# Mechanisms underlying mutational outcomes of DNA double-strand break repair

Kamp, J.A.

## Citation

Kamp, J. A. (2022, September 6). *Mechanisms underlying mutational outcomes of DNA double-strand break repair*. Retrieved from https://hdl.handle.net/1887/3455196

Version: Publisher's Version

License: Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden

Downloaded from: https://hdl.handle.net/1887/3455196

**Note:** To cite this publication please use the final published version (if applicable).

# CHAPTER 4

# Crosstalk between RNA surveillance and DNA repair via the moonlighting PIKK-family kinase SMG-1

J. A. Kamp*, B. B. L. G. Lemmens*, R. Gonzalez Prieto, J. Olsen, R. van Schendel, A.C. Vertegaal, M. Tijsterman

* equally contributed

Published in final form in *Nucleic Acids Research*, 2022

# ABSTRACT

The integrity and proper expression of genomes are safeguarded by DNA and RNA surveillance pathways. While many RNA surveillance factors have additional functions in the nucleus, little is known about the incidence and physiological impact of converging RNA and DNA signals. Here, using genetic screens and genome-wide analyses, we identified unforeseen SMG-1-dependent crosstalk between RNA surveillance and DNA repair in living animals. Defects in RNA processing, due to viable THO complex or PNN-1 mutations, induce a shift in DNA repair in dividing and non-dividing tissues. Loss of SMG-1, an ATM/ATR-like kinase central to RNA surveillance by nonsense-mediated decay (NMD), restores DNA repair and radio-resistance in THO-deficient animals. Mechanistically, we find SMG-1 and its downstream target SMG-2/UPF1, but not NMD *per se*, to suppress DNA repair by non-homologous end-joining in favour of single strand annealing. We postulate that moonlighting proteins create short-circuits *in vivo*, allowing aberrant RNA to redirect DNA repair.

# INTRODUCTION

DNA and RNA molecules are highly effective carriers of genetic information as demonstrated by their ubiquitous use in all domains of life. However, damages to the genetic code are inevitable and even the smallest error can derail cellular fate and impair an individual's fitness, as illustrated by the severe consequences of inherited diseases and cancer predispositions. To safeguard genetic integrity numerous conceptually different DNA and RNA surveillance mechanisms have evolved that detect, remove or resolve aberrant nucleic acid structures. One of the most detrimental damages a genome can encounter is a DNA double-strand break (DSB). In addition to disrupting the sequence of connected nucleic acid bases, DSBs expose fragile moieties within the nucleic acid structure that are vulnerable to nuclease attacks, often causing further loss of genetic information. Repair of DSBs requires multiple complex reactions, which makes it inherently error-prone and/or sensitive to external cues. The mutagenic nature of some DSB repair pathways is a fundamental hazard to human health, causing genetic scars ranging from a single base pair mutation to extensive chromosomal rearrangements involving thousands of base pairs(Hanscom and McVey 2020). The efficacy of DSBs to cause genetic damage can however also be exploited to improve health. Many anticancer therapies are based on induction of DSBs and DSBs are key intermediates for genome modification and gene therapy approaches (including those based on CRISPR Cas9 technology).

One of the earliest responses to a DSB is the activation of ATM and/or ATR, two well-conserved members of a family of phosphatidylinositol 3-kinase-related protein kinases (PIKK) that amplify the stress signal, delay cell proliferation and coordinate DNA repair. Recent evidence shows that crosstalk occurs between PIKK family members, and that deregulation of PIKK signalling changes DSB repair pathway choice(Zhou, Lee et al. 2017, Balmus, Pilger et al. 2019, Lanz, Dibitetto et al. 2019, Mladenov, Fan et al. 2019). Multiple DNA repair routes have evolved that sense and resolve DSBs depending on the nature of the lesion, the genomic and chromatin environment, the cell cycle stage, and the developmental context. For metazoan cells multiple distinct DSB repair routes have been described: Non-Homologous End-Joining (NHEJ), Homologous Recombination (HR), Single Strand Annealing (SSA), Microhomology Mediated End-Joining (MMEJ) and polymerase Theta-Mediated End-Joining (TMEJ)(Schimmel, van Schendel et al. 2019, Hanscom and McVey 2020). While a versatile toolbox promotes DNA repair flexibility, it also requires strict coordination between the different alternatives. In fact, defective DNA repair pathway choice can be highly toxic and cause lethal levels of chromosomal abnormalities(Lemmens, Johnson et al. 2013). How the different PIKK family members control the relative activities of these DSB repair routes is still poorly understood.

While DNA surveillance mechanisms detect a wide range of DNA damages, they do not detect mistakes in the DNA code (*e.g.*, base changes that do not prevent sealing of the break or distort the DNA helix). However, eukaryotes also evolved RNA surveillance mechanisms that inspect messenger RNA (mRNA) for errors before they have the opportunity to produce faulty proteins in bulk. For instance, the RNA surveillance mechanism called Nonsense-Mediated mRNA Decay (NMD) cleaves and eliminates mRNAs that have a premature termination codon (PTC). PTCs cause mRNAs to encode for truncated proteins that are often non-functional or have dominant-negative effects, making PTCs particularly harmful mutations that cause disease with high penetrance(Atkinson and Martin 1994). Interestingly, NMD requires a PIKK very similar to ATM/ATR called SMG-1 (Figure S1). To trigger the degradation of non-sense mRNAs, SMG-1 phosphorylates the DNA/RNA helicase SMG-2/UPF1, which recruits a protein complex that cleaves the target mRNA. Like ATM and ATR, SMG-1 preferentially phosphorylates serine and threonine residues upstream of glutamine residues (*i.e.*, S/TQ motifs). S/TQ motifs are enriched in the NMD target SMG-2/UPF1 but also in many factors involved in genome surveillance and DNA repair. Intriguingly, human ATM and SMG1 phosphorylate identical residues in tumour suppressor p53 and SMG-2/UPF1, and combined loss of ATM and SMG-1 increases haematopoietic cancer incidence(Ho, Luff et al. 2019). This redundancy in phosphorylation targets implies possible signalling interference between NMD and the DNA damage response (DDR). Large-scale proteomic analyses of proteins phosphorylated in response to DNA damage identified over 700 human proteins modified at S/TQ sites, of which surprisingly nearly a third was implicated in RNA processing and transcription (including various RNA splicing and NMD factors)(Matsuoka, Ballif et al. 2007). While these phosphorylations are mostly accredited to ATM and ATR activity, the impact of SMG-1 activity might be undervalued and explain the strong bias towards RNA metabolism, given that the list of DNA damage-induced phospho-proteins includes established SMG-1 targets (including SMG-2/UPF1), SMG-1 has similar substrate specificities and can be activated by DNA damage(Brumbaugh, Otterness et al. 2004). Moreover, SMG-1 deficiency prevents NMD and increases susceptibility to DNA damage in animals(González-Huici, Wang et al. 2017) as well as plants(Lloyd, Lang et al. 2018), advocating for a preserved function in both RNA and DNA surveillance.

During the past decades it has become apparent that many enzymes and signalling factors do not perform a single tailored reaction in the cell, but perform additional functions, a phenomenon termed moonlighting. Moonlighting functions can be surprisingly different and often involve different compartments within the cell or organism, which also explains why they are typically discovered by chance or via unbiased screens. While moonlighting proteins provide economical means to expand the range of cellular functions with a limited set of genes, cells have to tolerate the added dimension to cellular complexity and/or evolve regulatory mechanisms to switch between the distinct protein functions. Having moonlighting proteins controlling both RNA and DNA surveillance could promote functional crosstalk, but might also cause signalling conflicts that jeopardize genome stability. To date, surprisingly little is known about the occurrence or molecular impact of such short-circuits *in vivo*.

The existence of shared signalling nodes (like SMG-1) predicts that the cellular abundance, location and/or activity of NMD factors might influence critical processes beyond RNA surveillance. Here we provide empirical evidence of this prediction, and reveal that mild changes in RNA processing can activate the moonlighting kinase SMG-1, which causes a shift in DNA repair and compromises genome stability. A combination of unbiased DNA repair screens and genome-wide analyses in developing animals allowed us to identify a systemic, potentially harmful link between RNA processing and NHEJ, which is mediated by the PIKK SMG-1.

# RESULTS

## Dual reporter system to measure NHEJ activity in Caenorhabditis elegans

In recent years, *C. elegans* has become a powerful model to study DNA damage responses in a developmental context, leading to the discovery of new DSB repair mechanisms in somatic as well as germline tissues(Lemmens and Tijsterman 2011, Roerink, van Schendel et al. 2014). While germline tissues strictly use homology-directed DSB repair(van Schendel, Roerink et al. 2015, Kamp, van Schendel et al. 2020), somatic tissues use different and partly competing repair routes depending on the nature of the break and the proliferative state of the tissue(Clejan, Boerckel et al. 2006, Pontier and Tijsterman 2009, Lemmens, Johnson et al. 2013). NHEJ is considered the principle DSB repair route in somatic tissues, yet little is known about its regulation *in vivo*(Clejan, Boerckel et al. 2006). A key obstacle in NHEJ research is the existence of alternative end-joining mechanisms, which mask the defects in NHEJ activity and obscure interpretations of DSB repair assays. To study DSB repair in somatic tissues we previously created animals in which we induced site-specific DSBs in pathway-specific reporter systems(Pontier and Tijsterman 2009, Johnson, Lemmens et al. 2013). Here, we set out to create a system able to directly detect NHEJ efficacy in single animals and perform an unbiased search for NHEJ regulators. We designed a fluorescent reporter system based on two key features of classical NHEJ: its error-prone nature (often leaving small deletions/

insertions) and its prevalence in non-dividing tissues. In brief, the NHEJ reporter is based on mutagenic repair of a targeted DSB and consequent restored expression of a downstream GFP/LacZ open reading frame (Figure 1A). NHEJ reporter expression is driven by a tissue-specific *myo-2* promoter, limiting GFP/LacZ expression to pharyngeal muscle cells. Pharyngeal muscle cells are terminally differentiated at the first larval stage and thus rely heavily on classical NHEJ(Clejan, Boerckel et al. 2006). Site-directed DSB induction was achieved using a heat-shock inducible, mCherry-tagged I-SceI endonuclease (Figure 1B). One day after I-SceI induction, nearly 100% of the animals carrying a multicopy NHEJ reporter transgene showed GFP expression, which was fully dependent on the classical NHEJ factor LIG-4 (Figure 1C). The robust induction of GFP expression and the large difference in NHEJ activity between the genetic conditions provides a solid window for an unbiased search for NHEJ modulators. We reasoned, however, that whilst loss of GFP expression could reflect a specific loss of NHEJ, it could also reflect reduced I-SceI expression, loss of the reporter, or other general defects in DSB induction or repair. To discriminate between these scenarios and find bona fide NHEJ regulators, we made use of a previous observation demonstrating that defects in somatic NHEJ result in a stark increase in compensating DSB repair pathways such as SSA, provided that the DSBs is surrounded by stretches of identical sequence(Pontier and Tijsterman 2009). We thus generated a more advanced NHEJ reporter animal that carries both the reporter specific to monitoring NHEJ as well as a well-characterized SSA reporter(Pontier and Tijsterman 2009) (Figure 1D). These dual reporter animals detect NHEJ activity directly, by measuring NHEJ in nonreplicating pharyngeal cells, as well as indirectly, by measuring SSA in various replicating somatic tissues (Figure 1E). As predicted, loss of NHEJ activity in these animals resulted in a dramatic shift in LacZ staining pattern (Figure 1F). While control animals showed robust LacZ staining in the pharynx but rarely in other somatic tissues (reflecting relatively high NHEJ and low SSA activities, respectively), *lig-4* mutants showed no LacZ staining in the pharynx but strong LacZ restoration in other somatic tissues (reflecting NHEJ deficiency and an ensuing shift towards SSA). As both the NHEJ and SSA readouts depend on I-SceI activity, shifts in LacZ patterns are indicative of changes post DSB induction and also discriminate general DSB repair defects from specific changes in NHEJ activity.

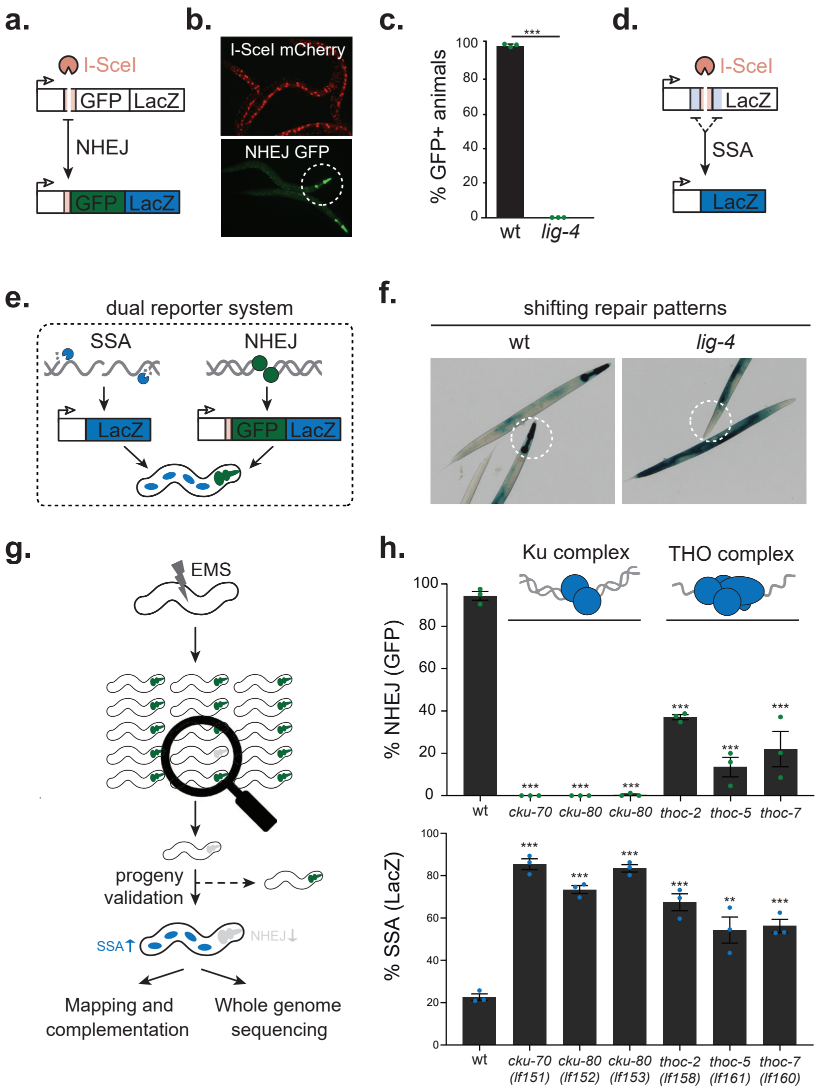

**Figure 1.** ***In vivo*** **NHEJ reporter identifies role of Ku and THO complex in DSB repair**

**a.** Schematic diagram of our NHEJ reporter based on heat-shock-inducible expression of mCherry::ISceI and break induction at the ISceI target site in a multi-copy integrated reporter transgene. Mutagenic NHEJ is measured by GFP/LacZ ORF correction. **b.** Representative pictures of animals expressing nuclear mCherry::I-SceI (6 hours post heat-shock), pharyngeal GFP (3 days post heat-shock). Circle indicates a representative pharynx. **c.** Quantification of GFP-positive pharynges in NHEJ-proficient wild-type and NHEJ-deficient *lig-4* deficient reporter animals, heat-shocked for 180 minutes at L1 stage and measured in adults. Average of three populations (n>200) is depicted. Dots indicate the average of each experiment (two-tailed t-tests *P < 0.05; **P < 0.01, ***P < 0.001). **d.** Schematic diagram of the Singe Strand Annealing (SSA) reporter that serves as an indirect NHEJ activity assessment tool. Heat-shock-inducible expression of mCherry::ISceI leads to break induction at the I-SceI target site in a multi-copy integrated reporter transgene. SSA is measured by LacZ ORF correction that is established by repair of the induced break using annealing of flanking stretches of homologous nucleotides. **e.** Schematic diagram of the dual reporter system. The NHEJ reporter is expressed in non-dividing pharyngeal muscle cells, leading to GFP and LacZ production in the pharynx of NHEJ-proficient nematodes after break induction and erroneous repair. The SSA reporter is expressed in dividing somatic cells, leading to LacZ expression in the soma after SSA of induced breaks. **f.** Representative pictures NHEJ-proficient (wild-type) and NHEJ-deficient (*lig-4*) dual reporter animals. In NHEJ-proficient animals, NHEJ reporter activity is visible as LacZ staining throughout the pharynx, while SSA staining is low in the dividing somatic cells in the body. In NHEJ-deficient animals, LacZ staining is absent in the pharynx, but SSA activity is increased in replicating cells, leading to increased LacZ staining in the soma. Circles indicate representative pharynges. **g.** Forward genetics screen set-up. Genomes of animals carrying the dual reporter system were mutagenized using ethyl methanesulfonate (EMS). NHEJ activity of the progeny was assessed using the NHEJ reporter and GFP-negative worms were isolated. Reporter animals of which NHEJ impairment was verified using LacZ staining were followed up to identify causal mutations. **h.** Quantification of GFP-positive pharynges and non-pharyngeal LacZ staining of identified Ku or THO complex mutants and wild-type animals. Average of three populations (n>200) is depicted. Dots indicate the average of each experiment. Error bars represent SEM. Asterisks indicate statistically significant difference compared to wildtype controls (two-tailed t-tests $^{*}P < 0.05$; $^{**}P < 0.01$, $^{***}P < 0.001$).

## Identifying novel regulators of NHEJ

Having validated the reagents, we next searched for genetic conditionals that control NHEJ activity. To this end, we performed a forward genetics screen in which we induced random genomic mutations in dual reporter animals by ethyl methanesulfonate (EMS) and assessed NHEJ activity among mutant progeny (*i.e.*, complex F3 populations were screened for animals with reduced pharyngeal GFP signal) (Figure 1G). These "GFP-low" animals were selected and their clonal progeny was again tested for NHEJ activity to identify heritable traits that affect pharyngeal GFP expression. To exclude false-positive NHEJ mutants we checked for unaffected mCherry-I-SceI expression and monitored SSA activity in all mutant candidates. Only NHEJ mutants that also showed increased SSA activity (indicative of a specific repair defects post DSB induction) were selected for further analysis, including genetic complementation assays and whole-genome sequencing (Figure 1G). By screening ~9000 unique haploid genomes we found seven *bona fide* NHEJ mutants: four showing reduced GFP expression and three showing no pharyngeal GFP at all. We quantified both NHEJ activity (GFP expression) and SSA activity (somatic LacZ expression) of synchronized clonal populations and observed the expected inverse correlation between NHEJ defect severity and increased compensatory SSA activity (Figure 1H). By combining whole-genome sequencing and classical complementation assays using null alleles of the known NHEJ genes *lig-4*, *cku-80* and *cku-70*, we identified the causal mutations in the three mutants devoid of NHEJ activity: alleles *lf152* and *lf153* were caused by nonsense mutations in *cku-80* and allele *lf151* was caused by a nonsense mutation in *cku-70* (Figure S2) The identification of the canonical NHEJ complex Ku using this unbiased approach validated both our dual reporter system as well as our screening setup. Interestingly, the modifier alleles *lf158*, *lf159*, *lf160* and *lf161* were not caused by mutations in known canonical NHEJ genes (Figure S3). To identify the chromosomal regions linked to the NHEJ defects we performed classical positional mapping using single nucleotide polymorphisms

(SNPs) between the Bristol dual reporter mutants and a Hawaiian mapping strain CB4856. In parallel, we performed genome-wide sequencing to find non-synonymous SNPs in the mapped regions. Strikingly, we found that three out of four modifier mutants mapped to a subunit of the very same multi-protein complex, the THO complex, strongly suggesting that defective THO function causes reduced NHEJ activity. Indeed, rescue experiments using extrachromosomal expression arrays confirmed that the alleles *lf158*, *lf160* and *lf161* were caused by defective *thoc-2*, *thoc-7* and *thoc-5* function, respectively (Figure S4). *Thoc-2* depletion by RNA interference also mirrored the changes in DSB repair (Figure S4). THO is a well-conserved RNA binding complex that controls processing and nuclear export of mRNA(Luna, Rondón et al. 2012). THO also has been implicated in safeguarding genome integrity, yet the mechanisms underlying the latter are still unclear(Salas-Armenteros, Pérez-Calero et al. 2017). THO deficiency causes genome instability in part through increased levels of RNA:DNA hybrids (or R-loops), which in dividing cells cause impaired DNA replication and DNA damage(Dominguez-Sanchez, Barroso et al. 2011, Castellano-Pozo, Garcia-Muse et al. 2012, Salas-Armenteros, Pérez-Calero et al. 2017, Marabitti, Lillo et al. 2019, Matos, Zhang et al. 2020). Notably, our data indicate that THO deficiency can change DNA repair in non-dividing cells, and thus may trigger genome instability independent of DNA replication.

## THO deficiency impairs NHEJ and RNA processing

To verify if the THO complex is needed for efficient repair of DSBs throughout the genome (and independent of I-SceI, GFP or LacZ expression), we exposed the mutant animals to ionizing radiation (IR). IR causes DSBs in a dose-dependent manner and can be used to study tissue-specific DSB repair(Craig, Moser et al. 2012). Different tissues have different DNA repair capacities and thus IR tolerance needs to be examined in a tissue-specific manner. In *C. elegans,* IR tolerance of germ cells can be examined via a so-called 'L4 assay', while somatic tissues can be tested via a 'L1 assay' (Figure S5). In a L4 assay germ cells of adolescent L4 larva are irradiated and progeny survival is analysed, while in a L1 assay synchronized L1 larva are irradiated and tracked for developmental defects (often focussing on well-defined organs such as the vulva, of which the precursor cells are not replicating during the L1 stage). Animals lacking canonical NHEJ are exceptionally sensitive to IR in the L1 assay, but not in the L4 assay, reflecting the key role of NHEJ in somatic cells (refs). In contrast, mutants defective in HR-mediated repair are typically hypersensitive to IR in the L4 assay, consistent with the strong HR bias in germ cells(Clejan, Boerckel et al. 2006, Robert, Davis et al. 2008, Johnson, Lemmens et al. 2013). Indeed, we find animals deficient for BRD-1 (ortholog of the human HR factor BARD1) to be hypersensitive to IR in the L4 assay, while animals lacking the NHEJ factor CKU-80 behaved like wild-type animals (Figure S5). Similar to canonical NHEJ mutants, germ cells of *thoc-5* and *thoc-7* animals did not show hypersensitivity to IR, indicating that defective THO function does not sensitize genomes to IR *per se* (Figure S5). Yet, when we subjected these animals to the L1 assay, we observed the inverse pattern. While irradiated *brd-1* mutants showed only minor somatic defects, irradiated *cku-80*, *thoc-5* and *thoc-7* mutants showed acute sensitivity, resulting into ~75%, ~65% and ~60% of animals having defective vulva development upon 60Gy of IR, respectively, while only ~10% of wild-type animals was affected at this dose (Figure 2B). Analogous to *cku-80* mutants, irradiated *thoc-5* mutants developed various IR-dependent somatic defects, including protruding vulvas, ruptured vulvas and so-called "bag-of-worms" where progeny hatches within the mother because of an egg laying defect (Figure 2C).

Epistasis analysis revealed that the *thoc-5* single mutant and *thoc-5*;*thoc-7* double mutant is equally sensitive to IR, implying that THOC-5 and THOC-7 act in the same pathway/complex. Similarly, *cku-80* single mutants and *cku-80; thoc-7* double mutants are equally sensitive to IR, indicating that the THO complex affects IR tolerance via NHEJ. Given that *thoc-7* deficiency does not increase the number of somatic defects in NHEJ null mutants (with or without IR), viable THO mutations appear to change DSB repair pathway choice without causing a large number of extra DSBs. In accordance with that notion, we find that the increased frequency of SSA in *thoc-7* mutants relies on I-SceI induced DSBs, and thus is not caused by more spontaneous DSBs (Figure S6).

The GFP reporter and IR studies also demonstrate that impaired THO function alters NHEJ efficacy in pharyngeal cells as well as vulval precursor cells (but not in germ cells), which together with the data from the somatic SSA reporter hints towards the existence of an unexplored near-systemic modulator of DNA repair fidelity. Our transgenic NHEJ reporter system creates targeted DSBs in nearly all somatic tissues, but only in pharyngeal cells this will be translated into GFP expression. We reasoned that THO deficiency likely altered DSB repair in many more cells than was visualized by the NHEJ reporter. To test this hypothesis, we developed a systemic PCR-based assay to quantify mutagenic repair at the I-SceI target site. In short, I-SceI expression was induced in synchronized L1 larva and DNA repair products were PCR-amplified and subjected to next-generation sequencing analysis (Figure 2D). In line with a widespread defect in NHEJ, both *lig-4* and *thoc-5* mutants showed a marked drop in the frequency of NHEJ-associated repair footprints (Figure 2E and 2F). Interestingly, *lig-4* mutant animals show a significant shift towards micro-homology driven DSB repair (i.e. 1-10bp homology at break junctions) while *thoc-5* mutants do not, suggesting that *thoc-5* deficiency causes a defect in NHEJ as well as micro-homology-driven repair routes such as TMEJ (Figure 2G). Thus DNA repair footprint analyses as well as the IR sensitivity studies indicate that THO deficiency causes a significant but selective defect in DSB repair, which limits NHEJ (and potentially TMEJ) but allow SSA and meiotic recombination – the latter as THO mutants display normal fertility and intact diakinesis chromosomes (Figure S5 and S7).

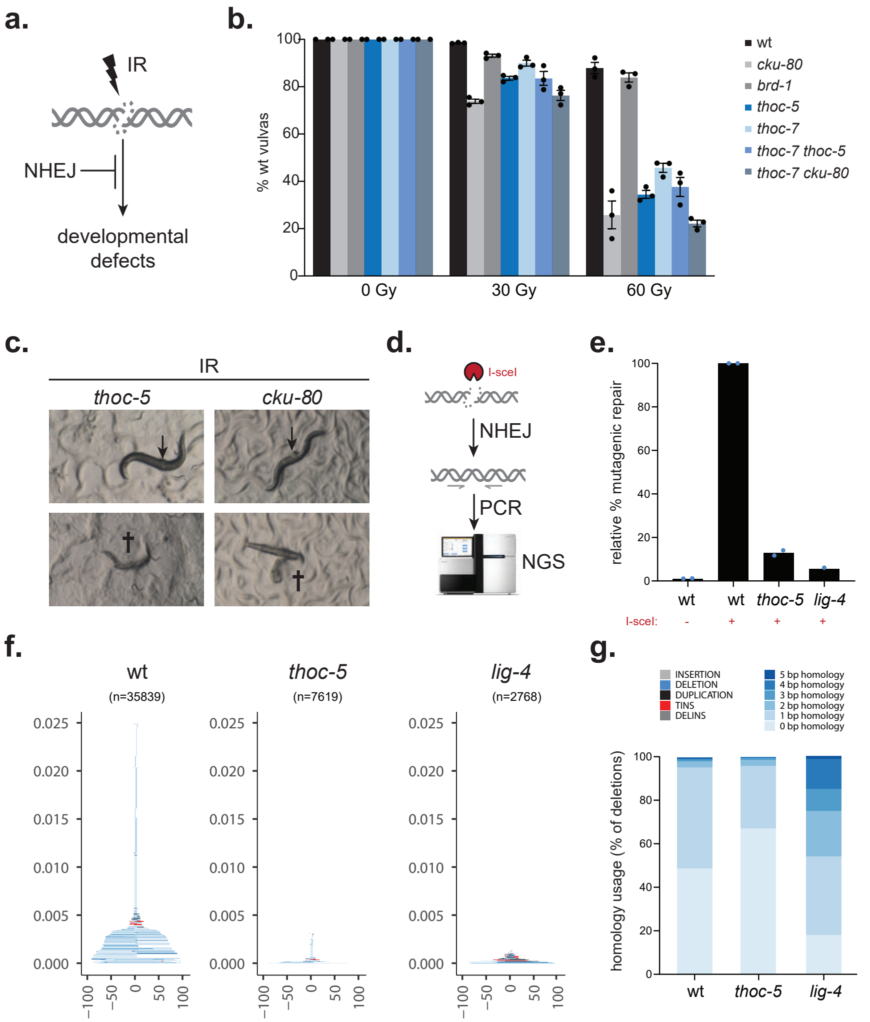

**Figure 2. THO deficiency causes radiation sensitivity and DNA end joining defects**

**a.** NHEJ protects nematodes from ionizing irradiation (IR)-induced developmental defects, including abnormal development of the vulva. **b.** Larvae were challenged with indicated doses of IR and vulva development was scored. Values depict averages of three independent experiments and error bars represent S.E.M. Dots indicate the average of each experiment. **c.** Examples of defective vulva development as a consequence of IR in *thoc-5* and *cku-80* animals. Arrows show protruding vulvas, crosses indicate lethality due to germline extrusions or internal hatching of progeny. **d.** Schematic representation of PCR-based assay to quantify mutagenic repair. Breaks were induced at the NHEJ reporter by expressing I-SceI. Genomic DNA was isolated and PCR reactions are carried out with primers surrounding the break site. Amplicons were analysed using next generation sequencing (NGS). **e.** Quantification of mutagenic DSB repair. Repair efficiency was normalized to the number of mutagenic repair events after I-SceI induction in wildtype worms. Dots represent averages of each replicate. **f.** Tornado plots depict type and size of mutagenic repair events identified by NGS (n indicates number of mutagenic events). Color code indicates type of event as in figure 3g. **g.** Relative distribution of deletion types found by NGS in wildtype, *thoc-5* or *lig-4* reporter animals.

Given the established role of THO in mRNA processing we searched the transcriptome for clues on how this complex might skew DNA repair activities. We performed RNAseq analysis on synchronized L1 animals and identified robust changes in general transcript levels and alternative transcript isoforms (Figure 3A). How THO selects its mRNA targets is currently unknown, but its selectivity is remarkably well preserved during metazoan evolution(Guria, Tran et al. 2011, Saran, Tran et al. 2013, Tran, Saran et al. 2013, Wang, Miao et al. 2013, Tran, Saran et al. 2014, Maeder, Kim et al. 2018). Analogous to other model organisms, *thoc-5* or *thoc-7* deficiency does not cause an overt change in total RNA levels (Figure S8), but alters the expression of a highly specific subset of genes (<5% of the transcriptome). As expected, we find a significant overlap between *thoc-5* and *thoc-7* mutants in both down- (67%) and upregulated mRNAs (71%) (Figure 3B). None of the shared downregulated transcripts, however, have been implicated in DDR signalling, DNA metabolism or repair  (Figure S9). Among the shared upregulated transcripts (Figure S10) are three factors implicated in DNA metabolism: *atl-1*(~1.7 fold), *gei-17* (~1.4 fold) and *rev-1* (~1.9 fold), possibly reflecting a response to DNA replication stress(Holway, Kim et al. 2006). Previous studies found elevated levels of ATL-1 foci and DNA replication stress-associated ubiquitination in sterile *thoc-2* deletion mutants(Castellano-Pozo, Garcia-Muse et al. 2012, Castellano-Pozo, García-Muse et al. 2012). Defective mRNA processing is often associated with R-loop formation, perhaps due the ineffective extraction of nascent RNA away from its DNA template(Salas-Armenteros, Pérez-Calero et al. 2017). We quantified global R-loop levels using dot blots and found that the genomes of our identified *thoc-2*, *thoc-5* and *thoc-7* mutants indeed have elevated levels of RNA:DNA hybrids (Figure 3B and 3C). While these R-loops can trigger DNA replication stress and possibly alter DNA repair, the mechanism by which THO deficiency controls NHEJ does not rely on DNA replication (as it occurs in differentiated pharyngeal cells and arrested L1 animals). Detailed analysis of RNA expression changes within individual transcripts confirmed widespread yet highly selective alterations in mRNA transcription, splicing and/or degradation in *thoc-5* or *thoc-7* animals (Figure S8). The most frequent RNA processing defects observed among *thoc-5* or *thoc-7* mutants involve skipped exons or retained introns, which supports a role for THO in mRNA maturation (Figure 3E). Multivariate analysis of transcript splicing (MATS) identified 90 skipped exons and 18 retained introns in *thoc-5* mutants, and 27 skipped exons and 10 retained introns in *thoc-7* mutants. More than half of these splicing defects were shared between the two THO mutants, yet none of the shared defects prevents expression of known DNA repair factors (Figure S11). Splicing defects often disrupt the reading-frame and thus are potent triggers of PTCs and subsequent NMD activation(Neu-Yilik, Gehring et al. 2004). Transcript analysis at the nucleotide level confirmed robust expression of PTC-

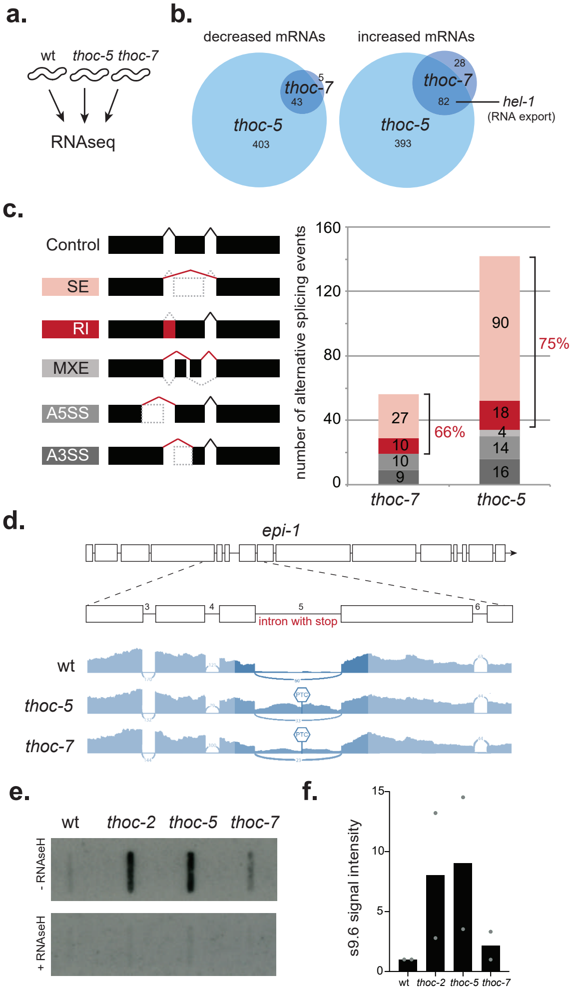

**Figure 3. THO deficient animals suffer from RNA splicing errors and R-loops**

**a.** RNA sequencing and direct comparison of NHEJ-deficient THO mutants. **b.** Venn diagrams of significantly (q ≤ 0.05) downregulated transcripts (left) and significantly (q ≤ 0.05) upregulated transcripts (right) identified using RNA sequencing. **c.** Representative slot blots to quantify R-loops in THO mutants. Nucleic acid extracts of THO mutant worms show increased s9.6 antibody staining. Specificity of the s9.6 antibody signal was confirmed by RNAse H treatment to digest DNA:RNA hybrids. **d.** Quantification of R-loop levels. s9.6 antibody signal intensity was normalized to WT and RNAse-H digested controls. Dots represent values of each experiment. **e.** Schematic representation of the different splicing events detected by rMATS: SE, skipped exon; RI, retained intron; MXE, mutually exclusive exon; A5SS, alternate 5' splice site; A3SS, alternate 3' splice site. Bar chart indicates number of splicing events detected and altered in *thoc-5* or *thoc-7* animals compared to wildtype controls. Color code is the same as in schematic on the left. **f.** Typical example of a selective splicing error in THO mutants, causing a retained intron and robust expression of PTC-bearing mRNAs. Gene model of *epi-1* and below (in blue) simplified sashimi plots for each genotype. Sashimi plots depict relative RNAseq read-depth around the highlighted region of *epi-1*; both *thoc-5* and *thoc-7* mutants selectively retain the 5th intron of *epi-1* transcripts. The predicted position of the first PTC is indicted.

bearing mRNAs in the THO mutants (Figure 3F and Figure S12). Our unbiased DSB repair screen thus identified viable THO mutants that suffer from a selective DNA repair defect and multiple RNA stresses (including RNA:DNA hybrids and non-sense RNAs).

## PNN-1 and the THO complex are needed for efficient NHEJ

The DSB repair screen yielded one additional mutant (allele *lf159*), which based on epistasis and mapping studies was not affected in known NHEJ genes nor in THO genes. Instead, we found a causal nonsense mutation in *pnn-1* (R186.7), the *C. elegans* ortholog of the mRNA splicing and export factor Pinin/DRS/memA (Figure S13). Animals with an independently derived *pnn-1* deletion allele (*ok1872*) confirmed a reduction in NHEJ, a shift towards SSA and hypersensitivity to IR (Figure 4A and 4B). Comparative analysis using *thoc-5; pnn-1* double mutants showed a mild, yet reproducible additive defect in NHEJ, hinting towards a shared trigger and a common theme between defective RNA processing and NHEJ regulation. To map the degree of functional overlap between PNN-1 and THO components we directly compared the transcriptome profiles of synchronized *pnn-1*, *thoc-5* and *thoc-7* animals by RNAseq. Similar to THO deficiency, *pnn-1* loss affects only a selective subset of the transcriptome (<3%). *pnn-1* mutants share nearly a third of the up-regulated transcripts and half of the down-regulated transcripts with the THO mutants (Figure 4C and 4D). Interestingly, *thoc-5*, *thoc-7* and *pnn-1* mutants show increased transcript levels of *hel-1*/UAP56 (Figure 4D), which is a well-conserved RNA-DNA helicase known to control mRNA export and interact with THO complex proteins(Shen 2009). In flies, UAP56 up-regulation is shown to be part of a feedback mechanism that responds to a block in mRNA maturation and export, and human UAP56 is recently shown to resolve RNA-DNA hybrids genome-wide(Herold, Teixeira et al. 2003, Pérez-Calero, Bayona-Feliu et al. 2020). Elevated *hel-1*/UAP56 expression thus might indicate a common RNA stress in these animals.

rMATS analysis indeed revealed various splicing defects in *pnn-1* mutants compared to N2 controls, including skipped exons and retained introns (Figure S13). Notably, pnn-1 mutants share more than a third of splicing alterations with thoc-5 and/or thoc-7 animals, substantiating a common role for PNN-1 and THO in RNA surveillance (Figure S8 and S13). In-depth analysis at the nucleotide level confirmed matching RNA expression changes in *pnn-1* and *thoc-5* mutants; e.g. both mutants show enhanced retention of the first intron of *tos-1*, which is a well-characterized endogenous reporter gene for splicing perturbations(Ma, Gao et al. 2012). These analyses also revealed gene-specific effects, as only THO mutants show an additional defect at the 3' end of *tos-1* transcripts (Figure 4E). PNN-1 and THO thus have non-identical yet highly similar functions in RNA processing, which parallels their effect on NHEJ. The identification of PNN-1 and THO in NHEJ regulation implies an extensive connection between RNA surveillance and DNA repair and suggests that unprocessed RNA molecules may confound DNA repair (Figure 4F).

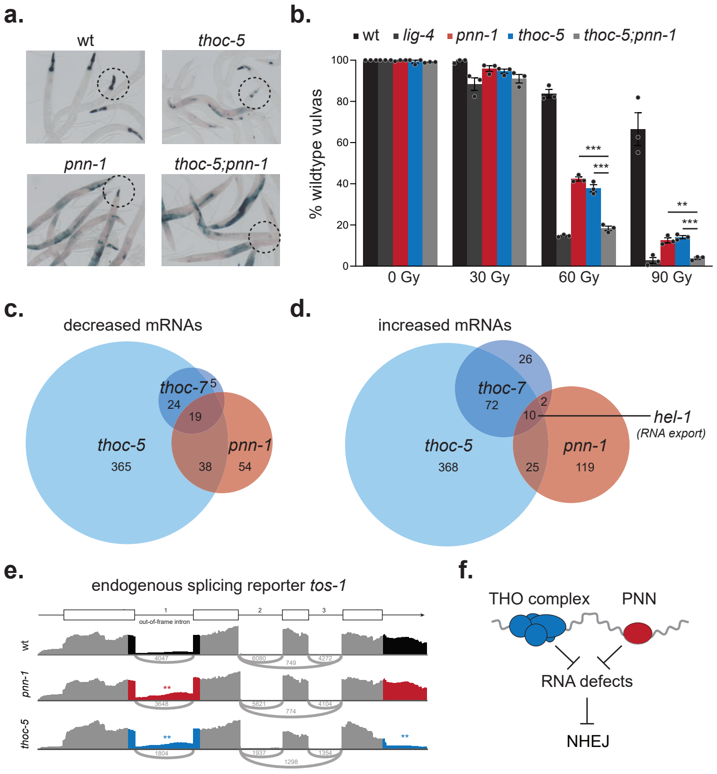

**Figure 4. THOC-5 and PNN-1 control RNA processing and NHEJ efficacy**

**a.** Representative pictures of LacZ staining patterns of dual reporter animals. Circles indicates representative pharynges; please note that *thoc-5;pnn-1* double mutants lack LacZ staining of pharynges while showing robust LacZ expression in other somatic tissues, indicative of a selective, highly penetrant NHEJ defect **b.** Larvae were challenged with indicated doses of IR and vulva development was scored. Values depict averages of three independent experiments and error bars represent S.E.M. Dots indicate the average of each experiment. Statistical analysis shows significant differences in vulva development (two-tailed t-tests *P < 0.05; **P < 0.01, ***P < 0.001). please note that *thoc-5;pnn-1* double mutants show radiosensitivity similar to *lig-4* null mutants **c.** Venn diagram of significantly ($q \leq 0.05$) downregulated transcripts identified using RNA sequencing. **d.** Venn diagram of significantly ($q \leq 0.05$) upregulated transcripts identified using RNA sequencing. One of the nineteen transcripts upregulated in all three mutants, *hel-1*, is highlighted. **e.** Examples of splicing errors in *thoc-5* and *pnn-1* mutants, enhancing expression of an out-of-frame intron. Gene model of *tos-1* and below simplified sashimi plots for each genotype. Sashimi plots depict relative RNAseq read-depth around the highlighted region of *tos-1*. Asterisks indicate significant alternative splicing events as detected by DEXSeq (Padjusted ≤ 0.05) **f.** Working model; The presence of the THO complex and PNN is necessary to prevent RNA defects that impede NHEJ.

## SMG-1 kinase impairs NHEJ in THO deficient animals

To identify the mechanism connecting RNA surveillance to NHEJ we performed an unbiased genetic suppressor screen (Figure 5A). We reasoned that if RNA stress is responsible for an altered DNA damage response, we could potentially restore normal NHEJ levels in THO mutants by mutating potential signalling components.

To this end, we mutagenized *thoc-5* mutant animals that carry the fluorescent reporter system and identified a suppressor mutant that restored NHEJ activity in *thoc-5* deficient animals. The identified suppressor mutation restored I-SceI-dependent GFP fluorescence and alleviated IR sensitivity in *thoc-5* deficient animals (Figure S14). Combining positional mapping and whole-genome sequencing revealed a candidate suppressor mutation in *smg-1,* a gene encoding a well-conserved kinase implicated in both NMD and DDR signalling(Gewandter, Bambara et al. 2011) (Figure 5B). The identified smg-1(lf238) allele causes a nonsense mutation and is predicted to express a severely truncated SMG-1 peptide lacking its well-conserved kinase and FACT domains (Figure S1). Notably, *smg-1* loss also fully rescues NHEJ activity in animals bearing a missense mutation in *thoc-*7, indicating that restored NHEJ activity does not rely on non-sense *thoc-5* mRNA, which is a possible NMD substrate (Figure 5C, 5D). Depleting *smg-1* using RNAi also restored NHEJ activity in *thoc-7* mutant animals and *smg-1* loss did not compromise NHEJ reporter specificity, confirming SMG-1 as a bona-fide NHEJ suppressor (Figure S14). Next-generation sequencing of DSB repair footprints revealed restored deletion formation in *thoc-5;smg-1* double mutants compared to *thoc-5* single mutants, including typical repair products of NHEJ (deletions without homology) or TMEJ (deletions with micro-homology and templated inserts) (Figure 5E). Independent phospho-proteome analysis also suggests a major role for SMG-1 in THO deficient animals, as *thoc-5* mutants show increased phosphorylation at more than 800 unique phosphorylation sites compared to wildtype controls, while their *smg-1* deficient counterparts have less than 300 unique phosphorylation events (Figure S15). Although this whole-animal proteomics approach is unable to identify direct SMG-1 targets (i.e. direct phospho-peptides are lost/undetectable in SMG-1 null mutants), it substantiates a key role for SMG-1 in altered signalling in THO mutants. SMG-1 phosphorylates numerous cellular proteins beyond those required for NMD. We thus wondered if the observed crosstalk between RNA surveillance and NHEJ activity requires SMG-1 specifically or NMD activity in general. Genetic dissection of the NMD pathway revealed that SMG-2, the major downstream target of SMG-1 also contributes to NHEJ suppression in THO deficient animals (Figure 5B, Figure S16). In contrast, loss of SMG-5, a critical NMD factor that controls the dephosphorylation of SMG-1 targets, did not restore NHEJ activity in THO deficient animals (Figure 5C-D and Figure S15). These observations have important implications for the mechanism of NHEJ interference: i) although NMD substrates are elevated in THO mutants, it is not the stabilisation of a particular NMD substrate that rescues the NHEJ defect, as the latter would also be the case in *smg-5* animals, and ii) SMG-1 impairs NHEJ via its established NMD target SMG-2, which much like SMG-1 has reported functions in both RNA

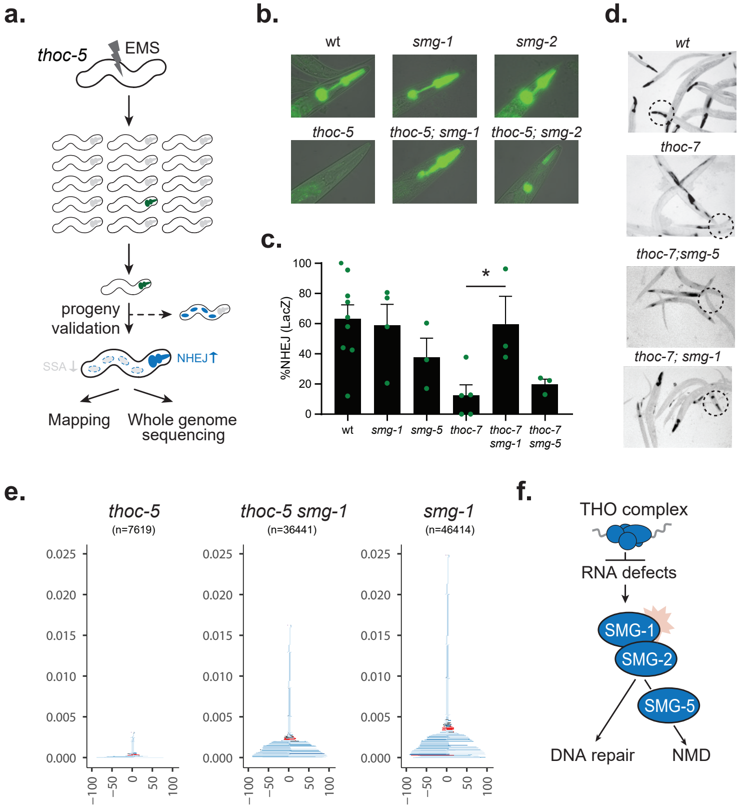

**Figure 5. Loss of SMG-1 or SMG-2, but not SMG-5, restores NHEJ in THO deficient animals**

**a.** NHEJ suppressor screen set-up. Genomes of *thoc-5* animals carrying the dual reporter system were mutagenized using ethyl methanesulfonate (EMS). NHEJ activity of the progeny was assessed and GFP-positive worms were isolated. Restored NHEJ activity was verified in *thoc-5* animals using LacZ staining patterns among clonal progeny. NHEJ-proficient candidates were followed up to identify the causal mutation. **b.** Representative pictures of pharyngeal GFP expression in NHEJ reporter animals. **c.** Quantification of GFP-positive pharynges. Average of three populations ($n>200$) is depicted. Dots indicate the average of each experiment. Error bars represent SEM (two-tailed t-tests *$P < 0.05$; **$P < 0.01$, ***$P < 0.001$). **d.** LacZ staining patterns; circles highlight representative pharynges (i.e. NHEJ proficiency). **e.** Tornado plots depict type and size of mutagenic repair events identified by NGS (n indicates number of mutagenic events). Colour code indicates type of event as in figure 3g. **f.** Working model; while RNA defects due to THO deficiency activate NMD by SMG-1, SMG-2 and SMG-5, only SMG-1 and SMG-2 have moonlighting functions in DDR and inhibit NHEJ.

surveillance and DNA integrity(Azzalin and Lingner 2006). THO deficiency thus causes a defect in DSB repair, which does not require NMD *per se*, but rather the specific activity of moonlighting NMD factors, i.e., the PIKK SMG-1 and the DNA/RNA helicase SMG-2 (Figure 5D).

# DISCUSSION

DNA and RNA fulfil very different functional niches within the cell, yet their chemical structures are highly similar. Due to their analogous structure, many proteins bind both DNA and RNA, which predicts a high level of crosstalk between DNA and RNA regulation. Indeed, DNA damage changes the activity of many RNA-binding factors and thus also dictates how a cell processes RNA(Botto, Muñoz et al. 2020). It is, however, unclear whether the reverse is also true. Can defects in RNA processing feed into the DDR and change the way a cell repairs DNA? The latter could have major consequence, as it would translate stochastic RNA expression changes into permanent scars in the genome. Here we describe how two well-conserved proteins, SMG-1 and SMG-2, can cause a short-circuit *in vivo* between RNA processing and DNA repair.

We created and used animal models to monitor DSB repair in somatic tissues and identified highly conserved RNA binding proteins that cause defective NHEJ (i.e. Pinin and the THO complex). Both THO and Pinin associate with spliceosomes and have been implicated in pre-mRNA processing and mRNA export(Li, Lin et al. 2003, Chi, Wang et al. 2013). Intriguingly, defective THO complex function is known to result in genome instability in various species, including yeast, worms and humans, which in part could be explained by its role in preventing the formation of RNA:DNA hybrids that hinder DNA replication throughout the genome(Dominguez-Sanchez, Barroso et al. 2011, Castellano-Pozo, Garcia-Muse et al. 2012). Our findings in terminally differentiated muscle cells demonstrate that either loss of PNN-1 or THO function can alter DSB repair in the absence of DNA replication. We find that PNN-1 and THO deficiency cause a strikingly similar change in RNA metabolism genome-wide, including altered transcript levels and splicing defects. Importantly, animals lacking PNN-1 or THO also shared multiple features of compromised NHEJ activity, including reduced NHEJ products, elevated SSA and IR hypersensitivity. Subsequent genetic screening revealed the signalling factor linking RNA processing to DNA repair: the moonlighting kinase SMG-1. SMG-1 was first identified to control RNA surveillance via NMD but later was found to also respond to DNA damage and control cell cycle arrest and p53 expression(Gehen, Staversky et al. 2008, Gewandter, Bambara et al. 2011, Gubanova, Issaeva et al. 2013). These latter functions in DNA surveillance are shared by the related kinases ATM and ATR, suggesting a possible ancestral role of PIKKs in genome stability. Interestingly, both ATM and ATR have recently been shown to be activated not only by broken DNA but also by R-loops(Marabitti, Lillo et al. 2019, Matos, Zhang et al. 2020), suggesting that intertwined DNA-RNA signalling might be a common theme among PIKKs. Our results indicate that pathological conditions causing RNA stress can derail DSB repair via hyperactivation of the PIKK SMG-1. Probable triggers for SMG-1 activity in THO mutants include elevated levels of non-sense RNAs and R-loops. In contrast to SMG-1 loss, loss of SMG-5 does not restore NHEJ in RNA processing mutants, indicating that the DNA repair defect is not caused by NMD activity *per se* and thus does not require degradation of a particular non-sense RNA. Instead we find that SMG-2, a well-established phosphorylation target of SMG-1, also impairs NHEJ in THO mutants. Interestingly, the human ortholog of SMG-2, UPF1, is a multifunctional DNA/RNA helicase with roles in NMD and DNA replication(Azzalin and Lingner 2006). Recent *in vitro* studies revealed that SMG-2/UPF1 binds both DNA and RNA, and can act as a highly progressive helicase on both substrates(Fiorini, Bagchi et al. 2015, Kanaan, Raj et al. 2018). Our results suggest that the dual roles of SMG-1 and SMG-2 in RNA and DNA surveillance can distort DNA repair and trigger genome instability, particularly in conditions that impair the processing and nuclear export of RNA (Figure 6).

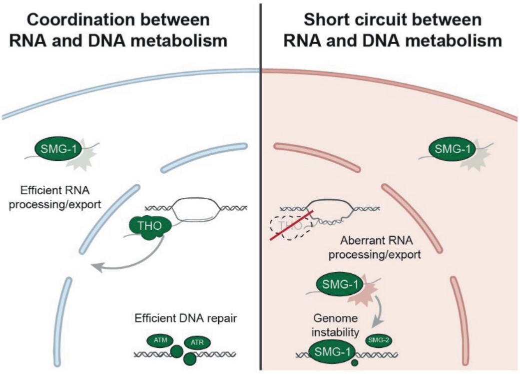

**Figure 6. Crosstalk between RNA defects and DNA repair**

In wildtype situations, the THO complex processes newly transcribed RNA, enabling efficient export and preventing RNA:DNA hybrids. SMG-1 functions as a NMD mediator to remove defective transcripts. In situations where efficient RNA processing is impaired, like in THO mutant cells, we hypothesize that defective transcripts accumulate and hyperactivate SMG-1. Deregulation of this moonlighting kinase alters the DNA damage response, which leads to a decrease in NHEJ efficiency. Thus, in conditions that cause aberrant RNA processing, protein moonlighting can create short-circuits between RNA and DNA maintenance.

Here we provide proof-of-concept that crosstalk exists between RNA surveillance and DNA repair and propose that stochastic defects in RNA processing can permanently shape the

DNA landscape via deregulation of SMG-1. Notably, genetic conditions causing elevated SMG-2/UPF1 phosphorylation have recently been linked to severe developmental retardation, microcephaly and variable facial and organ malformations, emphasizing the health risks associated with deregulated SMG-1 activity(Alzahrani, Kuwahara et al. 2020). Given that many RNA processing proteins have multiple functions, also in processes not related to NMD or DNA repair (e.g. PNN-1 controls splicing and cell adhesion), we predict that signalling crosstalk via moonlighting proteins is abundant *in vivo* and might have a major impact on basic physiology and pathology. Recent estimates indeed suggest that 78% of moonlighting proteins are involved in human diseases, which is significantly higher than the 18% disease association for proteins in general(Franco-Serrano, Huerta et al. 2018). We predict that the recent advances in unbiased genetic screening in both animal models and human cells will contribute to the discovery of hidden signalling connections and reveal additional vulnerabilities caused by moonlighting proteins.

# MATERIALS AND METHODS

## Genetics

All strains were cultured according to standard *C. elegans* procedures(Brenner 1974). Alleles used in this study include: *LGI: smg-1(lf238), smg-2(qd101), smg-5(r860) LGIII: cku-70(tm1524), cku-80(ok861), lig-4(ok716), brc-1(tm1145), cku-70(lf151), cku-80(lf152), cku-80(lf153), thoc-2(lf158), thoc-5/Y32H12A.2(lf161), LGIV: thoc-7/B0513.2(lf160), LGV: pnn-1/R186.7(lf159), pnn-1/R186.7(ok1872), LGX: pkIs2379 [Phsp-16.41::I-SceI-ORF; rol-6(su1006)], pkIs2170 [SSA reporter Phsp-16.41::ATG::LacZ::ISceI- site::stops::LacZ-ORF); unc-119(+)], IfIs104 [NHEJ reporter Pmyo-2::ATG::I-SceIsite: GFP-ORF::LacZ-ORF, Phsp-16.41::mCherry::I-SceI-ORF; rol-6(su1006)], lfEx164 [thoc-2 fosmid WRM0614bD12; Pmyo-3::mCherry; Prab-3::mCherry], lfEx166 [thoc-7 fosmid WRM0640bD11; Pmyo-3::mCherry; Prab-3::mCherry], lfEx164 [thoc-5 fosmid WRM0617bE04; Pmyo-3::mCherry; Prab-3::mCherry], lfEx190 [lig-4 fosmid WRM0634bF07; Pmyo-3::mCherry; Prab-3::mCherry], lfEx195 [Prpl-28::lig-4-cDNA; Pmyo-3::mCherry; Prab- 3::mCherry], lfEx196 [Prpl-28::lig-4-ORF; Pmyo-3::mCherry; Prab-3::mCherry].* All transgenic strains were obtained by microinjection of plasmid/fosmid DNA into the germ line and data presented are from a single representative transgenic line unless noted otherwise. The parental NHEJ reporter transgene *IfIs104* was obtained via IR-mediated genomic integration and combined with *pkIs2379* and *pkIs2170* to create the dual reporter strain XF540, which served as the starting strain for the forward genetics screen.

## DSB repair reporter assays

Synchronized L1 animals were obtained by harvesting eggs from hypochlorite-treated gravid adults and overnight starvation in M9 solution(Lewis and Fleming 1995). Hatched L1 larvae were transferred on NGM plates seeded with either *E. coli* OP50 or HT115 bacteria(Kamath, Fraser et al. 2003). In order to insure complete RNAi before DSB induction, L1 worms were cultured at 20°C for at least 20hrs. Heat-shock driven I-SceI expression was induced by putting the worms at 34°C for 60-180 minutes, as indicated. After the heat-shock procedure, worms were cultured at 20°C to allow DSB formation, DSB repair and worm development. NHEJ activity was measured by scoring pharyngeal GFP expression using a Leica M165FC fluorescence dissecting-microscope. Experiments were performed in triplicate with 50- 200 animals tested for each condition. After GFP quantification, ~25 adult animals were transferring onto microscope slides with 3% agarose pads and representative pictures were acquired using a Leica DM6000 microscope with 10X objective. SSA activity was measured by scoring animals showing LacZ positive cells in non-pharyngeal somatic tissues(Pontier and Tijsterman 2009). One hour prior fixation/LacZ staining, young adults were heat-shocked at 34°C for 120 minutes to induce SSA reporter expression.

## Forward genetics screen

Dual reporter larvae (XF540) were mutagenized with ethyl methanesulfonate (EMS) using standard procedures(Brenner 1974). Complex F2 populations, each derived from 50 mutagenized P0s, were bleached and synchronized L1 larvae (F3) were seeded on NGM/OP50 plates. On two consecutive days larvae were heat-shocked at 34°C for 180 minutes in order to maximize GFP ORF correction. GFPlow F3 animals were selected using a Leica M165FC fluorescence dissecting-microscope and clonal F4 populations were tested again for NHEJ activity. Populations showing reduced GFP expression were fixed and stained with X-gal as described previously(Pontier and Tijsterman 2009).

## Suppressor screen

Dual reporter larvae defective for *thoc-5* (XF677) were mutagenized with ethyl methanesulfonate (EMS) using standard procedures(Brenner 1974). Complex F2 populations, each derived from 10 mutagenized P0s, were bleached and synchronized L1 larvae (F3) were seeded on NGM/OP50 plates. Larvae were heat-shocked for 140 minutes, which was determined to be the optimal heat exposure to distinguish all NHEJ proficient worms spiked into a NHEJ deficient population. GFPhigh F3 animals were selected using a Leica M165FC fluorescence dissecting-microscope and clonal F4 populations were tested again for NHEJ activity. Populations showing GFP expression were fixed and stained with X-gal as described previously(Pontier and Tijsterman 2009).

## Positional cloning, genome-wide sequencing and transgenesis

Causal mutations in *thoc-2*, *thoc-5*, *thoc-7* and *smg-1* were mapped by crossing the respective mutants (Bristol) to the related Hawaiian strain CB4856 and performing single-nucleotidepolymorphism mapping on NHEJ proficient versus NHEJ deficient F2 lines(Davis, Hammarlund et al. 2005). Unique EMS-induced genetic alterations in the mapped regions were identified by comparing genome-wide paired-end sequencing data of the parental mutant strains using the Illumina Hiseq 2000 platform, the *C. elegans* reference genome (Wormbase version 225) and MaqGene software(Bigelow, Doitsidou et al. 2009). Causality was established by complementation analysis using wild-type fosmid arrays. Complemented regions spanned by the fosmids contained only one non-synonymous SNP (Figure S17). To create transgenic animals carrying fosmid arrays an injection mix containing 100 ng/µl pBluescript, 10 ng/µl pGH8 (*Prab-3::mCherry::unc-54-3'UTR), 5 ng/µl pCFJ104 (Pmyo-3::mCherry::unc-54-3'UTR) and 10-50 ng/µl fosmid DNA (lig-4 WRM0634bF07, thoc-2 WRM0614bD12, thoc-7 WRM0640bD11, thoc-5 WRM0617bE04,pnn-1 WRM0637aA06*) was injected into the gonads of young adults. For *lig-4* cDNA and *lig-4* ORF containing vectors, 5 ng/µl plasmid DNA was added instead of the fosmid DNA.

## IR sensitivity assays

All IR experiments were performed with a dose rate of 10-15 Gy/minute using an electronic X-ray generator (XYLON International). Figures provide mean values of three independent experiments. L4 assay: three L4 animals per plate were treated with various doses of IR (three plates per condition) and were removed after an egg-laying period of two days. Progeny survival was scored 1 day after removal of the irradiated p0. L1 assay: ~200 L1 larvae per plate were treated with various doses of IR (three plates per condition) and vulva phenotypes were scored 5 days post IR. Representative pictures of irradiated populations were acquired using a Leica DFC295 camera/M165FC microscope.

## Transcriptome sequencing

To obtain clean L1 populations and remove dead corpses, o/n starved L1 progeny from hypochlorite-treated gravid adults were filtered using 10µm nylon filters (Millipore). Total

RNA of >3000 L1s per sample was extracted as follows: L1 animas were collected in 100µl M9, 400µl Trizol was added, vortexed 2 minutes, followed by 4 snap freeze/thaw cycles at -196°C/37°C, 200µl Trizol was added, incubated 5 minute RT, 120µl chloroform was added, incubated 2 minute at RT, centrifuged 15 minutes at 4°C (16000 rcf), 350µl supernatant was mixed with 70% ethanol (1:1) and total RNA was purified using Purelink® RNA columns (Ambion) and stored at -80°C. Total RNA was DNAse treated (Turbo DNA-free, Ambion) and RNA quality was verified using RNA 6000 Pico kit (Agilent). RNA-Seq was performed on an Illumina HiSeq 2000 platform using standard reagents. Raw reads were aligned by TopHat on Wormbase assembly 238, which allows for reads to be split over splice junctions. Next, we used DESeq to identify differentially expressed transcripts ($q \leq 0.05$)(Trapnell, Roberts et al. 2012). For differential exon expression, DEXSeq was used using standard settings (padjust ≤ 0.05)(Anders, Reyes et al. 2012). For each mutant two or three samples were sequenced to account for variation between isogenic L1 populations. **r**MATS was used to determine alternatively spliced events. Since rMATS is unable to intron retention in introns that are not part of annotated alternative transcripts we also analyzed the data using iReads(Shen, Park et al. 2014, Li, Funk et al. 2020), which is specifically designed to find retained introns in introns that do not overlap annotated exons.

## PCR-based DSB repair assay

Synchronized L1 larvae were seeded on NGM/OP50 plates and heat-shocked at 34°C for 180 Minutes. Heat-shock was repeated 24 hours after seeding. Expression of mCherry-tagged ISceI was verified using a Leica M165FC fluorescence dissecting-microscope. Genomic DNA was extracted 24 hours after the second heat shock using a DNeasy Blood & Tissue kit (Qiagen). The first PCR on 18,5 ng isolated DNA was performed using Phusion High Fidelity Polymerase (ThermoFisher). Primers (forward: 5' AAAGTTATCTCCAGGCTCGC, reverse: 5' TTCACCCTCTCCACTGACAG) containing 5' adaptor sequences (forward: 5' GATGTGTATAAGAGACAG, reverse: 5' CGTGTGCTCTTCCGATCT) flanking the I-SceI target site of the NHEJ reporter were used. PCR products were purified using Ampure XP beads(Beckman Coulter). A second PCR of 5 cycles was performed to barcode the products using p5 and p7 index primers. Final PCR products were purified using Ampure XP beads (Beckman Coulter) and sequenced on an Illumina NovaSeq 6000. Reads were filtered and aligned to a reference sequence using a custom JAVA program (available upon request). Only reads of which both the start and end of the sequence lied within the specified primers were used. The custom JAVA program was used to assign the reads to different classifications: wildtype, SNVs, deletion, insertion, delins (deletion with insertion), tandem duplication(TD) or compound TD (TD with an additional mutation). Reads containing variation in the number of adenines in a polyA-tract were filtered and reads containing consequences of mutagenic repair (deletions, insertions, delins, TDs and compound TDs) were plotted using R.

## Phosphoproteome analysis

Worm populations from twenty plates per genotype were rinsed off and washed with M9 solution(Lewis and Fleming 1995). Lysis buffer (6M GdmCl (Guanidine hydrochloride), 100 mM Tris pH 8.5, 10 mM TCEP (Tris(2-carboxyethyl)phosphine) , 40mM CAA (2-Chloroacetamide) and 4µl benzonase (250U/µl, Novagen) was added after the populations were washed with autoclaved demi water. Worms were snap-frozen in liquid nitrogen and stored at -80 degrees Celsius. After thawing, worms were sonicated four times for 30 seconds in an ice-filled water bath. Lysis of worms was verified using a dissection microscope. Lysates were snap-frozen in liquid nitrogen. Lysates were thawed and protein concentrations were measured using a BCA kit (Thermo Scientific). Proteins were digested using Trypsin/Lys-C mix (Promega) and phosphopeptides were enriched, cleaned and eluted as described previously(Sampadi, Pines et al. 2020). Samples were analyzed using a Q-Exactive HF mass spectrometer. Phosphosites identified in all samples were included for statistical analysis and compared between genotypes using two-sided t-tests. Significance of global phosphorylation changes was determined by permutation of the set.

## RNA:DNA hybrid staining

Genomic DNA was isolated from worm populations grown on a NGM plate using a Blood and Tissue Culture Kit (Qiagen), without the RNAse A treatment step. Concentrations of DNA were measured using Qubit. Equal amounts of DNA per sample were treated with RNAse H (NEB) or mock-treated at 37 degrees Celsius for 1 hour. Samples were diluted in TE buffer with 6x SSC and blotted on a nitrocellulose membrane. The membrane is washed with TE0.1 and baked at 65 degrees Celsius for 1 hour. The membrane is blocked in 5% milk in PBST overnight. Staining is performed with 1:3000 s9.6 antibody (Kerafast) in 3% milk 1% BSA in PBST for 2 hours at room temperature. After washing (once 3% milk 1% BSA in PBST, twice with 1% BSA in PBST), the membrane was stained with anti-mouse$^{HRP}$ for 1 hour at room temperature, washed (thrice 1% BSA in PBST, twice with PBST and once with PBS) and visualized using enhanced chemiluminescence. Dot intensities were quantified using ImageJ.

### Fecundity assessment

L4 larvae (p0) were isolated on seeded NGM plates and transferred every 24 hours until no eggs were laid (day 5). Larvae and dead eggs on NGM plates were counted 24 hours after transfer of the p0 worm. Brood size was calculated as the sum of larvae and eggs on all plates per p0. Survival was calculated by dividing the number of alive worms by the total brood size.

## DATA AVAILABILITY

Raw sequences have been made publicly available at NCBI SRA (accession code PRJNA678136 and PRJNA674486). Sequencing data for N2 wild-type, *brc-1* and *lig-4* animals were published previously and can be found at NCBI SRA (accession codes PRJNA260487 and PRJNA599297. The phosphoproteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD022616.

## ACKNOWLEDGEMENTS

The authors thank Shohei Mitani (National Bioresource Project, Japan) and the Caenorhabditis Genetics Center for strains; Jennemiek van Arkel and Evelina Papaioannou for technical support; Bharath Sampadi for sharing his phosphoproteomics protocol and his help obtaining the phosphopeptides; Xanthe Matthijssen and Jelle Goeman for their help with statistical analysis. This study has been supported by the European Research Council (203379, DSBrepair), the European Commission (DDResponse) and ZonMW/NGI-Horizon (to M.T.) and the Swedish Research Council and the Swedish Cancer Society (to B.L.).

## CONTRIBUTIONS

B.L. and M.T. conceived and designed the study. J.K., B.L. and J.v.O. performed the experiments. R.v.S. and R.G.P. performed the bioinformatics analysis. All authors interpreted the experimental data. B.L., J.K. and M.T. wrote the manuscript.

## COMPETING INTERESTS

The authors declare no competing financial interests.

# REFERENCES

Alzahrani, F., H. Kuwahara, Y. Long, M. Al-Owain, M. Tohary, M. AlSayed, M. Mahnashi, L. Fathi, M. Alnemer, M. H. Al-Hamed, G. Lemire, K. M. Boycott, M. Hashem, W. Han, A. Al-Maawali, F. Al Mahrizi, K. Al-Thihli, X. Gao and F. S. Alkuraya (2020). "Recessive, Deleterious Variants in SMG8 Expand the Role of Nonsense-Mediated Decay in Developmental Disorders in Humans." Am J Hum Genet **107**(6): 1178-1185.

Anders, S., A. Reyes and W. Huber (2012). "Detecting differential usage of exons from RNA-seq data." Genome Res **22**(10): 2008-2017.

Atkinson, J. and R. Martin (1994). "Mutations to nonsense codons in human genetic disease: implications for gene therapy by nonsense suppressor tRNAs." Nucleic Acids Res **22**(8): 1327-1334.

Azzalin, C. M. and J. Lingner (2006). "The double life of UPF1 in RNA and DNA stability pathways." Cell Cycle **5**(14): 1496-1498.

Balmus, G., D. Pilger, J. Coates, M. Demir, M. Sczaniecka-Clift, A. C. Barros, M. Woods, B. Fu, F. Yang, E. Chen, M. Ostermaier, T. Stankovic, H. Ponstingl, M. Herzog, K. Yusa, F. M. Martinez, S. T. Durant, Y. Galanty, P. Beli, D. J. Adams, A. Bradley, E. Metzakopian, J. V. Forment and S. P. Jackson (2019). "ATM orchestrates the DNA-damage response to counter toxic non-homologous end-joining at broken replication forks." Nat Commun **10**(1): 87.

Bigelow, H., M. Doitsidou, S. Sarin and O. Hobert (2009). "MAQGene: software to facilitate C. elegans mutant genome sequence analysis." Nat Methods **6**(8): 549.

Botto, A. E. C., J. C. Muñoz, L. E. Giono, N. Nieto-Moreno, C. Cuenca, A. R. Kornblihtt and M. J. Muñoz (2020). "Reciprocal regulation between alternative splicing and the DNA damage response." Genet Mol Biol **43**(1 suppl. 1): e20190111.

Brenner, S. (1974). "The genetics of Caenorhabditis elegans." Genetics **77**(1): 71-94.

Brumbaugh, K. M., D. M. Otterness, C. Geisen, V. Oliveira, J. Brognard, X. Li, F. Lejeune, R. S. Tibbetts, L. E. Maquat and R. T. Abraham (2004). "The mRNA surveillance protein hSMG-1 functions in genotoxic stress response pathways in mammalian cells." Mol Cell **14**(5): 585-598.

Castellano-Pozo, M., T. Garcia-Muse and A. Aguilera (2012). "R-loops cause replication impairment and genome instability during meiosis." EMBO Rep **13**(10): 923-929.

Castellano-Pozo, M., T. García-Muse and A. Aguilera (2012). "The Caenorhabditis elegans THO complex is required for the mitotic cell cycle and development." PLoS One **7**(12): e52447.

Chi, B., Q. Wang, G. Wu, M. Tan, L. Wang, M. Shi, X. Chang and H. Cheng (2013). "Aly and THO are required for assembly of the human TREX complex and association of TREX components with the spliced mRNA." Nucleic Acids Res **41**(2): 1294-1306.

Clejan, I., J. Boerckel and S. Ahmed (2006). "Developmental modulation of nonhomologous end joining in Caenorhabditis elegans." Genetics **173**(3): 1301-1317.

Craig, A. L., S. C. Moser, A. P. Bailly and A. Gartner (2012). "Methods for studying the DNA damage response in the Caenorhabdatis elegans germ line." Methods Cell Biol **107**: 321-352.

Davis, M. W., M. Hammarlund, T. Harrach, P. Hullett, S. Olsen and E. M. Jorgensen (2005). "Rapid single nucleotide polymorphism mapping in C. elegans." BMC Genomics **6**: 118.

Dominguez-Sanchez, M. S., S. Barroso, B. Gomez-Gonzalez, R. Luna and A. Aguilera (2011). "Genome instability and transcription elongation impairment in human cells depleted of THO/TREX." PLoS Genet **7**(12): e1002386.

Fiorini, F., D. Bagchi, H. Le Hir and V. Croquette (2015). "Human Upf1 is a highly processive RNA helicase and translocase with RNP remodelling activities." Nat Commun **6**: 7581.

Franco-Serrano, L., M. Huerta, S. Hernández, J. Cedano, J. Perez-Pons, J. Piñol, A. Mozo-Villarias, I. Amela and E. Querol (2018). "Multifunctional Proteins: Involvement in Human Diseases and Targets of Current Drugs." Protein J **37**(5): 444-453.

Gehen, S. C., R. J. Staversky, R. A. Bambara, P. C. Keng and M. A. O'Reilly (2008). "hSMG-1 and ATM sequentially and independently regulate the G1 checkpoint during oxidative stress." Oncogene **27**(29): 4065-4074.

Gewandter, J. S., R. A. Bambara and M. A. O'Reilly (2011). "The RNA surveillance protein SMG1 activates p53 in response to DNA double-strand breaks but not exogenously oxidized mRNA." Cell Cycle **10**(15): 2561-2567.

González-Huici, V., B. Wang and A. Gartner (2017). "A Role for the Nonsense-Mediated mRNA Decay Pathway in Maintaining Genome Stability in Caenorhabditis elegans." Genetics **206**(4): 1853-1864.

Gubanova, E., N. Issaeva, C. Gokturk, T. Djureinovic and T. Helleday (2013). "SMG-1 suppresses CDK2 and tumor growth by regulating both the p53 and Cdc25A signaling pathways." Cell Cycle **12**(24): 3770-3780.

Guria, A., D. D. Tran, S. Ramachandran, A. Koch, O. El Bounkari, P. Dutta, H. Hauser and T. Tamura (2011). "Identification of mRNAs that are spliced but not exported to the cytoplasm in the absence of THOC5 in mouse embryo fibroblasts." Rna **17**(6): 1048-1056.

Hanscom, T. and M. McVey (2020). "Regulation of Error-Prone DNA Double-Strand Break Repair and Its Impact on Genome Evolution." Cells **9**(7).

Herold, A., L. Teixeira and E. Izaurralde (2003). "Genome-wide analysis of nuclear mRNA export pathways in Drosophila." Embo j **22**(10): 2472-2483.

Ho, U., J. Luff, A. James, C. S. Lee, H. Quek, H. C. Lai, S. Apte, Y. C. Lim, M. F. Lavin and T. L. Roberts (2019). "SMG1 heterozygosity exacerbates haematopoietic cancer development in Atm null mice by increasing persistent DNA damage and oxidative stress." J Cell Mol Med **23**(12): 8151-8160.

Holway, A. H., S. H. Kim, A. La Volpe and W. M. Michael (2006). "Checkpoint silencing during the DNA damage response in Caenorhabditis elegans embryos." J Cell Biol **172**(7): 999-1008.

Johnson, N. M., B. B. Lemmens and M. Tijsterman (2013). "A role for the malignant brain tumour (MBT) domain protein LIN-61 in DNA double-strand break repair by homologous recombination." PLoS Genet **9**(3): e1003339.

Kamath, R. S., A. G. Fraser, Y. Dong, G. Poulin, R. Durbin, M. Gotta, A. Kanapin, N. Le Bot, S. Moreno, M. Sohrmann, D. P. Welchman, P. Zipperlen and J. Ahringer (2003). "Systematic functional analysis of the Caenorhabditis elegans genome using RNAi." Nature **421**(6920): 231-237.

Kamp, J. A., R. van Schendel, I. W. Dilweg and M. Tijsterman (2020). "BRCA1-associated structural variations are a consequence of polymerase theta-mediated end-joining." Nat Commun **11**(1): 3615.

Kanaan, J., S. Raj, L. Decourty, C. Saveanu, V. Croquette and H. Le Hir (2018). "UPF1-like helicase grip on nucleic acids dictates processivity." Nat Commun **9**(1): 3752.

Lanz, M. C., D. Dibitetto and M. B. Smolka (2019). "DNA damage kinase signaling: checkpoint and repair at 30 years." Embo j **38**(18): e101801.

Lemmens, B. B., N. M. Johnson and M. Tijsterman (2013). "COM-1 promotes homologous recombination during Caenorhabditis elegans meiosis by antagonizing Ku-mediated non-homologous end joining." PLoS Genet **9**(2): e1003276.

Lemmens, B. B. and M. Tijsterman (2011). "DNA double-strand break repair in Caenorhabditis elegans." Chromosoma **120**(1): 1-21.

Lewis, J. A. and J. T. Fleming (1995). "Basic culture methods." Methods Cell Biol **48**: 3-29.

Li, C., R. I. Lin, M. C. Lai, P. Ouyang and W. Y. Tarn (2003). "Nuclear Pnn/DRS protein binds to spliced mRNPs and participates in mRNA processing and export via interaction with RNPS1." Mol Cell Biol **23**(20): 7363-7376.

Li, H. D., C. C. Funk and N. D. Price (2020). "iREAD: a tool for intron retention detection from RNA-seq data." BMC Genomics **21**(1): 128.

Lloyd, J. P. B., D. Lang, A. D. Zimmer, B. Causier, R. Reski and B. Davies (2018). "The loss of SMG1 causes defects in quality control pathways in Physcomitrella patens." Nucleic Acids Res **46**(11): 5822-5836.

Luna, R., A. G. Rondón and A. Aguilera (2012). "New clues to understand the role of THO and other functionally related factors in mRNP biogenesis." Biochim Biophys Acta **1819**(6): 514-520.

Ma, L., X. Gao, J. Luo, L. Huang, Y. Teng and H. R. Horvitz (2012). "The Caenorhabditis elegans gene mfap-1 encodes a nuclear protein that affects alternative splicing." PLoS Genet **8**(7): e1002827.

Maeder, C. I., J. I. Kim, X. Liang, K. Kaganovsky, A. Shen, Q. Li, Z. Li, S. Wang, X. Z. S. Xu, J. B. Li, Y. K. Xiang, J. B. Ding and K. Shen (2018). "The THO Complex Coordinates Transcripts for Synapse Development and Dopamine Neuron Survival." Cell **174**(6): 1436-1449 e1420.

Marabitti, V., G. Lillo, E. Malacaria, V. Palermo, M. Sanchez, P. Pichierri and A. Franchitto (2019). "ATM pathway activation limits R-loop-associated genomic instability in Werner syndrome cells." Nucleic Acids Res **47**(7): 3485-3502.

Matos, D. A., J. M. Zhang, J. Ouyang, H. D. Nguyen, M. M. Genois and L. Zou (2020). "ATR Protects the Genome against R Loops through a MUS81-Triggered Feedback Loop." Mol Cell **77**(3): 514-527.e514.

Matsuoka, S., B. A. Ballif, A. Smogorzewska, E. R. McDonald, 3rd, K. E. Hurov, J. Luo, C. E. Bakalarski, Z. Zhao, N. Solimini, Y. Lerenthal, Y. Shiloh, S. P. Gygi and S. J. Elledge (2007). "ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage." Science **316**(5828): 1160-1166.

Mladenov, E., X. Fan, R. Dueva, A. Soni and G. Iliakis (2019). "Radiation-dose-dependent functional synergisms between ATM, ATR and DNA-PKcs in checkpoint control and resection in G(2)-phase." Sci Rep **9**(1): 8255.

Neu-Yilik, G., N. H. Gehring, M. W. Hentze and A. E. Kulozik (2004). "Nonsense-mediated mRNA decay: from vacuum cleaner to Swiss army knife." Genome Biol **5**(4): 218.

Pérez-Calero, C., A. Bayona-Feliu, X. Xue, S. I. Barroso, S. Muñoz, V. M. González-Basallote, P. Sung and A. Aguilera (2020). "UAP56/DDX39B is a major cotranscriptional RNA-DNA helicase that unwinds harmful R loops genome-wide." Genes Dev **34**(13-14): 898-912.

Pontier, D. B. and M. Tijsterman (2009). "A robust network of double-strand break repair pathways governs genome integrity during C. elegans development." Curr Biol **19**(16): 1384-1388.

Robert, V. J., M. W. Davis, E. M. Jorgensen and J. L. Bessereau (2008). "Gene conversion and end-joining-repair double-strand breaks in the Caenorhabditis elegans germline." Genetics **180**(1): 673-679.

Roerink, S. F., R. van Schendel and M. Tijsterman (2014). "Polymerase theta-mediated end joining of replication-associated DNA breaks in C. elegans." Genome Res **24**(6): 954-962.

Salas-Armenteros, I., C. Pérez-Calero, A. Bayona-Feliu, E. Tumini, R. Luna and A. Aguilera (2017). "Human THO-Sin3A interaction reveals new mechanisms to prevent R-loops that cause genome instability." Embo j **36**(23): 3532-3547.

Sampadi, B., A. Pines, S. Munk, B. Mišovic, A. J. de Groot, B. van de Water, J. V. Olsen, L. H. F. Mullenders and H. Vrieling (2020). "Quantitative phosphoproteomics to unravel the cellular response to chemical stressors with different modes of action." Arch Toxicol **94**(5): 1655-1671.

Saran, S., D. D. H. Tran, S. Klebba-Färber, P. Moran-Losada, L. Wiehlmann, A. Koch, H. Chopra, O. Pabst, A. Hoffmann, R. Klopfleisch and T. Tamura (2013). "THOC5, a member of the mRNA export complex, contributes to processing of a subset of wingless/integrated (Wnt) target mRNAs and integrity of the gut epithelial barrier." BMC Cell Biology **14**(1): 51.

Schimmel, J., R. van Schendel, J. T. den Dunnen and M. Tijsterman (2019). "Templated Insertions: A Smoking Gun for Polymerase Theta-Mediated End Joining." Trends Genet **35**(9): 632-644.

Shen, H. (2009). "UAP56- a key player with surprisingly diverse roles in pre-mRNA splicing and nuclear export." BMB Rep **42**(4): 185-188.

Shen, S., J. W. Park, Z.-x. Lu, L. Lin, M. D. Henry, Y. N. Wu, Q. Zhou and Y. Xing (2014). "rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data." **111**(51): E5593-E5601.

Tran, D. D., S. Saran, O. Dittrich-Breiholz, A. J. Williamson, S. Klebba-Färber, A. Koch, M. Kracht, A. D. Whetton and T. Tamura (2013). "Transcriptional regulation of immediate-early gene response by THOC5, a member of mRNA export complex, contributes to the M-CSF-induced macrophage differentiation." Cell Death & Disease **4**(10): e879-e879.

Tran, D. D. H., S. Saran, A. J. K. Williamson, A. Pierce, O. Dittrich-Breiholz, L. Wiehlmann, A. Koch, A. D. Whetton and T. Tamura (2014). "THOC5 controls 3'end-processing of immediate early genes via interaction with polyadenylation specific factor 100 (CPSF100)." Nucleic Acids Research **42**(19): 12249-12260.

Trapnell, C., A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn and L. Pachter (2012). "Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks." Nat Protoc **7**(3): 562-578.

van Schendel, R., S. F. Roerink, V. Portegijs, S. van den Heuvel and M. Tijsterman (2015). "Polymerase Θ is a key driver of genome evolution and of CRISPR/Cas9-mediated mutagenesis." Nat Commun **6**: 7394.

Wang, L., Y. L. Miao, X. Zheng, B. Lackford, B. Zhou, L. Han, C. Yao, J. M. Ward, A. Burkholder, I. Lipchina, D. C. Fargo, K. Hochedlinger, Y. Shi, C. J. Williams and G. Hu (2013). "The THO complex regulates pluripotency gene mRNA export and controls embryonic stem cell self-renewal and somatic cell reprogramming." Cell Stem Cell **13**(6): 676-690.

Zhou, Y., J. H. Lee, W. Jiang, J. L. Crowe, S. Zha and T. T. Paull (2017). "Regulation of the DNA Damage Response by DNA-PKcs Inhibitory Phosphorylation of ATM." Mol Cell **65**(1): 91-104.

# SUPPLEMENTARY INFORMATION

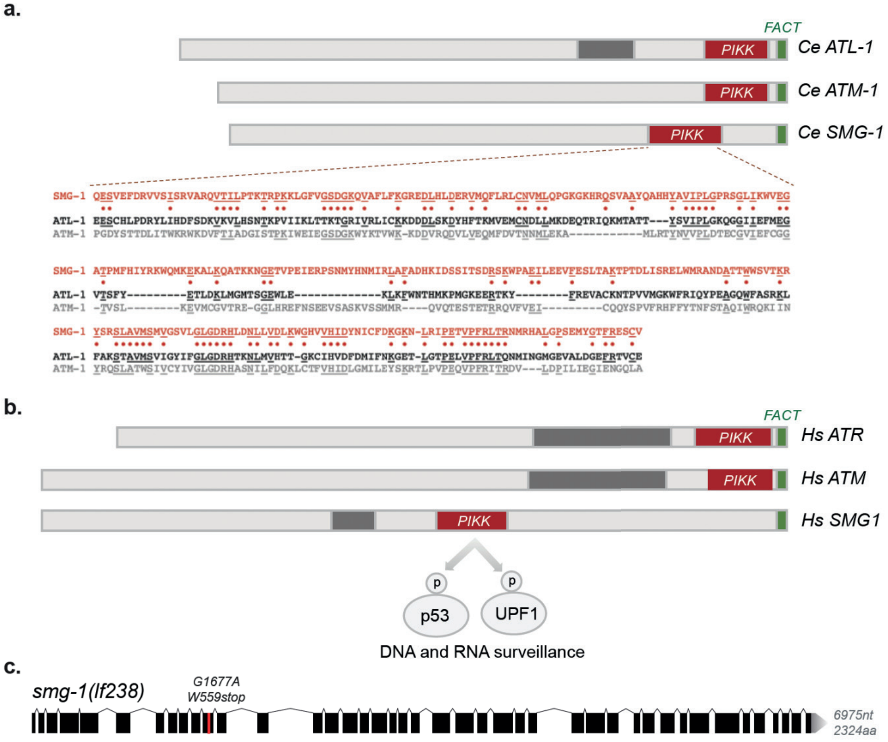

**Figure S1. SMG-1 shares conserved domains and phosphor-targets with DDR kinases**

**a.** Schematic diagram of protein domain structures of *C. elegans* PIKK-family kinases ATL-1, ATM-1 and SMG-1. Relative position of FACT domains (green), PIKK catalytic domains (red) and FAT domains (dark grey) are indicated. Lower panel shows sequence alignment of PIKK domains of the indicated proteins. Dots highlight homologous residues between SMG-1 (red) and ATL-1 (black) or ATM-1 (grey). **b.** Schematic diagram of protein domain structures of human PIKK-family kinases ATR, ATM and SMG1. Relative position of FACT domains (green), PIKK catalytic domains (red) and FAT domains (dark grey) are indicated. Arrows indicate two established SMG1 phosphorylation targets involved in the DNA damage response (p53) or NMD (UPF1). **c.** Gene model and newly identified non-sense allele of *smg-1*.

**Figure S2. Identified *cku-70* and *cku-80* alleles verify in vivo NHEJ screen**

**a.** Schematic diagram of conserved role of CKU-80/CKU-70 heterodimer in DSB recognition and repair via NHEJ. **b.** Gene models and newly identified alleles of *cku-70* and *cku-80*. **c.** LacZ expression patterns of *cku-80* deficient dual reporter animals. Synchronized animals were heat-shocked for 120 minutes at L1 stage to introduce I-SceI-induced DSBs and NHEJ/SSA reporter expression is detected in adults.

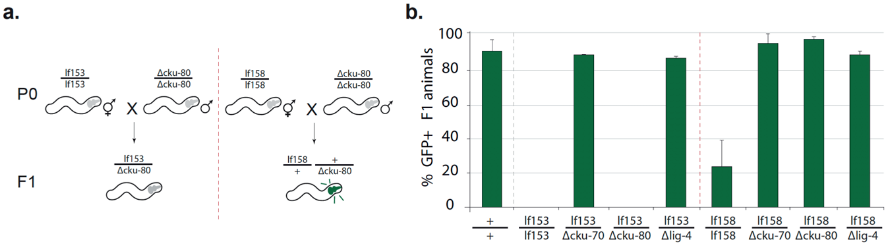

**Figure S3. Example complementation analysis revealing non-canonical NHEJ mutants**

**a.**Example of crossing scheme used for complementation analysis. Established and uncharacterized NHEJ mutants are crossed and trans-heterozygous F1 progeny is analysed for somatic NHEJ activity. **b.** Quantification of GFP-positive pharynges in trans-heterozygous F1 cross progeny, heat-shocked for 120 minutes to introduce I-SceI-induced DSBs and measured in adults. Average percentage of GFP-positive pharynges of three independent F1 populations is depicted.

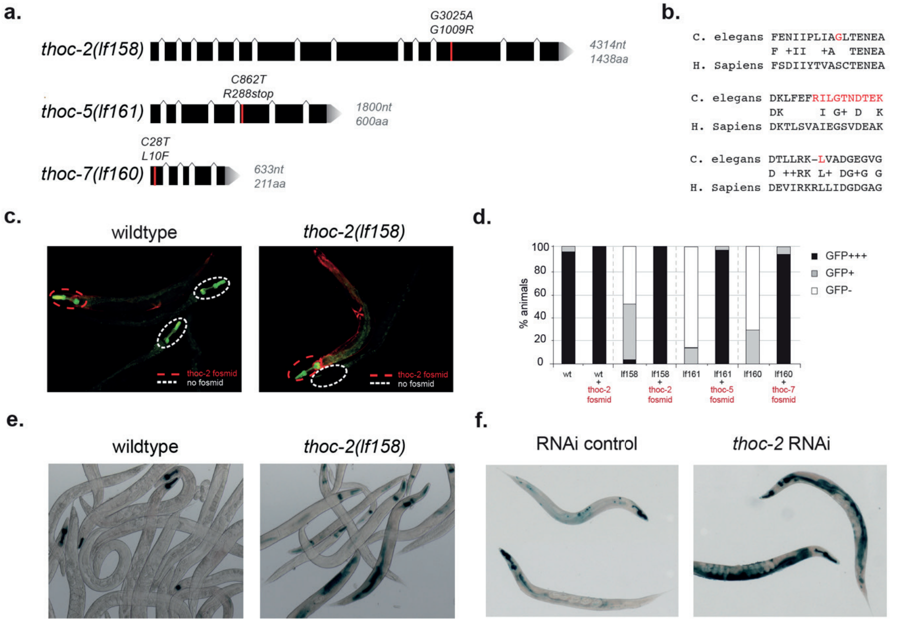

**Figure S4. Identified *thoc-2*, *thoc-5* and *thoc-7* alleles and causal effect validation**

**a.** Gene models and newly identified alleles of *thoc-2*, *thoc-5* and *thoc-7*. Right panel depicts amino acid context and conservation of THO mutations. **b.** Complementation analysis using fosmid arrays carrying wild-type THO genes and mCherry expression markers. Representative picture of synchronized populations of wildtype and lf158 animals with (red circle) or without (white circle) *thoc-2* fosmid array; animals were heat-shocked for 180 minutes at L1 stage. **c.** Histogram shows quantification of GFP-positive pharynges in adults of the different genetic backgrounds. NHEJ activity was restored to wild-type levels by complementing the THO mutants with functional THO genes. Average percentage of GFP-positive pharynges of three independent populations (n>150) is depicted. **d.** LacZ expression patterns of *thoc-2* deficient dual reporter animals and wildtype parental controls. Animals were heat-shocked for 120 minutes at L1 stage to introduce I-SceI-induced DSBs and NHEJ/SSA reporter expression is detected in adults. **e.** Representative pictures of LacZ staining patterns in synchronized populations of dual reporter animals fed bacteria carrying empty vectors of thoc-2 RNAi vectors. L4 animals were heat-shocked for 120 minutes to introduce I-SceI-induced DSBs and reporter expression is detected the next day. Depletion of *thoc-2* caused a three-fold increase in LacZ positive bodies (SSA).

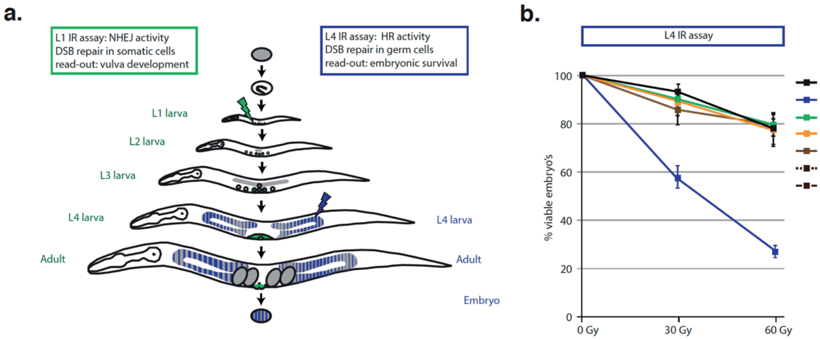

**Figure S5. Germ cells deficient for *thoc-5* or *thoc-7* are not hypersensitive to IR**

**a.** Schematic representation of two different IR assays in the context of the C. elegans life cycle and challenged tissues. While the L1 assay measures IR-resistance of arrested vulva precursor cells and mainly reflects NHEJ activity, the L4 assay measures IR-resistance of germ cells and typically reflects HR activity. **b.** L4 assay; L4 animals were challenged by the indicted dose of IR and percentage of viable progeny is plotted. Values depict the average of three independent experiments and error bars represent S.E.M.

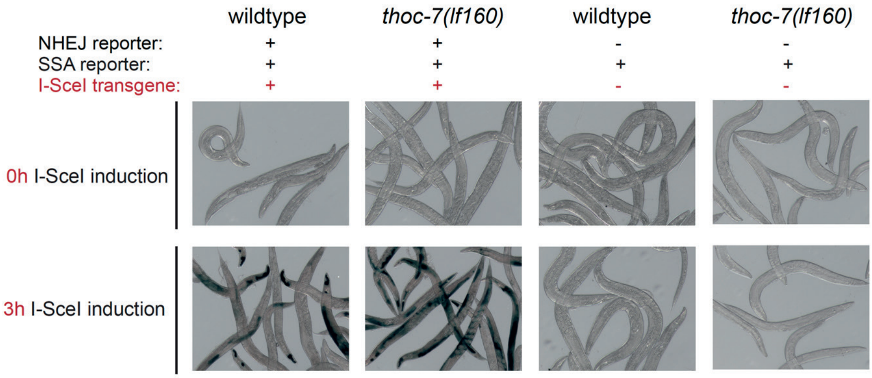

**Figure S6. Increased SSA in *thoc-7* mutants requires I-SceI induced DSB formation**

LacZ staining patterns in synchronized populations of animals carrying either the dual SSA/NHEJ reporter system or the SSA reporter only, the latter lacking the heat-shock inducible I-SceI expression transgene. Synchronized animals were heat-shocked 0 or 180 minutes at L1 stage and LacZ expression was detected in adults. Increased somatic LacZ (SSA) in *thoc-7* mutants required the presence of the I-SceI transgene and heat-shock-driven I-SceI expression.

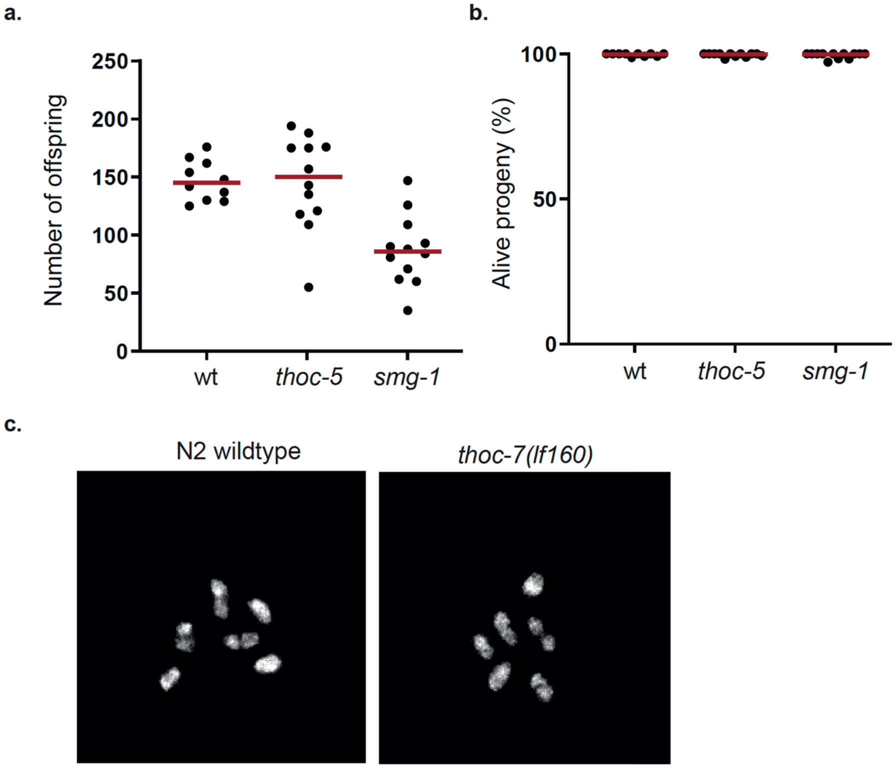

**Figure S7. Fecundity and meiotic recombination in *thoc-5* or *thoc-7* mutant animals**

**a.** Brood size quantification. Each dot depicts the number of offspring generated by a single worm. Lines depict median. **b.** Embryonic survival quantification. Each dot depicts the percentage of alive offspring of the brood generated by a single parent. **c.** DAPI-stained diakinesis nuclei (germline staining protocol as in [7]). The presence of six intact bivalents is indicative of proficient meiotic recombination.

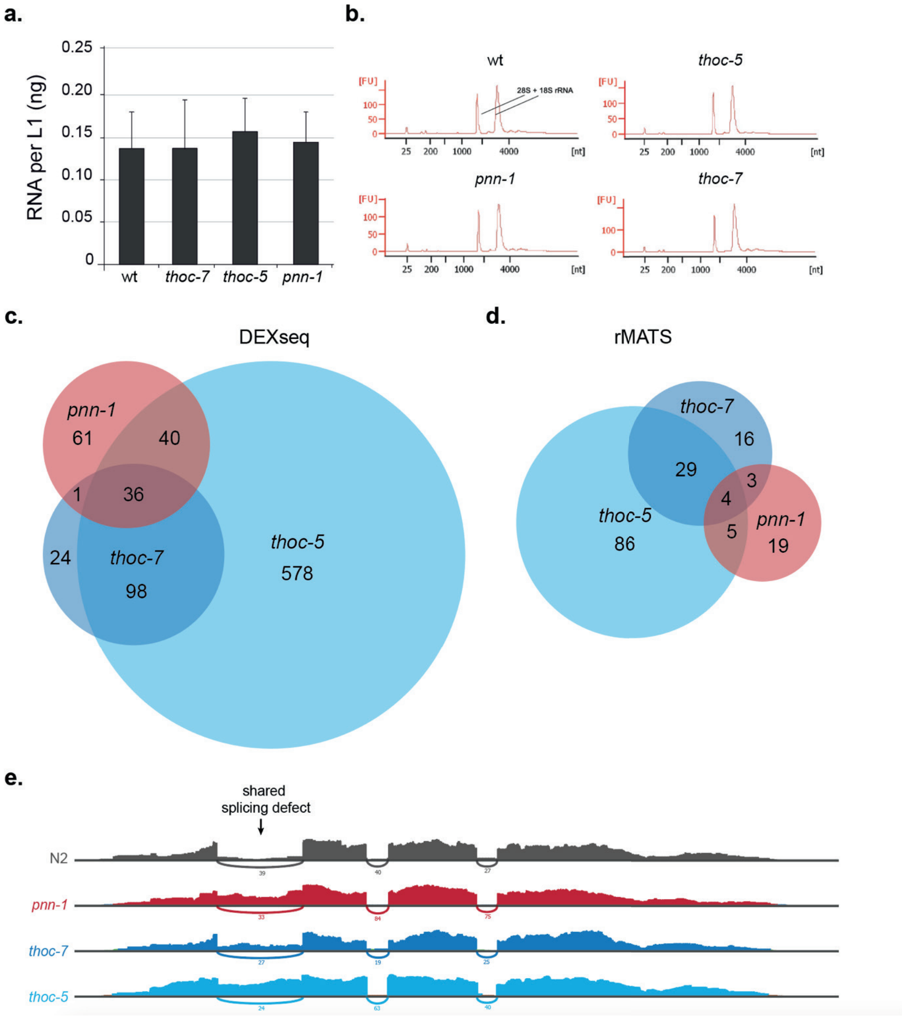

**Figure S8. RNAseq reveals shared splicing defects in THO and PNN-1 mutants**

**a.** Average total RNA yield per L1 animal determined from four independent L1 populations of the indicated genotype. **b.** RNA quality and size distribution determined using Agilent Eukaryote Total RNA Nano Lab-on-a-Chip. Panels depict representative RNA profiles for each genotype. Two well-defined peaks of 28S and 18S rRNA are indicative of intact RNA. **c.** Venn diagram of exon expression changes compared to N2 controls identified by DEXseq **d.** Venn diagram of alternative splicing events compared to N2 controls identified by rMATS. **e.** Example of shared intron retention in *nhr-263* mRNA identified by iREAD. Simplified sashimi plots depict relative RNAseq read-depth for each genotype. The number of junction reads in *nhr-263* are indicated at the arcs connecting the pair of exons.

**Figure S9. Gene Ontology ribbons of down-regulated genes in THO mutants**

Table visualizing the GO terms/biological processes annotated to the 43 genes of which mRNA expression was significantly decreased in *thoc-5* and *thoc-7* mutants compared to wildtype controls, detected by DEseq. Darker blue boxes indicate terms with the most annotations; white boxes represent terms that are not annotated for this gene. Red box highlights annotation to GO term 'DNA metabolic process', which includes DNA repair and DNA replication. On the right, genes annotated to the biological processes 'Response to Stimulus' or 'DNA metabolic process' are highlighted with light blue or red arrows, respectively.

**Figure S10. Gene Ontology ribbons of up-regulated genes in THO mutants**

Table visualizing the GO terms/biological processes annotated to the 82 genes of which mRNA expression was significantly increased in *thoc-5* and *thoc-7* mutants compared to wildtype controls, detected by DEseq. Labels/annotation as in Figure S9.

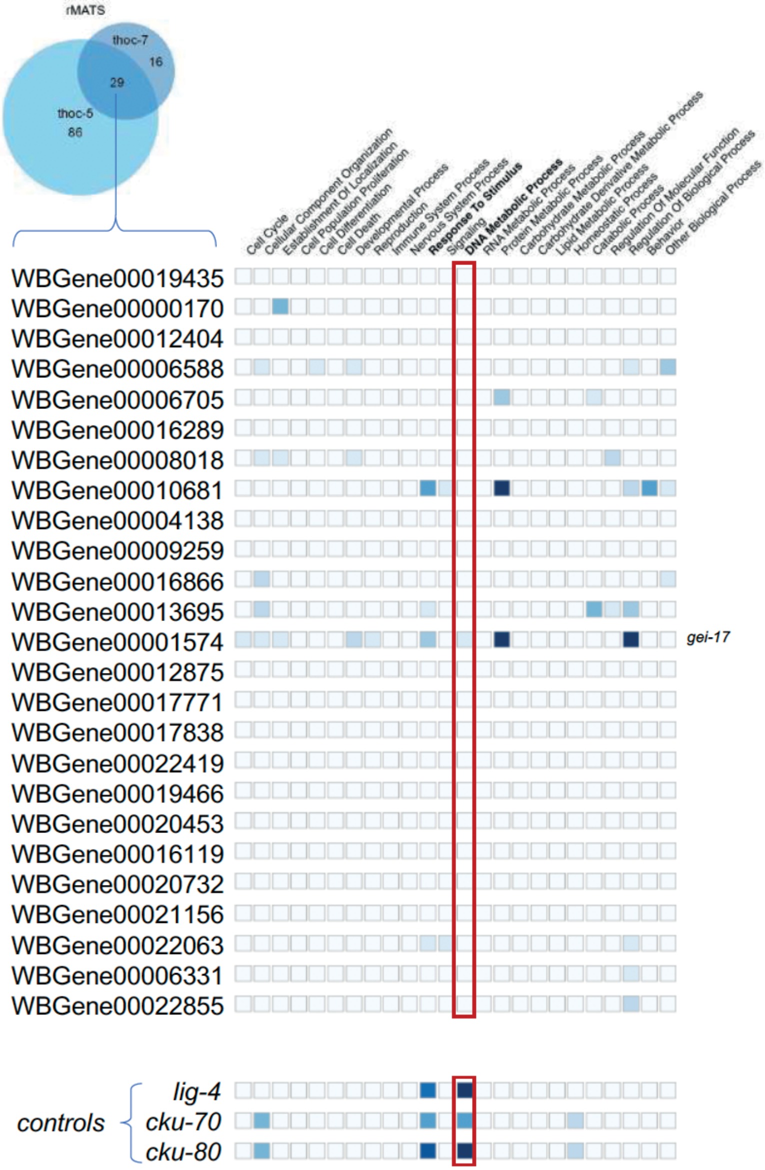

**Figure S11. Gene Ontology ribbons of mis-spliced transcripts in THO mutants**

Table visualizing the GO terms/biological processes annotated to the 29 genes of which mRNA expression was significantly increased in *thoc-5* and *thoc-7* mutants compared to wildtype controls, detected by rMATS. Darker blue boxes indicate terms with the most annotations; white boxes represent terms that are not annotated for this gene. Red box highlights annotation to GO term 'DNA metabolic process', which includes DNA repair and DNA replication. Control genes with known roles in NHEJ are depicted below. Although *gei-17* is implicated in DNA repair and DNA replication stress responses, only a minor fraction of *gei-17* transcripts is alternatively spliced and its overall expression is increased 1.4 fold.

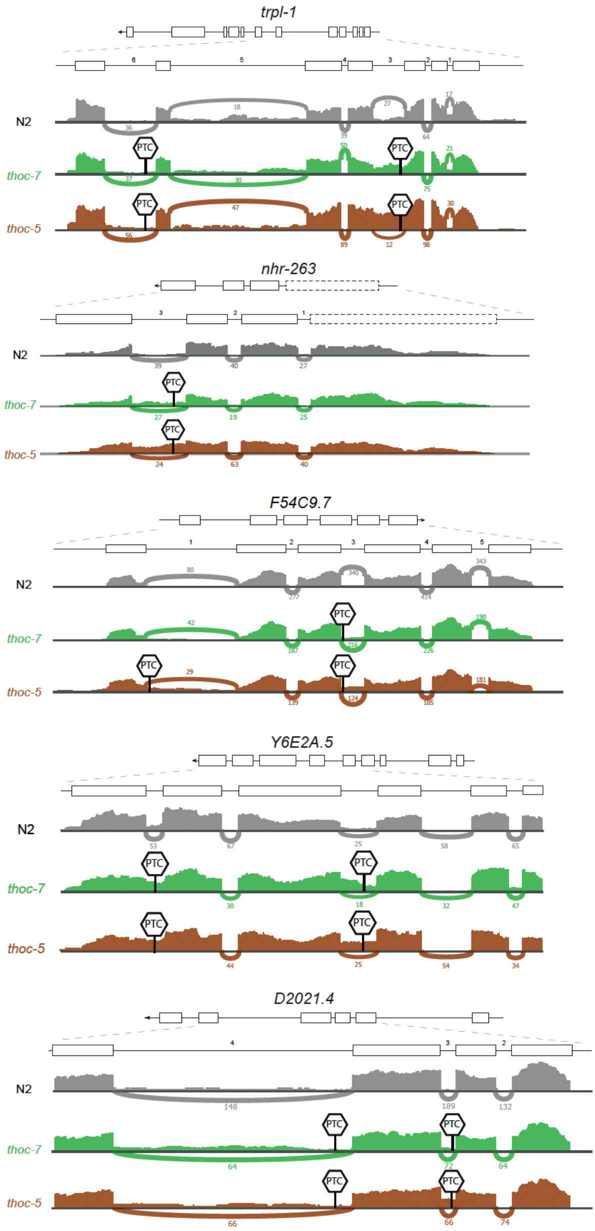

**Figure S12. Examples of PTC-bearing transcripts in *thoc-5* and *thoc-7* mutants**

Five genes exemplifying increased non-sense transcripts in *thoc-5* and *thoc-7* mutants. For each gene, simplified gene models and sashimi plots are depicted. Sashimi plots indicate relative RNAseq read-depth for each genotype and junction reads are shown as arcs connecting the pair of exons. PTC sign indicates position of a predicted premature stop codon.

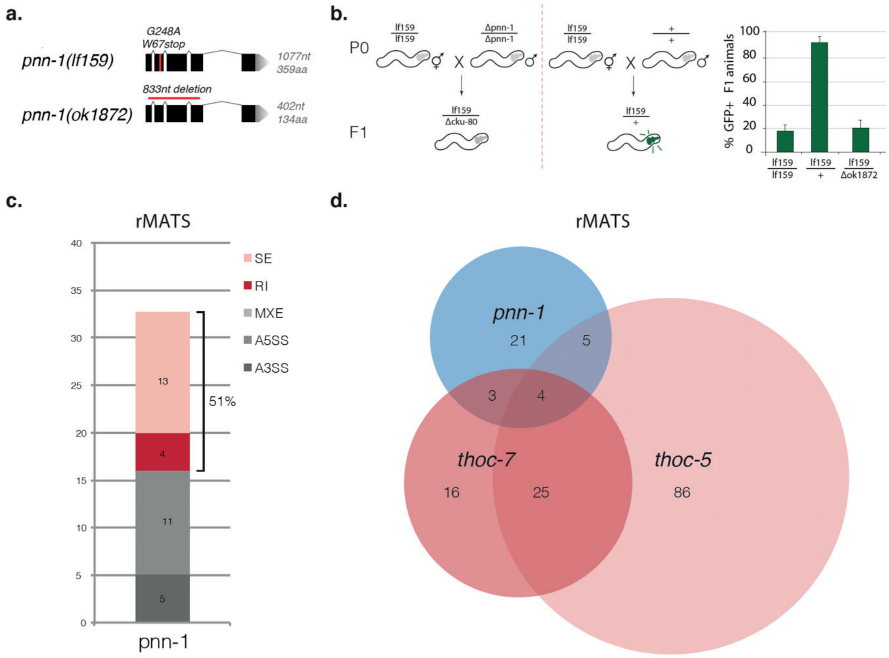

**Figure S13. Validation of NHEJ and RNA splicing defects in *pnn-1* deficient animals**

**a.** Gene models and loss-of-function alleles of *pnn-1*. **b.** Crossing scheme used for complementation analysis. Newly identified NHEJ mutants (homozygous for *lf159* allele) are crossed with either wildtype animals or *pnn-1* deletion mutants. trans-heterozygous F1 progeny is analysed for somatic NHEJ activity. Right graph shows quantification of GFP-positive pharynges in F1 cross progeny, heat-shocked for 120 minutes and measured in adults. **c.** Bar chart indicates number of splicing events detected and altered in *pnn-1* deletion mutants compared to wildtype controls. Splicing events as identified by rMTAS: SE, skipped exon; RI, retained intron; MXE, mutually exclusive exon; A5SS, alternate 5' splice site; A3SS, alternate 3' splice site. **d.** Venn diagram depicts overlap among differentially spliced genes in *pnn-1*, *thoc-7* and *thoc-5* mutants. Alternatively splicing of C53C7.5, *pdxk-1*, W02F12.4 and F07B10.4 is shared by all three mutants.

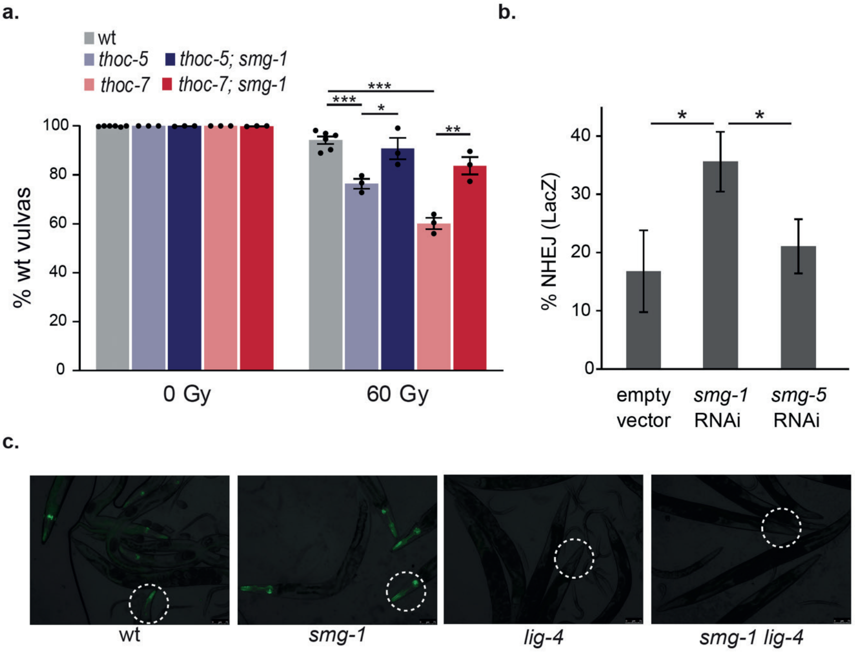

**Figure S14. Confirmation of SMG-1 as a suppressor of classical NHEJ**

**a.** The newly identified smg-1 allele (lf238) alleviates IR hypersensitivity of THO mutants. L1 animals were challenged with indicated doses of IR and vulva development was scored. Values depict averages of independent experiments and error bars represent S.E.M. Dots indicate the average of each experiment (two-tailed t-tests *P<0.05; **P<0.01; ***P<0.001) **b.** NHEJ activity in *thoc-7* mutants after RNAi-mediated knockdown of *smg-1* or *smg-5*. Bars indicate average fraction of LacZ-positive pharynges of three populations. Error bars depict standard deviation (two tailed t-tests, *P<0.05). **c.** GFP expression in heat-shocked dual reporter animals for each indicated genotype. Representative pharynges are encircled. Loss of *smg-1* does not restore GFP expression in *lig-4* mutants, indicating that *smg-1* deficiency does not alter the specificity of the NHEJ reporter.

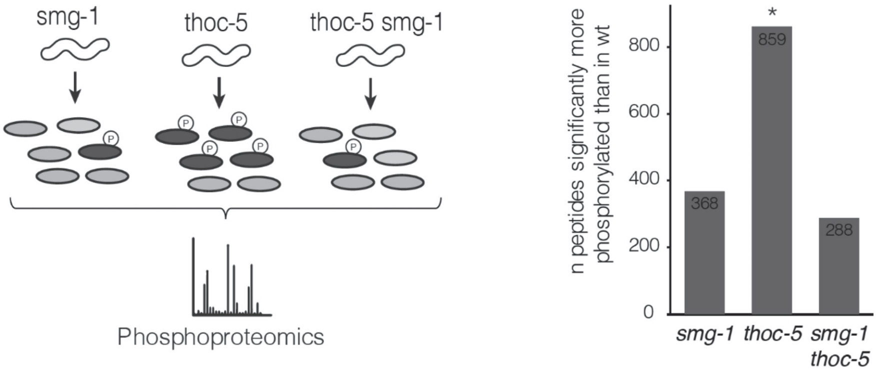

**Figure S15. Altered protein phosphorylation upon thoc-5 and/or smg-1 loss**

Schematic of proteomics approach (left). Proteins lysates were obtained from asynchronous populations of *smg-1*, *thoc-5* and *thoc-5; smg-1* mutants, along with wildtype controls, and enriched for phosphopeptides. For each genotype, unique phosphorylation events were detected and quantified using mass spectrometry. Bar graph represents the quantification of the amount of peptides significantly more phosphorylated than in wild-type worms (permutation test, *P<0.05).

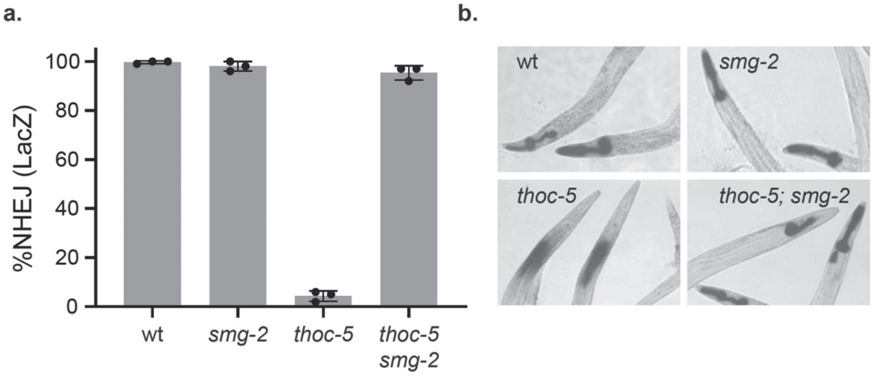

**Figure S16. SMG-2 deficiency restores NHEJ activity in *thoc-5* mutant animals**

**a.** Quantification of LacZ-positive pharynges. Dual reporter animals are heat-shocked for 120 minutes at L1 stage to induce I-SceI expression and LacZ reporter expression is analysed in adults. Dots indicate the average of independent populations. Error bars represent standard deviation. **b.** LacZ expression patterns of heat-shocked dual reporter animals (quantified in a.).

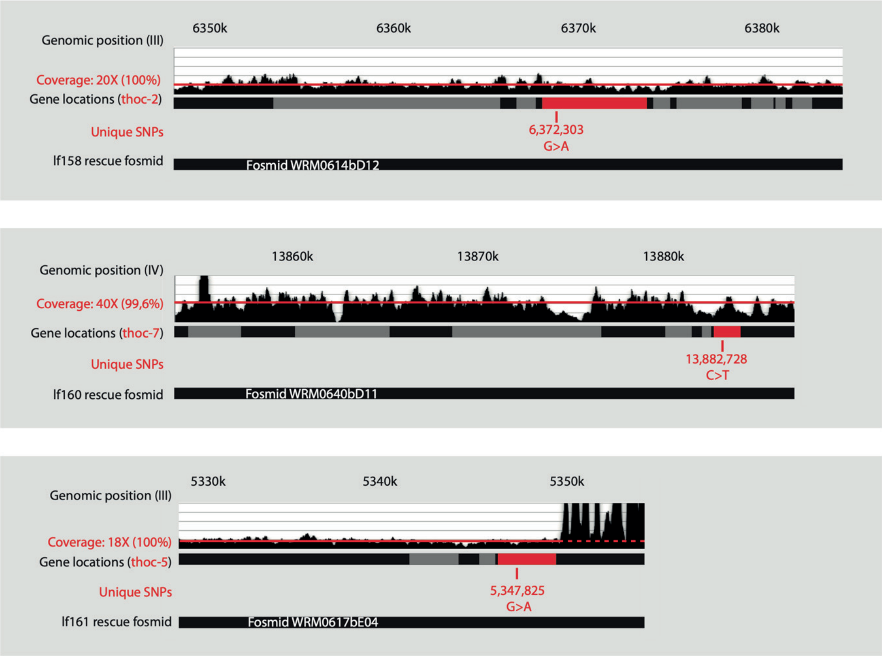

**Figure S17 Coverage and genomic context of regions spanned by rescue fosmids**

Unique non-synonymous SNPs in the genomic regions spanned by the injected fosmids (highlighted in red). Despite high sequence coverage, no other unique variants than those in the THO genes were found in the respective NHEJ mutants, advocating that we specifically complemented THO deficiency despite the use of 2-4kb fosmids.