G. Panzade, and A.Y. Zinovyeva, *microRNA strand selection: Unwinding the rules.* Wiley Interdiscip Rev RNA, 2021. **12**(3): p. e1627.
86. Miyoshi, K., et al., *Slicer function of Drosophila Argonautes and its involvement in RISC formation.* Genes Dev, 2005. **19**(23): p. 2837-48.
87. Tomari, Y., T. Du, and P.D. Zamore, *Sorting of Drosophila Small Silencing RNAs.* Cell, 2007. **130**(2): p. 299-308.
88. Guo, L. and Z. Lu, *The Fate of miRNA* Strand through Evolutionary Analysis: Implication for Degradation As Merely Carrier Strand or Potential Regulatory Molecule?* PLOS ONE, 2010. **5**(6): p. e11387.
89. Okamura, K., et al., *The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila.* Cell, 2007. **130**(1): p. 89-100.
90. Soares, Z.G., et al., *Viral RNA recognition by the Drosophila small interfering RNA pathway.* Microbes and Infection, 2014. **16**(12): p. 1013-1021.
91. Conte, D., Jr., et al., *RNA Interference in Caenorhabditis elegans.* Curr Protoc Mol Biol, 2015. **109**: p. 26.3.1-26.3.30.
92. Chen, W., et al., *A Genetic Network for Systemic RNA Silencing in Plants* Plant Physiology, 2018. **176**(4): p. 2700-2719.

93. Zamparini, A.L., et al., *Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in Drosophila.* Development, 2011. **138**(18): p. 4039-50.
94. Malone, C.D., et al., *Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary.* Cell, 2009. **137**(3): p. 522-35.
95. Malone, C.D. and G.J. Hannon, *Molecular evolution of piRNA and transposon control pathways in Drosophila.* Cold Spring Harb Symp Quant Biol, 2009. **74**: p. 225-34.
96. Watanabe, T. and H. Lin, *Posttranscriptional regulation of gene expression by Piwi proteins and piRNAs.* Mol Cell, 2014. **56**(1): p. 18-27.
97. Pillai, R.S. and S. Chuma, *piRNAs and their involvement in male germline development in mice.* Dev Growth Differ, 2012. **54**(1): p. 78-92.
98. Rojas-Rios, P., et al., *Aubergine and piRNAs promote germline stem cell self-renewal by repressing the proto-oncogene Cbl.* EMBO J, 2017. **36**(21): p. 3194-3211.
99. Khurana, J.S. and W.E. Theurkauf, *piRNA function in germline development*, in *StemBook*. 2008: Cambridge (MA).
100. Cornes, E., et al., *piRNAs initiate transcriptional silencing of spermatogenic genes during C. elegans germline development.* Developmental Cell, 2022. **57**(2): p. 180-196.e7.
101. Gutierrez, J., et al., *Evolutionary history of the vertebrate Piwi gene family.* PeerJ, 2021. **9**: p. e12451.
102. Sato, K. and M.C. Siomi, *The piRNA pathway in Drosophila ovarian germ and somatic cells.* Proc Jpn Acad Ser B Phys Biol Sci, 2020. **96**(1): p. 32-42.
103. Czech, B. and G.J. Hannon, *One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing.* Trends in Biochemical Sciences, 2016. **41**(4): p. 324-337.
104. Petit, M., et al., *piRNA pathway is not required for antiviral defense in Drosophila melanogaster.* Proc Natl Acad Sci U S A, 2016. **113**(29): p. E4218-27.
105. Ba, Z. and Y. Qi, *Small RNAs: emerging key players in DNA double-strand break repair.* Sci China Life Sci, 2013. **56**(10): p. 933-6.
106. Wei, W., et al., *A role for small RNAs in DNA double-strand break repair.* Cell, 2012. **149**(1): p. 101-12.
107. Bonath, F., et al., *Next-generation sequencing reveals two populations of damage-induced small RNAs at endogenous DNA double-strand breaks.* Nucleic Acids Res, 2018. **46**(22): p. 11869-11882.
108. Michalik, K.M., R. Bottcher, and K. Forstemann, *A small RNA response at DNA ends in Drosophila.* Nucleic Acids Res, 2012. **40**(19): p. 9596-603.
109. Vazquez, F., et al., *Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs.* Mol Cell, 2004. **16**(1): p. 69-79.
110. Peragine, A., et al., *SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis.* Genes Dev, 2004. **18**(19): p. 2368-79.
111. Rodrigues, D.F., et al., *In vitro mechanistic studies on α-amanitin and its putative antidotes.* Arch Toxicol, 2020. **94**(6): p. 2061-2078.
112. Hensold, J.O., D. Barth, and C.A. Stratton, *RNA polymerase II inhibitor, 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB) causes erythroleukemic differentiation and transcriptional activation of erythroid genes.* J Cell Physiol, 1996. **168**(1): p. 105-13.
113. Schmidts, I., et al., *Homology directed repair is unaffected by the absence of siRNAs in Drosophila melanogaster.* Nucleic Acids Res, 2016. **44**(17): p. 8261-71.
114. Merk, K., et al., *Splicing stimulates siRNA formation at Drosophila DNA double-strand breaks.* PLoS Genet, 2017. **13**(6): p. e1006861.
115. Liu, S., et al., *RNA polymerase III is required for the repair of DNA double-strand breaks by homologous recombination.* Cell, 2021. **184**(5): p. 1314-1329.e10.
116. Bottcher, R., et al., *RNA polymerase II is recruited to DNA double-strand breaks for dilncRNA transcription in Drosophila.* RNA Biol, 2022. **19**(1): p. 68-77.

117. Sekelsky, J., *DNA Repair in Drosophila: Mutagens, Models, and Missing Genes.* Genetics, 2017. **205**(2): p. 471-490.
118. Mota, M.B.S., et al., *DNA damage response and repair in perspective: Aedes aegypti, Drosophila melanogaster and Homo sapiens.* Parasites & Vectors, 2019. **12**(1): p. 533.
119. Békési, A., et al., *A novel fruitfly protein under developmental control degrades uracil-DNA.* Biochem Biophys Res Commun, 2007. **355**(3): p. 643-8.
120. Friedberg, E.C., *Cockayne syndrome--a primary defect in DNA repair, transcription, both or neither?* Bioessays, 1996. **18**(9): p. 731-8.
121. Sekelsky, J.J., et al., *Nucleotide excision repair endonuclease genes in Drosophila melanogaster.* Mutat Res, 2000. **459**(3): p. 219-28.
122. Nivard, M.J., A. Pastink, and E.W. Vogel, *Impact of DNA nucleotide excision repair on methyl methanesulfonate-induced mutations in Drosophila melanogaster.* Carcinogenesis, 1993. **14**(8): p. 1585-90.
123. Kovalchuk, I. and O. Kovalchuk, *Genome stability: from virus to human application*. Vol. 26. 2021: Academic Press.
124. Kunkel, T.A. and D.A. Erie, *Eukaryotic Mismatch Repair in Relation to DNA Replication.* Annu Rev Genet, 2015. **49**: p. 291-313.
125. Carpenter, A.T., *Mismatch repair, gene conversion, and crossing-over in two recombination-defective mutants of Drosophila melanogaster.* Proc Natl Acad Sci U S A, 1982. **79**(19): p. 5961-5.
126. Lopez-Martinez, D., C.C. Liang, and M.A. Cohn, *Cellular response to DNA interstrand crosslinks: the Fanconi anemia pathway.* Cell Mol Life Sci, 2016. **73**(16): p. 3097-114.
127. Yu, A.M. and M. McVey, *Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions.* Nucleic Acids Res, 2010. **38**(17): p. 5706-17.
128. Radford, S.J. and J.J. Sekelsky, *Taking Drosophila Rad51 for a SPiN.* Nature Structural & Molecular Biology, 2004. **11**(1): p. 9-10.
129. Yannuzzi, I., et al., *The Role of Drosophila CtIP in Homology-Directed Repair of DNA Double-Strand Breaks.* Genes, 2021. **12**(9): p. 1430.
130. Fernandez, J., et al., *Chromosome Preference During Homologous Recombination Repair of DNA Double-Strand Breaks in Drosophila melanogaster.* G3 (Bethesda), 2019. **9**(11): p. 3773-3780.
131. McVey, M., D. Radut, and J.J. Sekelsky, *End-joining repair of double-strand breaks in Drosophila melanogaster is largely DNA ligase IV independent.* Genetics, 2004. **168**(4): p. 2067-76.
132. Do, A.T., et al., *Double-Strand Break Repair Assays Determine Pathway Choice and Structure of Gene Conversion Events in Drosophila melanogaster.* G3 Genes|Genomes|Genetics, 2014. **4**(3): p. 425-432.
133. Gao, M., et al., *Ago2 facilitates Rad51 recruitment and DNA double-strand break repair by homologous recombination.* Cell Research, 2014. **24**(5): p. 532-541.
134. Wang, Q. and M. Goldstein, *Small RNAs Recruit Chromatin-Modifying Enzymes MMSET and Tip60 to Reconfigure Damaged DNA upon Double-Strand Break and Facilitate Repair.* Cancer Research, 2016. **76**(7): p. 1904-1915.
135. Shi, Y., *The Spliceosome: A Protein-Directed Metalloribozyme.* Journal of Molecular Biology, 2017. **429**(17): p. 2640-2653.
136. Jo, B.S. and S.S. Choi, *Introns: The Functional Benefits of Introns in Genomes.* Genomics Inform, 2015. **13**(4): p. 112-8.
137. Merkhofer, E.C., P. Hu, and T.L. Johnson, *Introduction to cotranscriptional RNA splicing.* Methods Mol Biol, 2014. **1126**: p. 83-96.
138. Borao, S., J. Ayté, and S. Hümmer, *Evolution of the Early Spliceosomal Complex-From Constitutive to Regulated Splicing.* Int J Mol Sci, 2021. **22**(22).

139. Zhan, X., et al., *Structures of the human pre-catalytic spliceosome and its precursor spliceosome.* Cell Research, 2018. **28**(12): p. 1129-1140.
140. Larson, J.D. and A.A. Hoskins, *Dynamics and consequences of spliceosome E complex formation.* eLife, 2017. **6**: p. e27592.
141. Hodson, M.J., et al., *The transition in spliceosome assembly from complex E to complex A purges surplus U1 snRNPs from alternative splice sites.* Nucleic Acids Res, 2012. **40**(14): p. 6850-62.
142. Agafonov, D.E., et al., *Molecular architecture of the human U4/U6.U5 tri-snRNP.* Science, 2016. **351**(6280): p. 1416-1420.
143. Nguyen, T.H.D., et al., *The architecture of the spliceosomal U4/U6.U5 tri-snRNP.* Nature, 2015. **523**(7558): p. 47-52.
144. Tholen, J. and W.P. Galej, *Structural studies of the spliceosome: Bridging the gaps.* Current Opinion in Structural Biology, 2022. **77**: p. 102461.
145. Bao, P., et al., *The RES complex is required for efficient transformation of the precatalytic B spliceosome into an activated B(act) complex.* Genes Dev, 2017. **31**(23-24): p. 2416-2429.
146. Robart, A.R., et al., *Crystal structure of a eukaryotic group II intron lariat.* Nature, 2014. **514**(7521): p. 193-197.
147. Golas, M.M., et al., *3D Cryo-EM Structure of an Active Step I Spliceosome and Localization of Its Catalytic Core.* Molecular Cell, 2010. **40**(6): p. 927-938.
148. Tresini, M., et al., *The core spliceosome as target and effector of non-canonical ATM signalling.* Nature, 2015. **523**(7558): p. 53-8.
149. Naro, C., et al., *The interplay between DNA damage response and RNA processing: the unexpected role of splicing factors as gatekeepers of genome stability.* Frontiers in Genetics, 2015. **6**.
150. Shkreta, L. and B. Chabot, *The RNA Splicing Response to DNA Damage.* Biomolecules, 2015. **5**(4): p. 2935-77.
151. Sherill-Rofe, D., et al., *Multi-omics data integration analysis identifies the spliceosome as a key regulator of DNA double-strand break repair.* NAR Cancer, 2022. **4**(2).
152. Oesterreich, F.C., et al., *Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II.* Cell, 2016. **165**(2): p. 372-381.
153. van den Boom, V., N.G.J. Jaspers, and W. Vermeulen, *When machines get stuck—obstructed RNA polymerase II: displacement, degradation or suicide.* BioEssays, 2002. **24**(9): p. 780-784.
154. Krietsch, J., et al., *PARP activation regulates the RNA-binding protein NONO in the DNA damage response to DNA double-strand breaks.* Nucleic Acids Res, 2012. **40**(20): p. 10287-301.
155. Edmond, V., et al., *Acetylation and phosphorylation of SRSF2 control cell fate decision in response to cisplatin.* Embo j, 2011. **30**(3): p. 510-23.
156. Matsuoka, S., et al., *ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage.* Science, 2007. **316**(5828): p. 1160-6.
157. Moumen, A., et al., *hnRNP K: an HDM2 target and transcriptional coactivator of p53 in response to DNA damage.* Cell, 2005. **123**(6): p. 1065-78.
158. Adamson, B., et al., *A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response.* Nat Cell Biol, 2012. **14**(3): p. 318-28.
159. Mastrocola, A.S., et al., *The RNA-binding protein fused in sarcoma (FUS) functions downstream of poly(ADP-ribose) polymerase (PARP) in response to DNA damage.* J Biol Chem, 2013. **288**(34): p. 24731-41.
160. Salton, M., et al., *Involvement of Matrin 3 and SFPQ/NONO in the DNA damage response.* Cell Cycle, 2010. **9**(8): p. 1568-76.

161. Haley, B., et al., *Response of heterogeneous ribonuclear proteins (hnRNP) to ionising radiation and their involvement in DNA damage repair.* Int J Radiat Biol, 2009. **85**(8): p. 643-55.
162. Filippov, V., M. Filippova, and P.J. Duerksen-Hughes, *The early response to DNA damage can lead to activation of alternative splicing activity resulting in CD44 splice pattern changes.* Cancer Res, 2007. **67**(16): p. 7621-30.
163. Colla, S., et al., *Telomere dysfunction drives aberrant hematopoietic differentiation and myelodysplastic syndrome.* Cancer Cell, 2015. **27**(5): p. 644-57.
164. Paulsen, R.D., et al., *A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability.* Mol Cell, 2009. **35**(2): p. 228-39.
165. Chanarat, S. and K. Strasser, *Splicing and beyond: the many faces of the Prp19 complex.* Biochim Biophys Acta, 2013. **1833**(10): p. 2126-34.
166. Bottcher, R., 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): p. e89.
167. Patel, V.J., et al., *A Comparison of Labeling and Label-Free Mass Spectrometry-Based Proteomics Approaches.* Journal of Proteome Research, 2009. **8**(7): p. 3752-3759.
168. Megger, D.A., et al., *Comparison of label-free and label-based strategies for proteome analysis of hepatoma cell lines.* Biochim Biophys Acta, 2014. **1844**(5): p. 967-76.
169. Cox, J., et al., *Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ.* Mol Cell Proteomics, 2014. **13**(9): p. 2513-26.
170. Hunter, S., et al., *InterPro: the integrative protein signature database.* Nucleic Acids Res, 2009. **37**(Database issue): p. D211-5.
171. Thomas, P.D., et al., *PANTHER: Making genome-scale phylogenetics accessible to all.* Protein Sci, 2022. **31**(1): p. 8-22.
172. Mi, H., A. Muruganujan, and P.D. Thomas, *PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees.* Nucleic Acids Res, 2013. **41**(Database issue): p. D377-86.
173. Thomas, P.D., et al., *PANTHER: a library of protein families and subfamilies indexed by function.* Genome Res, 2003. **13**(9): p. 2129-41.
174. Silva-Sousa, R., et al., *The Chromosomal Proteins JIL-1 and Z4/Putzig Regulate the Telomeric Chromatin in Drosophila melanogaster.* PLOS Genetics, 2012. **8**(12): p. e1003153.
175. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**(18): p. 3551-67.
176. Silva-Sousa, R., M.D. Varela, and E. Casacuberta, *The Putzig partners DREF, TRF2 and KEN are involved in the regulation of the Drosophila telomere retrotransposons, HeT-A and TART.* Mob DNA, 2013. **4**(1): p. 18.
177. Guruharsha, K.G., et al., *A protein complex network of Drosophila melanogaster.* Cell, 2011. **147**(3): p. 690-703.
178. Wong, K.H., Y. Jin, and Z. Moqtaderi, *Multiplex Illumina sequencing using DNA barcoding.* Curr Protoc Mol Biol, 2013. **Chapter 7**: p. Unit 7 11.
179. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.
180. Herai, R.H., P.D. Negraes, and A.R. Muotri, *Evidence of nuclei-encoded spliceosome mediating splicing of mitochondrial RNA.* Hum Mol Genet, 2017. **26**(13): p. 2590.
181. Lu, M. and W. Wei, *Proximity labeling to detect RNA–protein interactions in live cells.* FEBS Open Bio, 2019. **9**(11): p. 1860-1868.
182. Sharma, S., et al., *MRE11-RAD50-NBS1 Complex Is Sufficient to Promote Transcription by RNA Polymerase II at Double-Strand Breaks by Melting DNA Ends.* Cell Rep, 2021. **34**(1): p. 108565.

183. Todorova, T., D. Miteva, and S. Chankova, *DNA susceptibility of Saccharomyces cerevisiae to Zeocin depends on the growth phase.* Int Microbiol, 2019. **22**(4): p. 419-428.
184. Pfeifer, T.A., et al., *Baculovirus immediate-early promoter-mediated expression of the Zeocin resistance gene for use as a dominant selectable marker in dipteran and lepidopteran insect cell lines.* Gene, 1997. **188**(2): p. 183-90.
185. Ostling, O. and K.J. Johanson, *Microelectrophoretic study of radiation-induced DNA damages in individual mammalian cells.* Biochem Biophys Res Commun, 1984. **123**(1): p. 291-8.
186. Benyajati, C., A.R. Place, and W. Sofer, *Formaldehyde mutagenesis in Drosophila. Molecular analysis of ADH-negative mutants.* Mutat Res, 1983. **111**(1): p. 1-7.
187. Grafstrom, R.C., A. Fornace, Jr., and C.C. Harris, *Repair of DNA damage caused by formaldehyde in human cells.* Cancer Res, 1984. **44**(10): p. 4323-7.
188. Nadalutti, C.A., et al., *Mitochondrial dysfunction and DNA damage accompany enhanced levels of formaldehyde in cultured primary human fibroblasts.* Sci Rep, 2020. **10**(1): p. 5575.
189. Scurtu, F., et al., *Hemoglobin-albumin cross-linking with disuccinimidyl suberate (DSS) and/or glutaraldehyde for blood substitutes.* Artif Cells Nanomed Biotechnol, 2014. **42**(1): p. 13-7.
190. Hobro, A.J. and N.I. Smith, *An evaluation of fixation methods: Spatial and compositional cellular changes observed by Raman imaging.* Vibrational Spectroscopy, 2017. **91**: p. 31-45.
191. Tyanova, S. and J. Cox, *Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research.* Methods Mol Biol, 2018. **1711**: p. 133-148.
192. Ben Salem, K. and A. Ben Abdelaziz, *Principal Component Analysis (PCA).* Tunis Med, 2021. **99**(4): p. 383-389.
193. David, C.C. and D.J. Jacobs, *Principal component analysis: a method for determining the essential dynamics of proteins.* Methods Mol Biol, 2014. **1084**: p. 193-226.
194. Szklarczyk, D., et al., *STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets.* Nucleic Acids Res, 2019. **47**(D1): p. D607-D613.
195. Silva-Sousa, R., et al., *The chromosomal proteins JIL-1 and Z4/Putzig regulate the telomeric chromatin in Drosophila melanogaster.* PLoS Genet, 2012. **8**(12): p. e1003153.
196. Villasante, A., et al., *Drosophila telomeric retrotransposons derived from an ancestral element that was recruited to replace telomerase.* Genome Res, 2007. **17**(12): p. 1909-18.
197. Beagan, J.A. and J.E. Phillips-Cremins, *On the existence and functionality of topologically associating domains.* Nat Genet, 2020. **52**(1): p. 8-16.
198. Belokopytova, P. and V. Fishman, *Predicting Genome Architecture: Challenges and Solutions.* Front Genet, 2020. **11**: p. 617202.
199. Rowley, M.J. and V.G. Corces, *Organizational principles of 3D genome architecture.* Nat Rev Genet, 2018. **19**(12): p. 789-800.
200. Quinodoz, S.A. and M. Guttman, *Essential Roles for RNA in Shaping Nuclear Organization.* Cold Spring Harb Perspect Biol, 2022. **14**(5).
201. Zykova, T.Y., et al., *Polytene Chromosomes - A Portrait of Functional Organization of the Drosophila Genome.* Curr Genomics, 2018. **19**(3): p. 179-191.
202. Melnikova, L.S., et al., *The BEAF-32 Protein Directly Interacts with Z4/putzig and Chriz/Chromator Proteins in Drosophila melanogaster.* Dokl Biochem Biophys, 2021. **498**(1): p. 184-189.
203. Gan, M., et al., *The Chriz-Z4 complex recruits JIL-1 to polytene chromosomes, a requirement for interband-specific phosphorylation of H3S10.* J Biosci, 2011. **36**(3): p. 425-38.
204. Rath, U., et al., *The chromodomain protein, Chromator, interacts with JIL-1 kinase and regulates the structure of Drosophila polytene chromosomes.* J Cell Sci, 2006. **119**(Pt 11): p. 2332-41.
205. Eggert, H., A. Gortchakov, and H. Saumweber, *Identification of the Drosophila interband-specific protein Z4 as a DNA-binding zinc-finger protein determining chromosomal structure.* J Cell Sci, 2004. **117**(Pt 18): p. 4253-64.

206. Kugler, S.J. and A.C. Nagel, *putzig is required for cell proliferation and regulates notch activity in Drosophila.* Mol Biol Cell, 2007. **18**(10): p. 3733-40.
207. Kugler, S.J. and A.C. Nagel, *A novel Pzg-NURF complex regulates Notch target gene activity.* Mol Biol Cell, 2010. **21**(19): p. 3443-8.
208. Kugler, S.J., et al., *The Putzig-NURF nucleosome remodeling complex is required for ecdysone receptor signaling and innate immunity in Drosophila melanogaster.* Genetics, 2011. **188**(1): p. 127-39.
209. Zimmermann, M., et al., *Loss of putzig Activity Results in Apoptosis during Wing Imaginal Development in Drosophila.* PLoS One, 2015. **10**(4): p. e0124652.
210. Hoser, S.M., et al., *Intronic tRNAs of mitochondrial origin regulate constitutive and alternative splicing.* Genome Biol, 2020. **21**(1): p. 299.
211. Rector, R.S., R.M. Payne, and J.A. Ibdah, *Mitochondrial trifunctional protein defects: clinical implications and therapeutic approaches.* Adv Drug Deliv Rev, 2008. **60**(13-14): p. 1488-96.
212. Gibson, G.E., et al., *The alpha-ketoglutarate-dehydrogenase complex: a mediator between mitochondria and oxidative stress in neurodegeneration.* Mol Neurobiol, 2005. **31**(1-3): p. 43-63.
213. Ummethum, H. and S. Hamperl, *Proximity Labeling Techniques to Study Chromatin.* Front Genet, 2020. **11**: p. 450.
214. Samavarchi-Tehrani, P., R. Samson, and A.C. Gingras, *Proximity Dependent Biotinylation: Key Enzymes and Adaptation to Proteomics Approaches.* Mol Cell Proteomics, 2020. **19**(5): p. 757-773.
215. Roux, K.J., et al., *A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells.* J Cell Biol, 2012. **196**(6): p. 801-10.
216. Kim, D.I., et al., *An improved smaller biotin ligase for BioID proximity labeling.* Mol Biol Cell, 2016. **27**(8): p. 1188-96.
217. Wu, S.C., et al., *Design, production, and characterization of an engineered biotin ligase (BirA) and its application for affinity purification of staphylokinase produced from Bacillus subtilis via secretion.* Protein Expr Purif, 2002. **24**(3): p. 357-65.
218. Branon, T.C., et al., *Efficient proximity labeling in living cells and organisms with TurboID.* Nat Biotechnol, 2018. **36**(9): p. 880-887.
219. Cho, K.F., et al., *Proximity labeling in mammalian cells with TurboID and split-TurboID.* Nature Protocols, 2020. **15**(12): p. 3971-3999.
220. Rhee, H.W., et al., *Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging.* Science, 2013. **339**(6125): p. 1328-1331.
221. Di Marzo, N., E. Chisci, and R. Giovannoni, *The Role of Hydrogen Peroxide in Redox-Dependent Signaling: Homeostatic and Pathological Responses in Mammalian Cells.* Cells, 2018. **7**(10).
222. Ward, A.J. and T.A. Cooper, *The pathobiology of splicing.* J Pathol, 2010. **220**(2): p. 152-63.
223. Mills, J.D. and M. Janitz, *Alternative splicing of mRNA in the molecular pathology of neurodegenerative diseases.* Neurobiol Aging, 2012. **33**(5): p. 1012 e11-24.
224. Berger, K., et al., *Altered splicing associated with the pathology of inflammatory bowel disease.* Hum Genomics, 2021. **15**(1): p. 47.
225. Sune-Pou, M., et al., *Innovative Therapeutic and Delivery Approaches Using Nanotechnology to Correct Splicing Defects Underlying Disease.* Front Genet, 2020. **11**: p. 731.
226. Tang, Z., et al., *RNA-Targeting Splicing Modifiers: Drug Development and Screening Assays.* Molecules, 2021. **26**(8).
227. Monteys, A.M., et al., *Regulated control of gene therapies by drug-induced splicing.* Nature, 2021. **596**(7871): p. 291-295.
228. Sato, M., et al., *High antitumor activity of pladienolide B and its derivative in gastric cancer.* Cancer Sci, 2014. **105**(1): p. 110-6.

229. Miller-Wideman, M., et al., *Herboxidiene, a new herbicidal substance from Streptomyces chromofuscus A7847. Taxonomy, fermentation, isolation, physico-chemical and biological properties.* J Antibiot (Tokyo), 1992. **45**(6): p. 914-21.
230. Lagisetti, C., et al., *Pre-mRNA splicing-modulatory pharmacophores: the total synthesis of herboxidiene, a pladienolide-herboxidiene hybrid analog and related derivatives.* ACS Chem Biol, 2014. **9**(3): p. 643-8.
231. Butler, M.S., *Remediating cancer via splicing modulation.* J Med Chem, 2013. **56**(17): p. 6573-5.
232. Gamboa Lopez, A., et al., *Herboxidiene Features That Mediate Conformation-Dependent SF3B1 Interactions to Inhibit Splicing.* ACS Chem Biol, 2021. **16**(3): p. 520-528.
233. Yokoi, A., et al., *Biological validation that SF3b is a target of the antitumor macrolide pladienolide.* FEBS J, 2011. **278**(24): p. 4870-80.
234. Huang, F. and A.V. Mazin, *Targeting the homologous recombination pathway by small molecule modulators.* Bioorg Med Chem Lett, 2014. **24**(14): p. 3006-13.

7. Acknowledgements

I would like to thank my supervisor Professor Dr. Klaus Förstemann for his guidance during my work and his support of my scientific journey with many interesting discussions and a project that was very stimulating.

My dear friend and colleague Romy Böttcher I would like to thank them for her support from day one and her wish that I succeed. To Volker Nitschko and Selina Mußgnug, I would like to thank you for being role models and friends who have always been there for me. The sense of belonging inside our laboratory was very special.

I would like to thank Thomas Fröhlich and Julian Stingele for following and supporting my work throughout the years. Many of the colleagues from LMU I would like to thank, and working with them was a pleasure: Aleksandra Sarman-Grilc, Michael Engelschall, Gabriela Bittner, Michael Till, Louisa Papatheodorou, Johanna Turck, Andrea Gilmozzi, Joana Musial, Heidemarie Sieber, Birgitta Beatrix, Roland Beckmann, Tea Speljko, Mia Potocnjak, Chandni Kumar, Jennifer Wells, Ting Su, Josef Fischbock, Leona Chitou, Shuang Shuang, Cristian Rosales, and Jan Stöckl among many.

Big gratitude to my friends in Munich who gave me a sense of home: Simone Schymura, Sabrina Bartholl, Anja Knauseder, Annika Herrmann, Elisabeth Ringsgwandl, Alexandro Lagkadinos. For special moments I would like to thank my UNICEF family Charlotte Wendland, Mandy Seidenstrecker, and Charlotte von Rosenstiel. I would like to thank my dear friend from McKinsey time Lara Martin. Life in Munich gave me the privilege to know all of them and spend a wonderful time with them. I would also like to thank very much for all the years my childhood and university friends Jovana Milovanovic, Milena Stanojevic, Dusan Obradovic, Katarina Radisavljevic, and Mina Petrovic.

For the privilege to share their scientific vision today and wonderful projects each day I want to thank Prof. Dr. Ugur Sahin and Prof. Dr. Özlem Türeci. Jasmina Alatovic, my sister from another mother, for finding the best in me and helping me grow.

I want to thank my mother Milosava and my father Bojan for the very first moments I remember them picking me up from kindergarten. Thank you for letting me know I am valued, I am loved, and I am cherished. I want to thank my sister Jelena, for always being my biggest role model, a golden light, a spark of joy and meaning. Thank you.