THE CONSERVED ROLE OF RNA SPLICING FACTORS IN GENOME MAINTENANCE

by

ANNIE S. TAM

B.Sc., Simon Fraser University, 2011

M.Sc., University of British Columbia, 2014

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(MEDICAL GENETICS)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

March 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

THE CONSERVED ROLE OF RNA SPLICING FACTORS IN GENOME MAINTENANCE

submitted by Annie S. Tam in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

in Medical Genetics

Examining Committee:

Peter Stirling, Medical Genetics
Supervisor

Vivien Measday, Medical Genetics
Supervisory Committee Member

LeAnn Howe, Biochemistry
University Examiner

Peter Lansdorp, Medical Genetics
University Examiner

Karim Mekhail, Laboratory Medicine and Pathobiology, University of Toronto
External Examiner

Additional Supervisory Committee Members:

Gregg Morin, Medical Genetics
Supervisory Committee Member

Michael Kobor, Medical Genetics
Supervisory Committee Member

Abstract

RNA splicing mutants have been broadly implicated in genome stability, but mechanistic links are often unclear. Two predominant models have emerged: one involving changes in gene expression that perturb other genome maintenance factors and another in which genotoxic DNA:RNA hybrids, called R-loops, impair DNA replication. Recent efforts in whole genome sequencing have identified splicing factor mutations in several cancers, suggesting that splicing disruption may be a common mechanism involved in oncogenesis.

To understand how splicing factor mutations contribute to genetic instability (GIN) in budding yeast, I selected strains with mutations in core snRNP complexes involved in establishing the splicing reaction and characterized GIN phenotypes to find that mitotic defects, and in some cases R-loop accumulation, are causes of GIN. I observed evidence of R-loop induced DNA damage in some cases, while all splicing mutants tested caused GIN through aberrant splicing of the *TUB1* transcript, the protein product of which, α-tubulin, is critical in forming the mitotic spindle. GIN is exacerbated by loss of the spindle-assembly checkpoint protein Mad1, and moreover, removal of the intron from the α-tubulin gene *TUB1* restores genome integrity.

To gain functional insights to how *HSH155* could influence GIN in the context of cancer progression, I studied five cancer-associated *SF3B1* point mutations in the yeast ortholog *HSH155*. While the splicing activity in Hsh155 and SF3B1 were conserved, I did not observe measurable phenotypes in the yeast mutant strains. Thus, I used isogenic NALM6 human leukemia cell lines to investigate how a specific *SF3B1* hotspot mutation, H662Q contributes to GIN. My data indicate that GIN occurs in two ways: 1) by induction of R-loop-mediated replication stress either directly or indirectly through suboptimal expression of an R-loop modulating factor, and 2) aberrant splicing of the multifunctional protein DYNLL1, which may potentially perturb double strand break repair pathway choice.

The results of my study show how differing penetrance and selective effects on the transcriptome in yeast and human splicing factors contribute to GIN through R-loop accumulation and altered gene expression, adding to a growing body of evidence that splicing factors play a key role in genome maintenance across species.

Lay Summary

RNA splicing factors, which are proteins that cut, paste, and re-assemble the “messages” encoded by DNA to direct protein assembly, are frequently mutated in many types of cancers. It is thought that these mutations could be causing cancer by increasing genetic instability, which is when mutations or changes in DNA occur more often than normal. Although we know mutations in RNA splicing factors seem to increase genetic instability from yeast to humans, how this happens is not clear.

I used both yeast and human cancer cells with mutations in specific RNA splicing factors to determine how this could lead to genetic instability. I found that these mutations cause genetic instability by forming DNA damaging nucleic acid structures called R-loops, which occur when newly made RNA inappropriately attaches to DNA. In addition, I also found that these mutations affect how the “message” between DNA and protein is re-assembled, leading to changes to the proper amount of protein made that is required for cells to function properly.

This work identifies distinct causes of genetic instability when specific RNA splicing factors are mutated in yeast and human cells.

Preface

This dissertation is the original intellectual product of Annie S. Tam, unless otherwise noted below.

A version of Chapter 1 has been published as a review article written collaboratively by myself and Peter C. Stirling as Tam, A.S. and Stirling, P.C. "Splicing, genome stability and disease: splice like your genome depends on it!" (2019) *Curr Genet.* 65(4):905-912.

A version of Chapter 2 has been published as Tam, A.S., Sihota, T.S., Milbury, K.L., Zhang, A., Mathew, V., and Stirling, P.C. (2019) "Selective defects in gene expression control genome instability in yeast splicing mutants" *Mol Biol Cell.* 30(2):191-200. I designed, performed, and analyzed all of the experiments with guidance from P. Stirling, except for the following: P. Stirling performed the synthetic genetic array experiment, GO biological process and cellular component enrichment analysis in Figure 2.7. Double mutants used to validate the synthetic genetic array hits were made by A. Zhang and V. Mathew (Figure 2.8). K.L. Milbury performed the Western blot shown in Figure 2.14B. T.S. Sihota helped generate the *snu114-60* and *hsh155-1 tub1*Δ*i* double mutants and helped perform subsequent colony PCR verification in Section 2.3.5. The model schematic in Figure 2.18 was designed and illustrated by V. Mathew and is used with permission.

A version of Chapter 3 formed the basis of a manuscript currently in preparation, titled "SF3B1 cancer-associated mutations cause replication stress and alter fork dynamics." I designed, performed, and analyzed all of the experiments with guidance from P. Stirling, except for the following: Hsh155 cancer-associated yeast mutants used in Section 3.3.1 – Section 3.3.2 were generated by V. Mathew using site-directed mutagenesis.

Table of Contents

List of Tables

List of Figures

List of abbreviations and acronyms

3'SS	3' splice site
5'SS	5' splice site
ALF	A-like faker
AML	Acute myeloid leukemia
BER	Base excision repair
BIR	Break-induced replication
BP	Branch point
CEN	Centromere
CIN	Chromosome instability
CLL	Chronic lymphocytic leukemia
CMML	Chronic myelomonocytic leukemia
CTF	Chromosome transmission fidelity
DNA	Deoxyribonucleic acid
DSB	Double strand break
dsDNA	Double strand DNA
EdU	5-Ethynyl-2´-deoxyuridine
GFP	Green fluorescent protein
GO	Gene ontology
HEAT	Huntingtin, elongation factor 3, protein phosphatase 2A, TOR1
HR	Homologous recombination
MDS	Myelodysplastic syndrome
mRNA	Messenger RNA
NHEJ	Nonhomologous end joining
OD	Optical density
PARP	Poly (ADP-ribose) polymerase
PBS	Phosphate-buffered saline
PVDF	Polyvinylidene difluoride
RNA	Ribonucleic acid
RNaseH1	Ribonuclease H1
RT-qPCR	Reverse transcription - quantitative polymerase chain reaction
SC	Synthetic complete
SDS-PAGE	Sodium dodecyl sulfate- polyacrylamide gel electrophoresis
SGA	Synthetic genetic array
snRNP	Small nuclear ribonucleoproteins
ssDNA	Single strand DNA
TRC	Transcription replication conflict
WT	Wildtype
YPD	Yeast Extract-Peptone-Dextrose

Acknowledgements

I am grateful to the scientific community and colleagues who were incredibly generous with their help, advice, and discussions. I would like to acknowledge key people who helped me so much along the way: Dr. Veena Mathew, Dr. Libin Abraham, Dr. George Chung, Dr. Emily Chang, Dr. Shuhe Tsai and my committee members Dr. Gregg Morin, Dr. Vivien Measday, and Dr. Michael Kobor. I would also like to acknowledge the Department of Medical Genetics, in particular Cheryl Bishop, colleagues from the Terry Fox Laboratory, in particular Dr. Samuel Gusscott, Rod Docking, and Dr. Avinash Thakur, and my labmates over the years: Dr. Karissa Milbury, Arun Kumar, Alynn Shanks, Carolina Novoa, Romulo Segovia, Michael Yuen, Tianna Sihota, Anni Zhang, Kyra West, Dr. Sean Minaker, and all the Stirling Lab members past and present.

I'd like to especially thank the patience and encouragement of my family: Mandy, Dave, Turtle, Millie, and my parents.

Finally, I would like to thank my supervisor Peter Stirling for taking me on as a student, for pushing me to do better, and for providing a positive and nurturing environment to allow me to grow as a scientist.

Dedication

The journey of a PhD student can sometimes be very lonely. It was made much more pleasant by a host of incredible fellow students I had the pleasure of meeting along the way. This thesis is dedicated to the current graduate students of the Stirling Lab: Arun Kumar, Hilary Brewis, James Wells, Louis-Alexandre Fournier, Justin White, and Leticia Dinatto.

Chapter 1: Introduction

1.1 Genome instability, RNA processing, and disease

Genome maintenance pathways encompass all cellular activities whose functions prevent or respond to DNA damage, or enforce faithful DNA replication or mitotic chromosome segregation. The effect of genome maintenance in dividing cells is to prevent the accumulation of mutations or chromosome structural changes, within tissues and across generations. The maintenance of a relatively stable genome is critical for cell and organism viability, and cells engineered to have excessive mutation rates become inviable [1]. Moreover, robust genome maintenance pathways are important to resist environmental challenges, such as radiation or chemicals, that themselves damage cellular DNA, requiring repair and stressing endogenous systems.

The identification of genome maintenance regulators is perhaps of greatest interest in the context of cancer. Cellular defects that increase mutation rates or the rate of chromosomal instability are known to drive tumorigenesis [2, 3]. For example, Indeed, genome instability is now considered a hallmark of cancer [4]. Cells with increased genome instability are more likely to give rise to clones with mutations in cancer driver genes. In addition, genomically unstable tumors may be more likely to give rise to drug-resistant clones. At the same time, some defects in genome maintenance create specific susceptibility to cancer therapies. For example, PARP (poly(ADP-ribose) polymerase) inhibitors are highly effective on tumors with defects in homology-directed repair of DNA, a crucial genome maintenance pathway [5]. PARP proteins facilitate single strand break repair by the base excision repair (BER) pathway, and inhibition of the BER pathway by PARP inhibition results in accumulation of unrepaired single strand breaks, leading to the formation of double strand breaks that require HR-mediated repair [6]. Alternatively, mutation burden itself can be exploited for therapies based on immunological recognition of tumor neo-antigens. It is now clear in several cancer settings that genomically unstable tumors elicit better responses to immune checkpoint blockade therapies that stimulate anti-tumor immunity [7-9]. Given the importance of genome instability in cancer, there is a need to understand the full scope and mechanism of genome maintenance pathways in human cells.

Simplistically, there are three direct effector pathways controlling genome maintenance: DNA replication, DNA repair, and mitosis. However, work in model systems and human cells have revealed an incredible diversity of seemingly peripheral cellular activities that play roles in the fidelity of these three core pathways [10-12]. Over the past 2 decades, each step of RNA production and processing has been linked to genome maintenance. There are examples of transcriptional machinery [13], splicing factors [14, 15], transcription termination machinery [12, 16], RNA transport factors [17], and RNA degradation factors [18, 19] all being linked to increased rates of chromosome missegregation or to excess DNA damage. This may not be surprising; RNA production begins on the DNA template, and ultimately influences the levels of all cellular proteins. Indeed, there are numerous reported functional links between RNA processing and DNA repair machinery [20]. Nevertheless, it is an ongoing challenge to ascribe mechanisms of action to RNA processing defects that cause genome instability.

In this thesis, my focus is on RNA splicing as there is growing evidence that a general consequence of splicing defects is genome instability. Mutations in splicing proteins are of particular interest to human health as they are associated with both rare inherited human diseases and with various cancers. Human diseases have been directly linked to mutations in spliceosome components or splicing regulators (**Table 1.1**). Haploinsufficiency of core components such as SF3B4 or EFTUD2 leads to developmental defects [21, 22], while mutations in tri-snRNP can cause retinitis pigmentosa [23]. In contrast, many cancers seem to be associated with disruption of normal U2 snRNP function which can lead to alternative splice site selection, or changes to exon skipping and inclusion [24]. A pan-cancer analysis of splicing factor mutations in 33 cancer types reveals frequent splicing factor mutations [25], suggesting that this is a common mechanism involved in oncogenesis. Exactly how splicing factor mutations drive cancer, or rare diseases in such a tissue-specific manner remains an open question.

Gene	Function in Spliceosome	Disease	References
SF3B4	U2 snRNP	Nager Syndrome	[21]
EFTUD2	U4/U6.U5 tri-snRNP	Mandibulofacial dystosis with microcephaly	[22]
PRPF31	U4/U6.U5 tri-snRNP	Retinitis pigmentosa	[23]
U4 atac snRNA	U12 spliceosome	Microcephalic osteodysplastic primordial dwarfism	[26, 27]
SNRPB	U1, U2, U4/U6 and U5-snRNP	cerebro-costo-mandibular syndrome	[28]
SF3B1	U2 snRNP, branch point	MDS, CLL, AML, BrCa, UM and other cancers	[29]
SRSF2	Ser-Arg rich protein	MDS, AML, CMML	[29]
U2AF1	U2 auxiliary factor	MDS, AML	[29]
ZRSR2	U2 auxiliary factor	MDS	[29]
PRPF8	U4/U6-U5 tri-snRNP	Intraductal papillary mucinous neoplasm, AML	[30, 31]
CDK12	Required for splicing	Breast, ovarian and prostate cancer	[32, 33]
PRMT5	Regulate splicing proteins by arginine methylation	Leukemia, lymphoma	[34, 35]

Table 1. 1 A partial list of splicing-defect-associated diseases and cancers
Abbreviations: MDS (myelodysplastic syndrome), CLL (chronic lymphocytic leukemia), AML (acute myeloid leukemia), BrCa (breast cancer), UM (uveal melanoma), CMML (chronic myelomonocytic leukemia)

1.2 The splicing machinery

1.2.1 The splicing reaction

In eukaryotes, DNA is transcribed into pre-mRNAs that often contain noncoding intron sequences, the removal of which, and subsequent ligation of exon sequences, are essential steps in gene expression. This process is carried out by an evolutionarily conserved large ribonucleoprotein complex called the spliceosome, and is a collection of 5 small nuclear ribonucleoproteins (snRNPs U1, U2, U4, U5, and U6) containing small stable RNA bound to proteins, and hundreds of associated proteins [36, 37]. The splicing reaction is completed through two successive transesterification reactions that produce either constitutive or alternative splicing patterns (**Figure 1.1**).

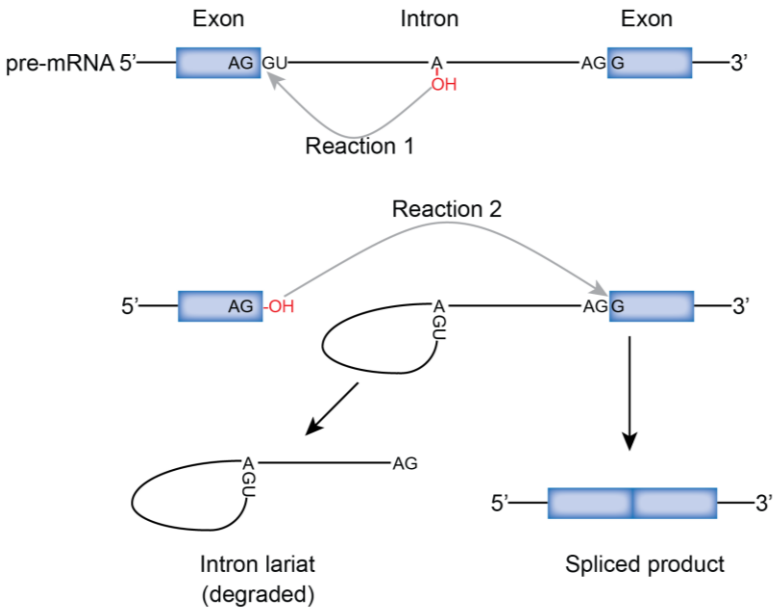

Figure 1. 1 The biochemical splicing reaction.

Splicing involves removal of introns and ligation of exons from the pre-mRNA. Intron removal occurs by two successive transesterification reactions. The 5' end of the intron is first cleaved, then during the first transesterification reaction, the hydroxyl group (-OH group depicted in red) on adenine binds to the guanine nucleotide at the 5' splice site (depicted by GU nucleotides), forming a lariat structure ("Reaction 1"). Subsequently, the hydroxyl group at the 3' end of the exon binds to the 3' splice site ("Reaction 2"), resulting in covalently bound exons. The resulting cleaved intron lariat is subjected to degradation.

1.2.2 Alternative splicing

Constitutive splicing is the process of intron removal and ligation of exons in the order they appear in a gene to form a mature mRNA product. Alternative splicing is the process in which exons can be included or excluded from a single pre-mRNA, leading to a diverse combination of mature mRNA transcripts called isoforms, that serves to increase transcriptome complexity, and plays important roles in cellular differentiation and development [38, 39]. Alternative splicing events are classified into five main types: (i) exon skipping, which occurs when an exon and the flanking introns are spliced out; (ii) mutually exclusive exon, which occurs when different combinations of exons are included or excluded resulting in distinct products; (iii) alternative 3' and (iv) 5' splice site, which occurs when two or more splice sites at the end of an exon are recognized by the splicing machinery; and (v) intron retention, which occurs when an intron is not spliced out of the mature mRNA [39] (**Figure 1.2**). Alternative splicing is a key mechanism in generating transcriptome diversity, and many human diseases have been linked to dysregulation of alternative splicing caused by both *cis*-acting mutations in RNA sequence elements that alter splicing patterns [40], and splicing factor mutations (**Table 1.1**).

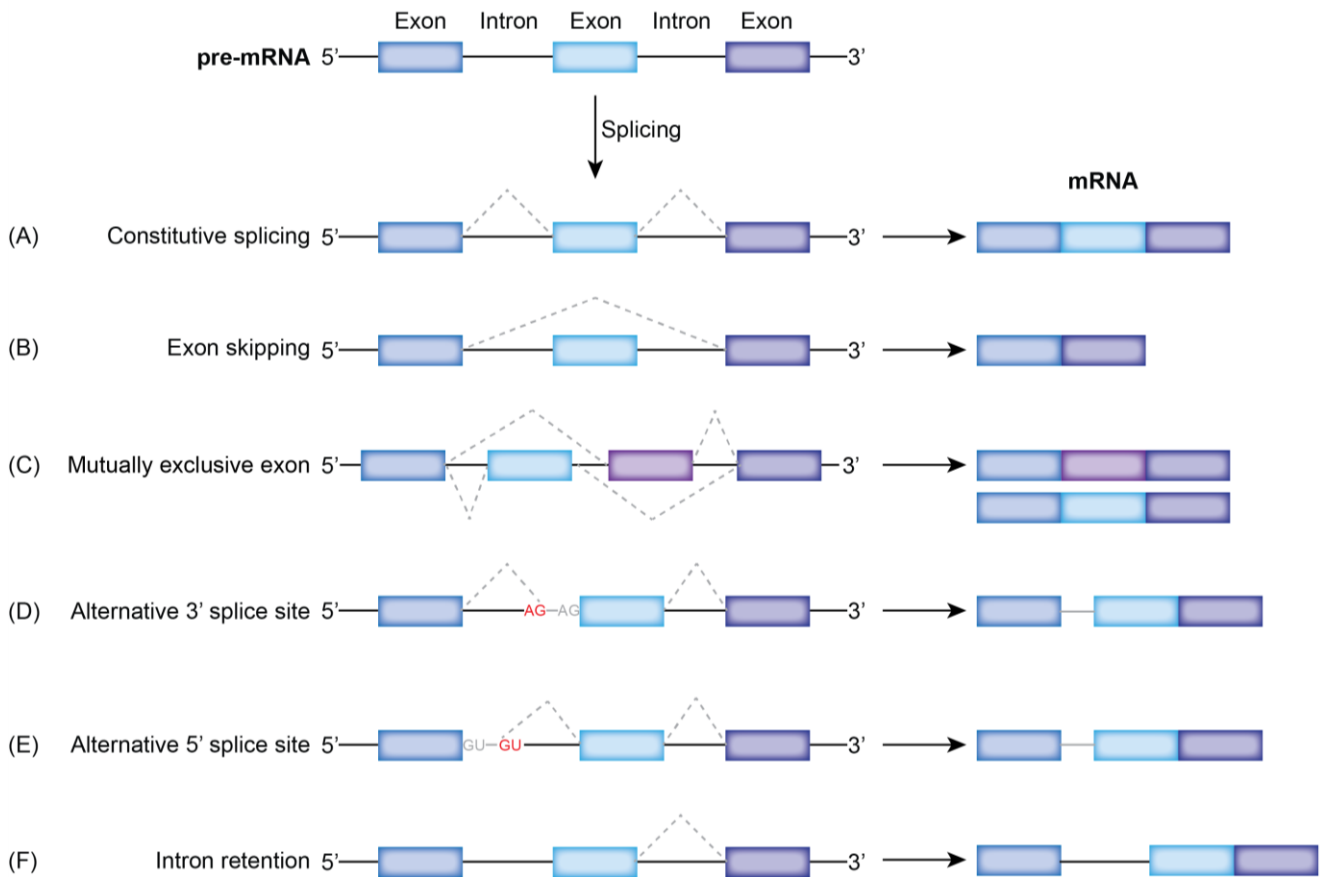

Figure 1. 2 Alternative splicing events.

Types of alternative splicing events that lead to a diverse combination of transcripts called isoforms: (A) Constitutive splicing (B) Exon skipping (C) Mutually exclusive exon (D) Alternative 3' splice site (E) Alternative 5' splice site (F) Intron retention. mRNA isoforms contribute to proteome diversity. Exons are represented by blue rectangles. Gray dotted lines represent splicing events. In (D) and (E), gray nucleotides represent canonical splice sites while red nucleotides represent cryptic splice sites. Adapted and modified from Wang Y. *et al*, 2015 *Biomed Rep* [39].

1.2.3 Conserved splice site sequences in introns

Splicing is an essential cellular process, and is critical for gene expression in eukaryotic cells. It has been estimated that >90% of human genes undergo splicing, and in *Saccharomyces cerevisiae* (budding yeast), while only ~5% of genes undergo splicing, the bulk of mRNAs are

transcribed from intron containing genes since the most highly expressed genes are intron-containing ribosomal protein genes [41]. Due to the strong conservation of the core spliceosome in yeast and human cells, yeast has been particularly useful in identifying components of the splicing machinery, and as a model to understand the basic mechanisms of splicing. As splicing proceeds, the spliceosome assembles in a step-wise manner onto nascent pre-mRNA, and recognizes short sequences in the intron called splice site sequences via RNA-RNA and RNA-protein interactions [42]. Ordered assembly of the various spliceosome components configures into a catalytically active structure that can coordinate precise excision of introns and ligation of exons to form mRNA. These short intron sequences are: the 5' splice site (5'SS), the branch site or branch point (BP), the 3' splice site (3'SS), and in humans a polypyrimidine tract (PPT) upstream to the 3'SS [43]. These splice sequences can be variable across species – in yeast, the splice site sequences are conservative, while in human cells there is a higher degree of sequence degeneracy (**Figure 1.3**) [44].

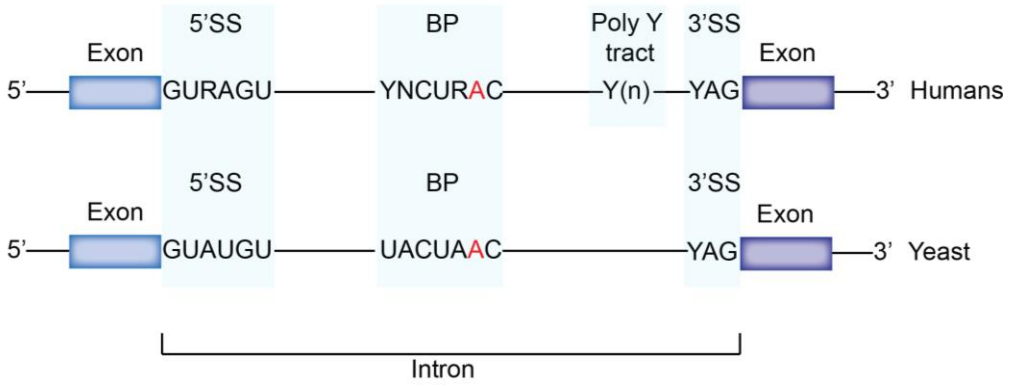

Figure 1. 3 Splice site sequences in human and yeast introns.

Schematic representation of conserved sequences in human and yeast introns. Y/R = pyrimidine/purine; 5'SS = 5' splice site; 3'SS = 3' splice site; Poly Y tract = polypyrimidine tract; BP = branch point; Red nucleotides depict the branch site adenosine. Adapted and modified from Will C.L. and Lührmann R., 2011 *Cold Spring Harb Perspect Biol* [42].

1.2.4 Spliceosome assembly

The splicing reaction described in Section 1.2.1 is chemically simple, but catalyzed by a complex macromolecular spliceosome. The spliceosome must assemble in a step-wise manner involving various ribonucleoproteins, and must undergo extensive conformational changes to generate a catalytically active complex [42, 45, 46]. Splicing is facilitated by interactions between the 5 snRNPs previously mentioned (U1, U2, U4, U5, and U6), small stable RNA bound to proteins, an array of associated proteins, and also the conserved intron sequences described in Section 1.2.3. During spliceosome assembly, the U1 snRNP recognizes the 5' splice site (5'SS) through complementarity between the U1 snRNA and 5'SS sequence. In yeast the branch point (BP) is first recognized by BP binding protein (BBP), and in human cells, the BP, polypyrimidine tract, and 3' splice site (3'SS) are recognized by splicing factor 1 (SF1), U2AF2, and U2AF1, respectively. Subsequently, the U2 snRNP replaces BBP and SF1 at the BP due to complementarity between the U2 snRNA and the BP sequence, leading to the formation of the A complex. The U4/U6-U5 tri-snRNP then associates with the spliceosome, resulting in a pre-B complex, and dissociation of the U1 snRNP by ATPases generate a stable B complex. Additional ATP activity then dissociates U4 snRNA from the spliceosome, enabling U6 and U2 snRNAs to pair with the 5' end of the intron, resulting in catalytic activation (Activated B complex). Association between the 5'SS and BP results in a catalytically activated B complex, and the first transesterification reaction which generates an intron lariat and free 3' exon hydroxyl group. The spliceosome is remodeled through ATP activity, U5 snRNA aligns the 5' and 3' exons, and the resulting spliceosomal C complex performs the second transesterification reaction which results in exon-exon ligation and freeing of the intron lariat. The mRNA is released, and ATPases and the GTPase Snu114 disassembles the spliceosome. This simplified description of spliceosome assembly and disassembly is summarized in **Figure 1.4**.

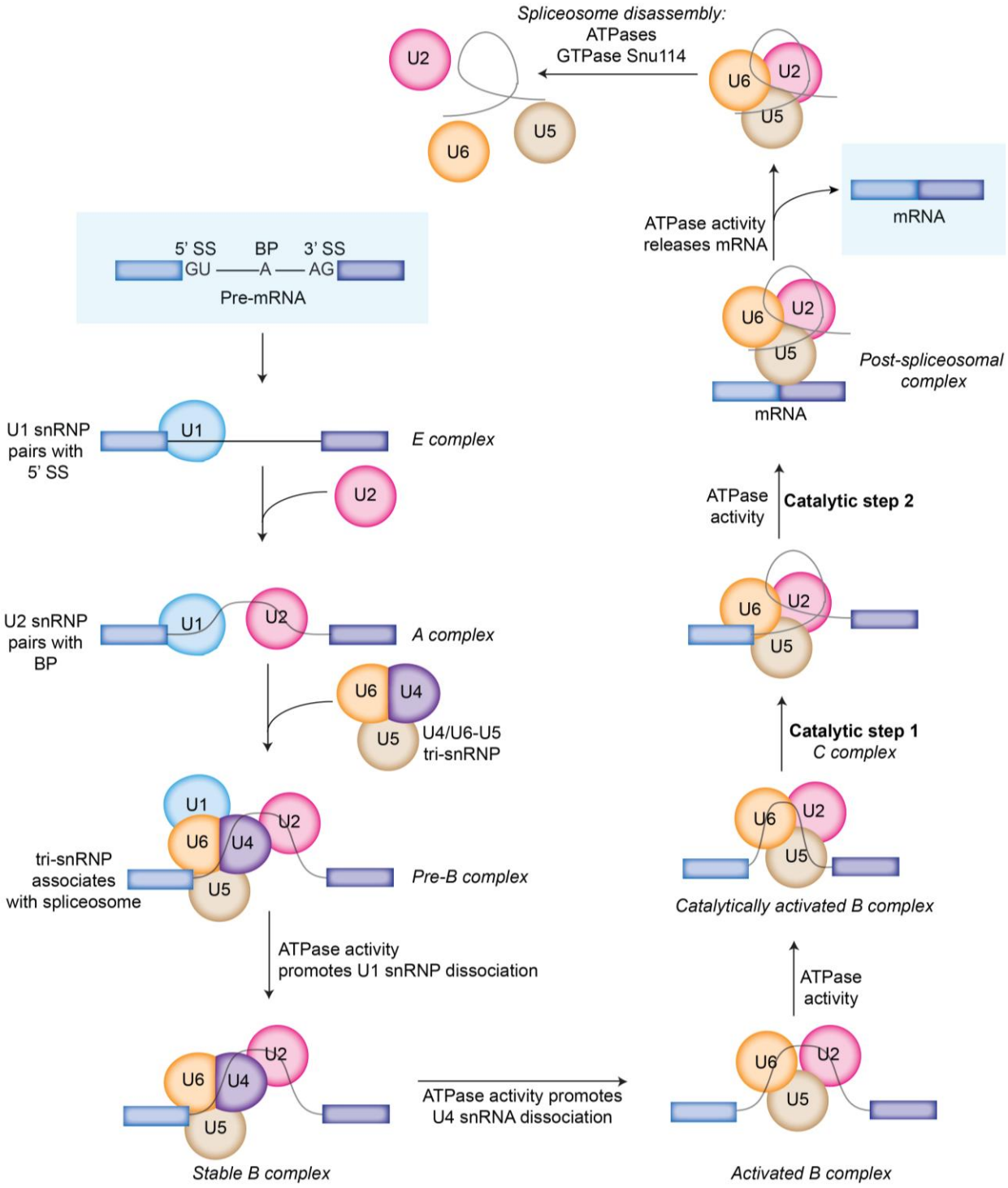

Figure 1. 4 Summary of spliceosome assembly.

Please refer to Section 1.2.4 for detailed description. Exons are represented by blue rectangles. 5'SS = 5' splice site; 3'SS = 3' splice site; BP = branch point. Adapted and modified from Will C.L. and Lührmann R., 2011 *Cold Spring Harb Perspect Biol* [42]. snRNP and pre-mRNA cartoons were drawn by V. Mathew.

1.3 Spliceosome mutation mediated genetic instability

An R-loop is a three-stranded nucleic acid structure composed of newly transcribed RNA, complementary DNA, and the displaced single stranded non-template DNA (**Figure 1.4**). Unscheduled or excessive R-loop accumulation has been linked to DNA damage and genome instability. The R-loop structure itself can be targeted by nucleases [47], or in some contexts by cytidine deaminase enzymes [48, 49], that directly damage the R-loop. Perhaps more commonly, R-loops are associated with transcription-replication conflicts (TRCs), where they can stall replication forks, leading to DNA replication stress [50].

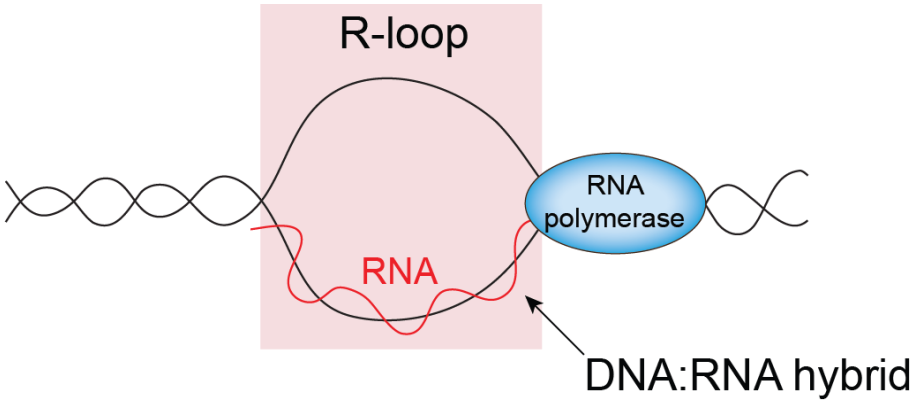

Figure 1. 5 Schematic representation of an R-loop.
An R-loop is a three-stranded nucleic acid structure formed when nascent RNA hybridizes to complementary DNA, resulting in a DNA:RNA hybrid and displaced single stranded DNA.

The act of splicing out intronic sequences should dramatically reduce the likelihood that R-loops will form because there will simply be less complementary sequence between the mature transcript and the DNA template. Accordingly, one of the earliest descriptions of R-loop induced genome instability was in cells lacking the splicing accessory factor, and Ser-Arg rich protein, ASF/SF2 [51]. These authors show that depletion of ASF/SF2 leads to DNA damage and the accumulation of R-loop structures. Importantly, they employ what has become a common control experiment: RNaseH1 overexpression as a means to reduce R-loop levels also reduced DNA damage in ASF/SF2 depleted cells [51]. Later, unbiased screens began to reveal that perturbing any of a large number of splicing factors leads to DNA damage, and much of this damage can be

suppressed by overexpression of RNaseH, suggesting a role for R-loops in splicing-factor loss mediated genome instability [52-54]. Chemical inhibition of splicing with diospyrin D1, or pladienolide B was also shown to increase R-loop levels [55, 56]. Other groups identified R-loop formation and genome instability accompanying the SNM1 splicing factor mutations in neuromuscular disease models [57]. Another study found that depletion of the tri-snRNP component SLU7 lead to R-loop induced DNA damage indirectly through alterations of a splicing accessory factor SRSF3 [58]. As a result, a broader appreciation that splicing factor perturbations can lead to R-loop associated DNA damage has been gained.

One critical study for interpreting the observations linking R-loops to splicing was published by the group of Benoit Palancade in 2017 [59]. The authors identified that introns play a special role in suppressing R-loop formation in *cis* to co-transcriptional splicing reactions, and are required to suppress DNA damage at highly transcribed genes. By artificially placing introns within R-loop prone genes, they were able to reduce their accumulation of R-loops, indicating that introns are sufficient to control R-loop occupancy *in vivo*. Interestingly, this study also found that it was recruitment of the spliceosome, not the act of splicing itself that was required to mitigate R-loops [59]. Indeed, this RNA packaging type function is similar to that proposed for the THO complex (made up of subunits Hpr1, Tho2, Mft1, Thp2, and Tex1 in *Saccharomyces cerevisiae*), which is also a critical suppressor of R-loop associated DNA damage [60, 61]. Finally, broader analysis of R-loop occupancy in human datasets showed that there also appeared to be a protective effect of introns on R-loop suppression in human cells [59].

Given the prevalence of R-loop accumulation in splicing factor deficient human cells, it has also become of interest to determine whether human diseases associated with splicing factor mutations have higher levels of R-loops and associated DNA damage. At least two studies have directly tested cancer-associated spliceosomal mutations for their impact on R-loop accumulation. Point mutations in the U2 snRNP component U2AF1, and the Ser-Arg rich splicing factor SRSF2, are drivers of myelodysplastic syndrome (MDS) [62-64]. Two studies demonstrated that point mutations in U2AF1 and SRSF2 led to R-loop accumulation and DNA damage in cell lines derived from MDS patients, [63, 64], and in normal CD34+ hematopoietic progenitor cells [63]. Thus, dominant cancer-associated point mutations in splicing factors drive R-loop formation and associated DNA damage due to replication stress.

1.4 Spliceosome regulation of genome maintenance gene expression

While R-loop-induced DNA damage is one explanation for genome instability in cells with defective splicing, a simpler explanation might simply be that some key transcripts are more sensitive to impaired splicing than others. Studies have highlighted that the efficiency of pre-mRNA splicing is dictated by not only the protein complexes present within the spliceosome, but also features of the transcript itself [65]. Splicing signals are highly important for canonical splice isoforms, as there are numerous cryptic splice sites that are present in almost every pre-mRNA molecule [66]. Adding to this complexity, several RNA splicing factors, in particular the Serine-Arginine (SR) proteins autoregulate their own expression by promoting intron retention or alternative exon events within their own mRNA [67, 68]. Due to the distinct function of each splicing factor, it has been proposed that the contribution of splicing factor mutations to the state of disease could be caused by 1) aberrant splicing of a specific gene or a subset of genes, or 2) global deregulation of splicing, which would lead to accumulation of aberrant splice isoforms and suboptimal levels of canonical transcripts [24] . Thus, genome instability could simply arise due to selective defects in splicing of key transcripts encoding canonical genome maintenance proteins.

Indeed, there have been several instances of identification of genes or subsets of genes that may be selectively influenced by defective splicing in certain contexts, leading to genome instability. It was previously shown that in yeast, many splicing genes when mutated lead to genetic instability [53]. Additionally, as I will show in Chapter 2, for certain splicing factors like Hsh155 (human homolog SF3B1) and Snu114 (human homolog EFTUD2) , the cause of genetic instability is due to suboptimal splicing of *TUB1*, the protein product of which, α-tubulin is critical in forming the mitotic spindle. This is consistent with previous research which shows that defective splicing leads to suboptimal expression of *TUB1*, ultimately resulting in cell cycle arrest phenotypes and loss of chromosome transmission fidelity [14, 69, 70]. Other research found that inefficient splicing of *TAF14* transcripts in specific Cdc40/Prp17 yeast mutants strains led to cell cycle arrest [71]. Several other groups have shown that splicing factors are essential for chromosome cohesion maintenance, and therefore proper chromosome segregation in mammalian systems [72-75]. The results of these studies indicate that when various components of the spliceosome were knocked down, this resulted in intron retention in the CDCA5 transcript,

and thus decreased levels of the protein sororin, which is essential for sister chromatid cohesion. As noted above, depletion of the splicing regulator SLU7 was found to create a truncated form of SRSF3, leading to R-loops. However, in the same study sororin splicing and sister chromatid cohesion was also impaired [58]. Thus, single splicing factor disruption can lead to genome instability through multiple parallel mechanisms [58, 76] (Chapter 2). Taken together, these studies indicate that in specific contexts, disruption of splicing factors can lead to expression changes of a gene that has a direct, measurable impact on genome maintenance. Why sororin transcripts, and not those of other chromosome cohesion genes, are so sensitive to splicing perturbations remains unclear.

While it may seem remarkable that a splicing factor mutation can lead to a phenotype via downregulation of a specific transcript, what is more commonly found in human diseases and cell lines is that splicing factor mutations lead to global deregulation of splicing, affecting the expression of a range of genes with functions in various cellular processes. How the aberrant expression of this broader portion of the transcriptome relates to genome maintenance and human disease is more difficult to interpret. Nevertheless, by identifying pathways involving factors that are differentially spliced due to disease associated splicing factor mutations, many groups have provided insight into how genome maintenance may be impacted. For example, Darman *et al* found that SF3B1 cancer hotspot mutations are gain of function mutations that induce aberrant 3' splice site selection [77]. The consequence of the SF3B1 hotspot mutations is that around half of the aberrant mRNAs produced are putative targets of nonsense mediated decay and thus expression of the canonical splice isoforms are globally downregulated [77]. The authors note that while aberrantly spliced mRNAs are common across different cancers, there are certain transcripts that appear to be cancer-specific. Importantly, the impact of these expression changes creates defects in DNA damage signaling which destabilize the genome [78, 79]. Work by Wang *et al* identified that in SF3B1-mutated CLL patient samples, several cancer-related pathways were dysregulated as a result of altered expression of transcripts including *TERC* and *TERT*, which encode the RNA and reverse transcriptase components of telomerase, respectively [79]. The authors also use a leukemia cell line model to show that altered expression of *TERC* and *TERT* was correlated with higher telomerase activity, which is interesting because activation of telomerase is among the most common alterations in cancer, since it enables replicative immortality by maintaining chromosome ends [80]. In parallel, work by Ilagan *et al* found that

U2AF1 mutations led to differential splicing of hundreds of genes including *DNMT3B, H2AFY, ATR* and *FANCA,* and *CASP8,* genes that have roles in DNA methylation, X chromosome inactivation, the DNA damage response, and apoptosis, respectively [81]. The authors speculate that due to the wide range of non-overlapping cellular functions of the targeted transcripts, it is likely that the splicing factor mutations deregulate many different genes that ultimately impact various aspects of cell physiology, including genome maintenance pathways.

Other indirect regulators of global splicing are also implicated in genome stability. For example, loss of PRMT5, an arginine methyltransferase with many substrates including the spliceosome, deregulates the splicing of hundreds of genes with an enrichment for chromatin modifiers [34]. The authors go on to show that aberrant splicing of the histone acetyltransferase *TIP60* is a driver of DNA damage [34]. Another example is the transcription-associated kinase CDK12, whose disruption causes global effects on alternative splicing, in particular alternative last exon splicing, in a gene and cell type specific manner [82]. CDK12 is of particular interest here because mutations in this kinase have been shown to result in deregulation of expression of DNA damage response genes [83] and generation of a unique signature of mutations associated with tandem duplications in prostate cancer [84]. How this genome instability might be related to the splicing function of CDK12 is currently unclear [82]. Thus, global changes in splicing can also lead to different biological consequences, including genome instability, and more subtle regulation of sets of transcripts by this mechanism can likely also give rise to instability.

1.5 Direct links to DNA damage response proteins and splicing factors

Splicing defects can have *cis*-effects during transcription, leading to R-loops at sites of aberrant splicing, or they can deregulate gene expression, indirectly depleting genome maintenance factors. However, hints of more direct interactions with DNA repair machinery exist for some splicing factors. For example, BRCA1 functions directly in the regulation of mRNA splicing in response to DNA damage through protein interactions with BCLAF1, which then forms a complex with splicing factors such as PRP8, U2AF1, U2AF2, and SF3B1 [85]. This leads to upregulation of genes involved in the DNA damage response such as *ATRIP*, *BACH1*, and *EXO1* [85]. The authors propose that chromatin-bound BRCA1 potentially recruits the splicing machinery to a subset of gene promoters required for efficient response to DNA

damage, thus acting as a tumour suppressor complex. Another study found that BCLAF1 and the RNA processing factor THRAP3 selectively regulate the expression of a specific subset of DNA damage response genes by promoting splicing of the transcripts (by BCLAF1) or mRNA export (by THRAP3) [86] . Another example is the moonlighting function of the splicing factor, and ubiquitin ligase, Prp19 which interacts directly with the ATR (Ataxia telangiectasia and Rad3 related) kinase, a major cellular sensor of single stranded DNA (ssDNA) [87, 88]. Prp19 mediated ubiquitylation of RPA (replication protein A) was shown to enforce ATR recruitment and activation at ssDNA, promoting genome stability independent of its splicing function [87]. Thus, direct physical engagement with DNA damage repair machinery, separable from global splicing defects, may need to be considered when interpreting genome instability phenotypes associated with disruption of certain splicing factors.

1.6 DNA double strand break repair

As a way of organizing a solid foundation for understanding a complex disease like cancer, Hanahan and Weinberg proposed that all cancers share six common traits, or hallmarks, that dictate how normal cells transform to cancer cells: 1) sustained proliferative signaling (cells stimulate their own growth); 2) evading growth suppressors (cells resist inhibitory signals); 3) activating invasion/metastasis (cells invade distant sites); 4) replicative immortality (cells multiply indefinitely); 5) induction of angiogenesis (cells stimulate blood vessel growth to supply nutrients to tumour); and 6) evading apoptosis (cells resist programmed cell death) [4]. Underlying the acquisition of these hallmarks is genome instability, which leads to accumulation of DNA alterations that can then generate genetic diversity and promote clonal evolution of cancer cells. DNA damage is a significant factor that causes genome instability, and one of the most toxic DNA damage lesions are DNA double strand breaks (DSBs), which can trigger cell death, and when not repaired properly can cause DNA structural rearrangements [89]. Once a DSB occurs within a cell, under normal circumstances most DSBs are quickly sensed and resolved by the DNA damage response (DDR), a series of rapid signaling cascades that serves to fix DSBs [90]. Although specific responses differ for different types of DNA lesions, in general the DDR involves a general program consisting of DNA damage sensors, recruitment of mediators, signal amplification, transducers and effectors, and ultimately diverse cellular responses [91].

Cells have evolved two major mechanisms to repair DNA DSBs: nonhomologous end joining (NHEJ) and homologous recombination (HR). NHEJ is an error prone process of reattaching broken DNA ends, and is active throughout the cell cycle [92]. HR in comparison involves exchange of genetic information between homologous DNA sequences, and is therefore generally active in S and G2 phases of the cell cycle when the sister chromatid is available as a repair template, resulting in accurate repair [93]. Repair of DSBs produced by replication fork collapse and fork restart are primarily mediated by HR proteins [94-96], although NHEJ proteins can play a lesser role [94, 97, 98]. Replication fork collapse produces one ended double strand breaks, also described as a "double strand end" (DSE), and it has been suggested that because forks are often widely spaced, NHEJ-mediated response to DSEs could result in large scale deletions or genomic rearrangements [99]. Therefore the balance between HR and NHEJ-mediated response to replication stress is important for maintaining genetic integrity. A major determinant of HR/NHEJ pathway choice is regulation of DNA end resection, which is initiated by the MRN (MRE11-RAD50-NBS1) complex [100]. 53BP1 is a key component in DSB repair and signaling in mammalian cells, and is also a determinant of DSB repair choice [101]. During G1, 53BP1 promotes NHEJ by antagonizing DNA end resection, which is essential for repair by HR [102-104]. HR depends on resected DNA around the double strand break to produce a 3' single strand overhang, that eventually becomes coated with RAD51 recombinase to form a nucleoprotein filament that is required for homology search [105]. During S/G2 phase, BRCA1 and its interacting partners counteract 53BP1 to promote DNA end resection and HR mediated repair [104].

1.7 Replication stress

Cellular defects that interfere with replication and subsequently slow down replication forks is termed replication stress. Persistent replication stress can completely stall replication forks and these can be processed to DNA DSBs, or simply fail to restart leading to under-replicated DNA [106]. Both of these phenomena can promote genetic instability. In general, when replication stress is sensed by the cell, this leads to activation of the replication checkpoint, cell cycle arrest, and the DDR to promote stabilization of the replication fork and to ensure replication completion in dividing cells [90]. Replication stress usually results in accumulation of stretches of exposed

single stranded DNA (ssDNA), which is formed when the replicative helicase MCM continues to unwind DNA despite polymerase stalling [107]. Stretches of ssDNA are coated by replication protein A (RPA), and this together with the stalled newly replicated double stranded DNA serves as a signaling platform to recruit DNA damage response proteins like the serine/threonine protein kinases ataxia telangiectasia mutated (ATM) and ataxia telangiectasia mutated Rad3 related (ATR) [108, 109]. The subsequent signaling cascades include phosphorylation of checkpoint kinase 2 (CHK2) [110, 111] and checkpoint kinase 1 (CHK1) [112-114], respectively. Simplistically, ATM responds to DNA double strand breaks while ATR is activated by the presence of RPA-coated single strand DNA due to replication stress [108, 115]. The ATR-CHK1 signaling cascade leads to cell cycle arrest and replication fork stabilization/restart, although the exact mechanisms by which this occurs are complex [116].

1.8 Thesis objectives

A growing body of work has provided evidence that splicing factor mutations play a role in genome maintenance across species. The observation that splicing factors are frequently mutated in a broad range of cancers suggests spliceosome dysfunction is a driver of disease progression. While many splicing factors that when mutated in yeast and human cells have been ascribed a role in genome instability, the mechanisms by which this occurs is not well understood. The goal of my thesis is to understand how splicing factor mutations in yeast and human cells contribute to genetic instability.

Thesis hypothesis: Splicing factor mutations in yeast and mammalian cells cause genetic instability by fluctuations in gene expression and aberrant R-loop accumulation

The objective of Chapter 2 was to determine the cause of genetic instability in select splicing factor mutants in yeast. This was accomplished by characterizing chromosome instability phenotypes, assessing R-loop accumulation, measuring cell cycle dynamics, and using a synthetic genetic array screen to identify positive or genetic interactors of a splicing factor mutant. The outcome of this work showed that select splicing factor mutants in yeast cause genetic instability by suboptimal splicing of a gene important for genome maintenance, and aberrant R-loop accumulation.

The objective of Chapter 3 was to determine whether a common cancer-associated SF3B1 mutation (H662Q) causes genetic instability, and to understand the mechanism by which this occurs. This was accomplished by characterizing the homolog in yeast bearing the cancer-associated mutations, and in mammalian cells measuring DNA damage and replication stress phenotypes, and assessing replication dynamics. The outcome of this work showed that SF3B1 H662Q mutations cause R-loop mediated replication stress, and changes in gene expression that potentially shifts the way these cells respond to replication stress and associated DNA damage.

The objective of Chapter 4 was to summarize the data, and to discuss the observations from Chapters 2 and 3 in the context of what is known about splicing factor mutations and genetic instability.

Chapter 2: Selective defects in gene expression control genome instability in yeast splicing mutants

Note:

This chapter is based on a published article in *Molecular Biology of the Cell* in 2019 as cited here:

Tam AS, Sihota TS, Milbury KL, Mathew V, Zhang A, **Stirling PC**. Selective defects in gene expression control genome instability in yeast splicing mutants. *Mol Biol Cell* 30: 191-200, 2019.

The contributions of other authors to any work shown in the thesis are clearly indicated in the relevant Figure Legends.

2.1 Background

Genome stability maintenance is a complex process involving the coordination of essentially all DNA transactions, including transcription, chromatin state, DNA replication, DNA repair and mitosis. Regulation of genome stability is critical to prevent cancer [4]. While screens across model organisms and human cells have implicated numerous genes as regulators of genome maintenance, in many cases we do not understand the mechanism of action [52, 53].

Defects in RNA processing have been implicated in genome instability across species, and in both human cancer and repeat-expansion diseases [117]. Indeed, RNA splicing factors are frequently mutated in cancers where they shift gene expression landscapes, favoring oncogenesis [77, 78, 118]. Previous work has suggested that loss of splicing factors like ASF/SF2 [51], or treatment with splicing inhibitors [55], cause the accumulation of DNA:RNA hybrids in genomic DNA. These three-stranded R-loop structures contribute to genome instability by exposing single-stranded DNA (ssDNA) and by blocking replication forks, causing replication stress induced genome instability [119, 120]. Indeed, splicing has been ascribed a protective role in genome maintenance in yeast [59]. More recently, cancer-associated mutations in splicing

factors such as SRSF2 and U2AF1 have been attributed roles in R-loop prevention related to genome instability in myelodysplastic syndromes [64]. These data compound other observations indicating that transcription termination [121], 3'-end processing [122], and mRNA packaging and export mutants [123] together create a robust R-loop prevention system.

Other data have suggested that splicing factor disruption, such as loss of CDK12, causes changes in gene expression, which reduce the activity of canonical genome maintenance factors [83]. Indeed, cancer-associated mutations in splicing factor SF3B1 have been shown to disrupt splicing of DNA damage response related transcripts in myelodysplastic syndromes [78]. There is little evidence for gene expression changes driving genome instability in yeast splicing mutants, but cell cycle delays, possibly linked to genome instability, have been previously connected to defective tubulin mRNA splicing in some mutants [69]. Whether altered gene expression, R-loops, or both contribute to genome instability in splicing mutants is unclear.

It has previously been identified that many spliceosome components when disrupted in yeast leads to genome instability [53]. Subsequent work found that only a handful of these genome destabilizing splicing mutants caused detectable increases in R-loop levels [54]. Importantly, only 5% of yeast genes encode introns reducing the complexity of interpreting specific splicing changes as drivers of genome instability [124]. Here my goal was to test the contribution of R-loops versus gene expression changes in splicing-loss induced genome instability in yeast. While I observed evidence of R-loop induced DNA damage in a mutant allele of *SNU114*, all splicing mutants appear to cause aberrant splicing of the α-tubulin transcript from the *TUB1* gene. A mitotic defect arising from Tub1 depletion is therefore a common driver of chromosome loss in yeast splicing mutants.

2.2 Materials and methods

2.2.1 Yeast strains, growth and CIN assays

All yeast strains were in the s288c background, and were grown under standard conditions in the indicated media and growth temperature. To assess benomyl sensitivity, strains were compared between YPD+0.2% DMSO (control) or YPD+15μg/mL benomyl (Sigma-Aldrich cat#45339). Growth curves were conducted in YPD using a Tecan M200 plate reader and were

compared using the area under the curve as previously described [125]. Briefly, logarithmic phase cultures were diluted to OD 0.05 in a 96-well plate and grown for 48 hours, with OD600 readings taken every 30 minutes.

The chromosome transmission fidelity (CTF) and a-like faker (ALF) assays were performed as described [53, 126]. In the ALF assay involving *tub1*Δ, *tub1*Δ mutant strain carrying a galactose-inducible *TUB1* plasmid was grown in either SC+2% galactose or 2% dextrose media for 24 hours, mixed with an excess of mating tester and plated on to SD media + 2% galactose or 2% dextrose at 30°C. The CTF assay measures inheritance of an artificial chromosome fragment as monitored by colony color to indicate chromosome instability – an induced *ade2-101* mutation pigments colonies red, and when supplemented with a chromosome fragment carrying suppressor mutation *SUP11*, the chromosome instability read-out was white (indicating no loss of fragment) or red (loss of fragment). The ALF assay measures the frequency of *MATα* locus loss in α-mating type haploid yeast, which indicates de-differentiation to a-mating type haploid yeast. Cells that have lost this locus were detected by selection for mated diploid yeast. For plasmid loss, strains carrying *pRS313::HIS3* were grown overnight in SC-histidine, then plated on YPD and allowed to form colonies without selection before replica plating onto SC-histidine. The frequency of colonies that could not grow on SC-histidine is reported as the plasmid loss rate. For all experiments significance of the differences was determined using Prism7 (GraphPad Software). For all experiments, sample means were compared with Fisher's exact test, Student's t-test or ANOVA for multiple comparisons as indicated.

2.2.2 Synthetic Genetic Array and validation

SGA screening was performed as described previously for *URA3*-marked query strains [53, 127] and scored using the Balony software package [128]. From raw colony scores, a cut-off of $p<0.05$ was used and an experiment-control score of $\geq|0.5|$ for list analysis and GO term enrichment using the Princeton generic GO term-finder (http://go.princeton.edu/cgi-bin/GOTermFinder) and visualized with REVIGO [129]. *snu114-60* and *yhc1-1* interactions are available at http://thecellmap.org [130]. For hit validation, fresh double mutant strains were made by tetrad dissection and tested in quantitative growth curves, or by spot dilution assays. Observed area under the curve for double mutants was compared to a multiplicative model of the predicted fitness based on the fitness of the two single mutants.

2.2.3 Image and cell cycle analysis

Imaging of budding index by differential interference contrast (DIC), or fluorescence of Rad52-GFP and GFP labeled chromosome III was conducted on a Leica DMi8 microscope using an HCX plan apochromat 1.4 NA oil immersion 100x lens. The images were captured at room temperature by an ORCA Flash 4.0 V2 camera (Hamamatsu Photonics), using MetaMorph Premier acquisition software (Molecular Devices). Scoring was done in ImageJ (National Institutes of Health). For Rad52 foci, all cells were scored as negative or positive for a focus. Imaging of LacO-LacI-GFP labeled chromosome III was performed as describe, using asynchronous yeast cultures [131]. For scoring, first unbudded cells were selected in the DIC channel, then scored for the presence of 1 or ≥2 LacI-GFP dots. Chromosome spreads were performed exactly as described; primary DNA-RNA Hybrid [S9.6] (Kerafast cat#ENH001); secondary Alexa Fluor® 568 goat anti-Mouse IgG (Invitrogen cat#A-11004) [132]. Flow cytometry was done using the BD FACSCalibur™ platform. Cells were arrested in G1 by α-factor arrest using 10 μg/mL α-factor (Cedarlane, cat#Y1001) for 2.5 hours, and samples were collected at 30min, 90min, and 150min after wash-out to account for more than one complete cell cycle. Cells were fixed with 70% ethanol and stained with 16 μg/mL propidium iodide (Sigma-Aldrich cat#287075). The proportion of cells in G1, S and G2/M cell cycles was quantified using FlowJo.

2.2.4 Splicing efficiency assay

Splicing assay protocol was performed as described [133]. Briefly, yeast cells were transformed with lithium acetate transformation. All measurements were taken with individual transformants in triplicate. Cells were struck as a patch on SC-leucine and then replica plated to glycerol-lactate–containing SC medium without leucine (GGL-leu). Cells from each patch were inoculated in liquid GGL-leu media for 2 h at 30°C and then were induced with final 2% galactose for 4 h. Cells carrying reporters were lysed and assayed for β-galactosidase assay using a Gal-Screen β-galactosidase reporter gene assay system for yeast or mammalian cells (Applied Biosystems) as per the manufacturer's instructions and read with a SpectraMax i3 (Molecular Devices). Relative light units were normalized to cell concentration as estimated by measuring OD600.

2.2.5 Recombination assays

To construct the recombination plasmids, pRS314GLB [134] was linearized with BglII, and ligated with *URA3* sequences with BglII sites [135]. Following lithium acetate transformation and plasmid isolation, presence of *URA3* was confirmed by Sanger sequencing. Direct repeat recombination assays were performed as described [125].

2.2.6 Western blot

Whole-cell extracts were prepared by 100% trichloroacetic acid extraction ($2x10^6$ cells per extraction). Lysates were separated on a 10% or 15% SDS-PAGE gel, transferred to 0.45μm PVDF membranes and probed with the following antibodies: anti-Tub1 (Invitrogen cat#32-2500) (1:500 dilution), anti-Pgk1 (Santa Cruz cat#sc-130335) (1:1000 dilution) as a loading control, Yra1 antibody (a gift from David Bentley, University of Colorado, Denver) (1:10000 dilution), Clb2 antibody (Santa Cruz Biotechnology cat#sc-6697, discontinued) (1:1000 dilution). ImageJ software was used to quantify protein bands [136].

2.2.7 RNA isolation, cDNA preparation, and reverse transcription–quantitative PCR analysis

Total RNA was isolated from 0.5-1 OD cell cultures shifted to 34°C for 3.5 h, using the yeast RiboPure RNA Purification kit (Ambion). 1 μg of cDNA was reverse transcribed using anchored-oligo(dT)18 primer and Transcriptor Reverse transcription (Roche). Reverse transcription–quantitative PCRs were performed and analyzed using SYBR green PCR Master Mix and a StepOnePlus Real-Time PCR system (Applied Biosystems). RQ values were normalized to an unspliced *SPT15* transcript and expressed relative to WT.

2.3 Results and discussion

2.3.1 Splicing factor mutations lead to chromosome instability

Previous screens have identified at least 25 splicing proteins that, when disrupted in yeast, lead to chromosomal instability (CIN) [53]. To begin to understand whether R-loops or other mechanisms drove genome instability, I conducted CIN assays in strains with mutations in each of the core snRNP complexes involved in establishing the splicing reaction [137]. Since each

spliceosomal snRNP is essential, I used temperature-sensitive (ts) alleles of *YHC1* (U1), *HSH155* (U2), and *SNU114* (U4/U6.U5), and used a splicing reporter assay as well as qPCR to measure splicing efficiency. The splicing reporter assay is a plasmid-based splicing efficiency assay where expression of *LacZ* protein product β-galactosidase was compared using chemiluminescence. Briefly, one plasmid with intronless *LacZ* was used to normalize expression levels, while another plasmid contains an intron in *LacZ*, resulting in an out of frame gene unless the intron is spliced out. Therefore, if splicing is functional, the intron is spliced out of *LacZ* and β-galactosidase is expressed. Each of the mutants exhibited strong splicing defects as measured with a *LacZ* splicing reporter (**Figure 2.1A**) or by directly measuring intron retention at an endogenously spliced gene, *RPL33B* (**Figure 2.1B**).

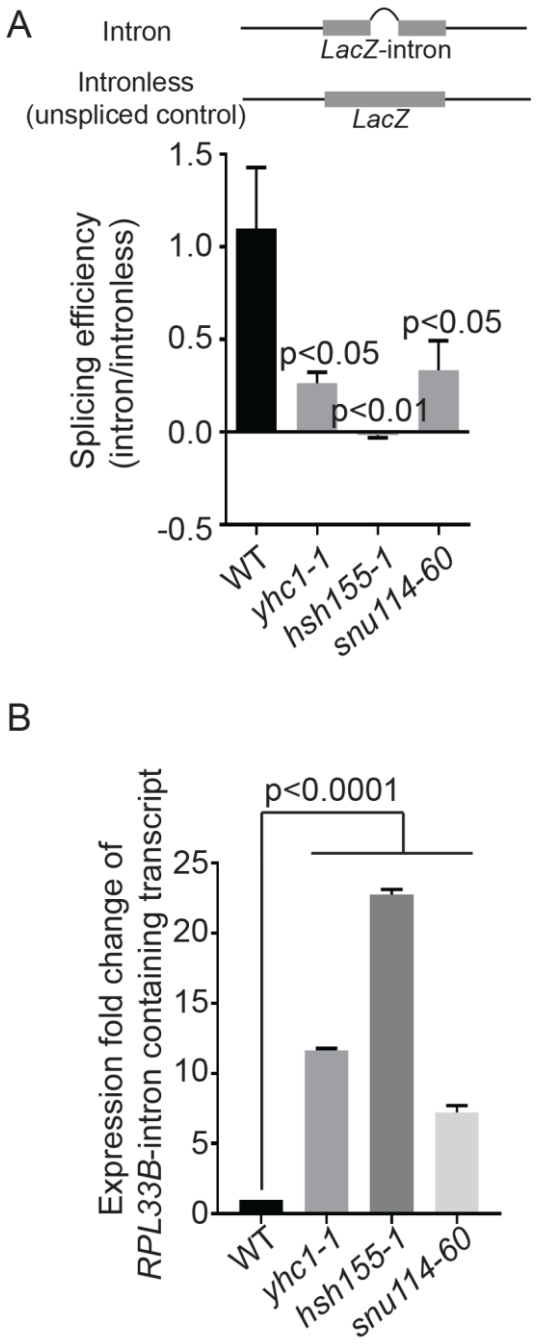

Figure 2. 1. Level of splicing deficiency in select yeast splicing factor mutants.

(A) Splicing efficiency of a LacZ reporter, values are relative to WT splicing level. Schematic of reporter constructs are presented above panel; Student's t-test (B) Quantification of *RPL33B* mRNA transcript levels from intron region in WT, *yhc1-1, hsh155-1* and *snu114-60* by RT-qPCR normalized to *SPT15* and relative to WT; p-values calculated from ΔΔCt levels (ANOVA). Mean values with S.E.M. error bars are shown, *n = 3*.

In all three splicing mutants, I observed increases in artificial chromosome loss by the chromosome transmission fidelity (ctf) assay, which measures loss of an artificial chromosome fragment, confirming previous findings [53] (**Figure 2.2A**). I next measured the stability of a CEN plasmid and found that splicing-defective cells have increased plasmid loss relative to WT (**Figure 2.2B**). A CEN plasmid is one that contains a yeast centromere, therefore these plasmids replicate like endogenous chromosomes and their copy number is regulated by functional mitotic spindle attachment to the centromere on the plasmid. Since these two assays rely on the loss of episomes, I tested the stability of endogenous chromosomes by monitoring a strain with an integrated LacO-array on chromosome III in cells expressing LacI-GFP, which creates a green fluorescent spot marking the chromosome. I studied this system in yeast haploid G1 cells, because if I observed more than one green fluorescent dot, this would indicate chromosome instability rather than the loss of chromosome cohesion or loss of the homologous chromosome. In haploid unbudded G1 cells, only a single GFP spot should be present, however, I observed that all splicing mutants tested showed an increased rate of gain of a LacI-GFP marked chromosome III, suggesting that a chromosome gain event has taken place (**Figure 2.2C**). Thus, multiple assays indicate splicing mutants led to a significant increase in chromosomal instability.

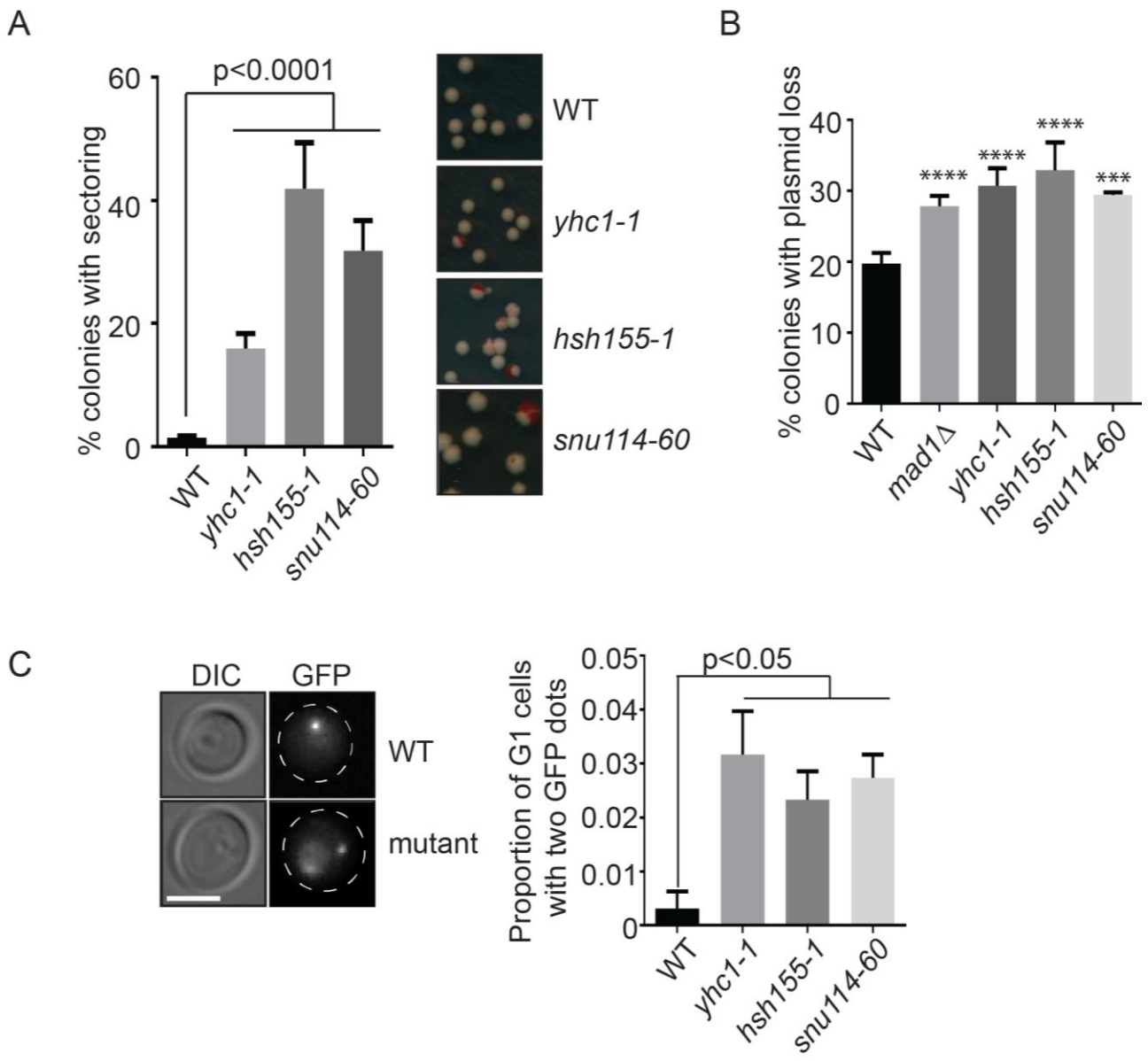

Figure 2. 2. Genome instability phenotypes of splicing mutants.

(A) CTF phenotypes indicated by the percent of sectored colonies. Right, representative images. (B) Plasmid loss frequency. ***p=0.0005; ****p<0.0001 (C) Frequency of unbudded cells with >1 GFP-marked LacO array. Left: representative images. Dashed lines indicate cell outlines, scale bar =2μm. Fisher's exact test was used to calculate statistical significance. For all figures where applicable mean values with S.E.M. error bars are shown, *n = 3*.

2.3.2 R-loop accumulation and associated DNA damage in *snu114-60*

The CIN phenotypes observed in splicing mutants could arise by several mechanisms including the formation of transcription coupled R-loops [51, 59]. To test this model I first performed chromosome spreads and used the S9.6 antibody to detect DNA:RNA hybrid levels in these spreads with immunofluorescence [132]. As reported previously, *yhc1-1* and *snu114-60* alleles showed higher levels of R-loop accumulation compared to WT, while the *hsh155-1* allele had no increase in R-loops (**Figure 2.3**) [54]. Ectopic expression of yeast RNaseH1 was used to verify S9.6 signal.

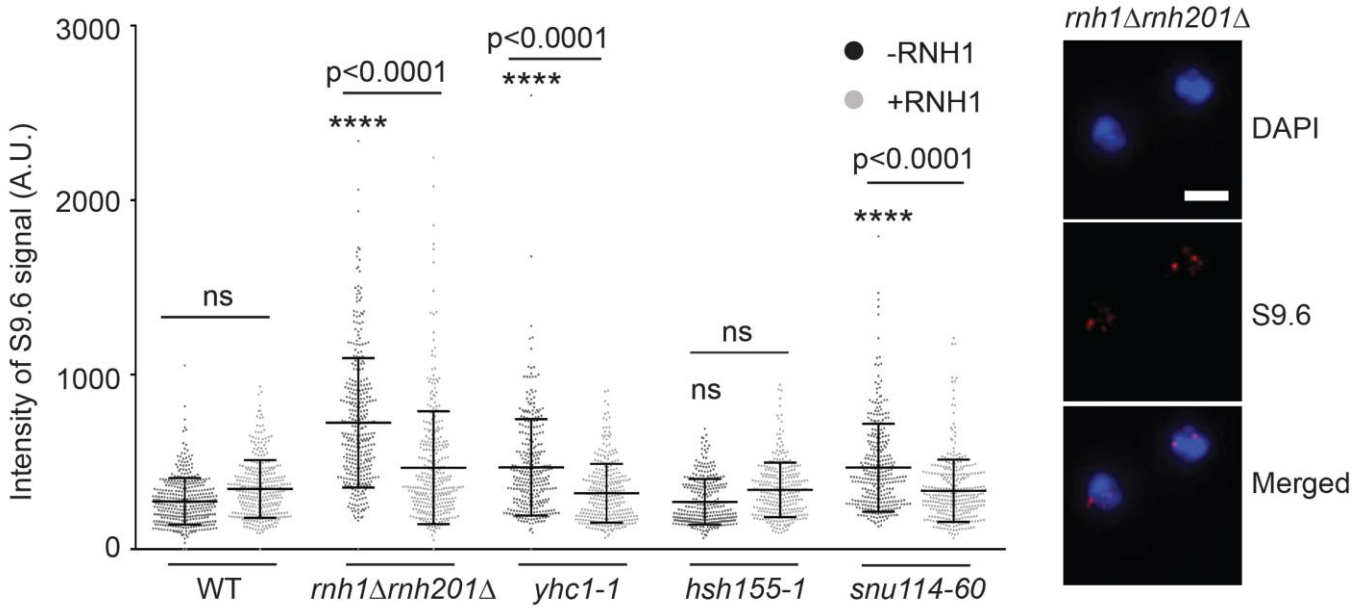

Figure 2. 3. R-loop detection in splicing mutants.

s9.6 antibody staining intensities in chromosome spreads. ****p<0.0001 relative to WT -RNH1. Right, representative images, scale bar = 2μm; one-way ANOVA.

I elected to pursue comparison of *snu114-60* and *hsh155-1* since both alleles disrupt splicing and exhibit strong CIN phenotypes but have opposing effects on R-loop levels. In addition *yhc1-1* cells are very sick across temperatures, making some assays technically unreliable. Consistent with the lack of R-loop accumulation in *hsh155-1*, when I analyzed Rad52-GFP foci, a marker of DNA damage repair that should increase if R-loops are driving genome instability through DNA breaks, I found that only *snu114-60* increased Rad52 foci number (**Figure 2.4A**). Importantly,

this DNA damage phenotype could be partially suppressed by ectopic expression of yeast RNaseH1 (*RNH1*), similar to a *rnh1*Δ*rnh201*Δ control strain (**Figure 2.4B**).

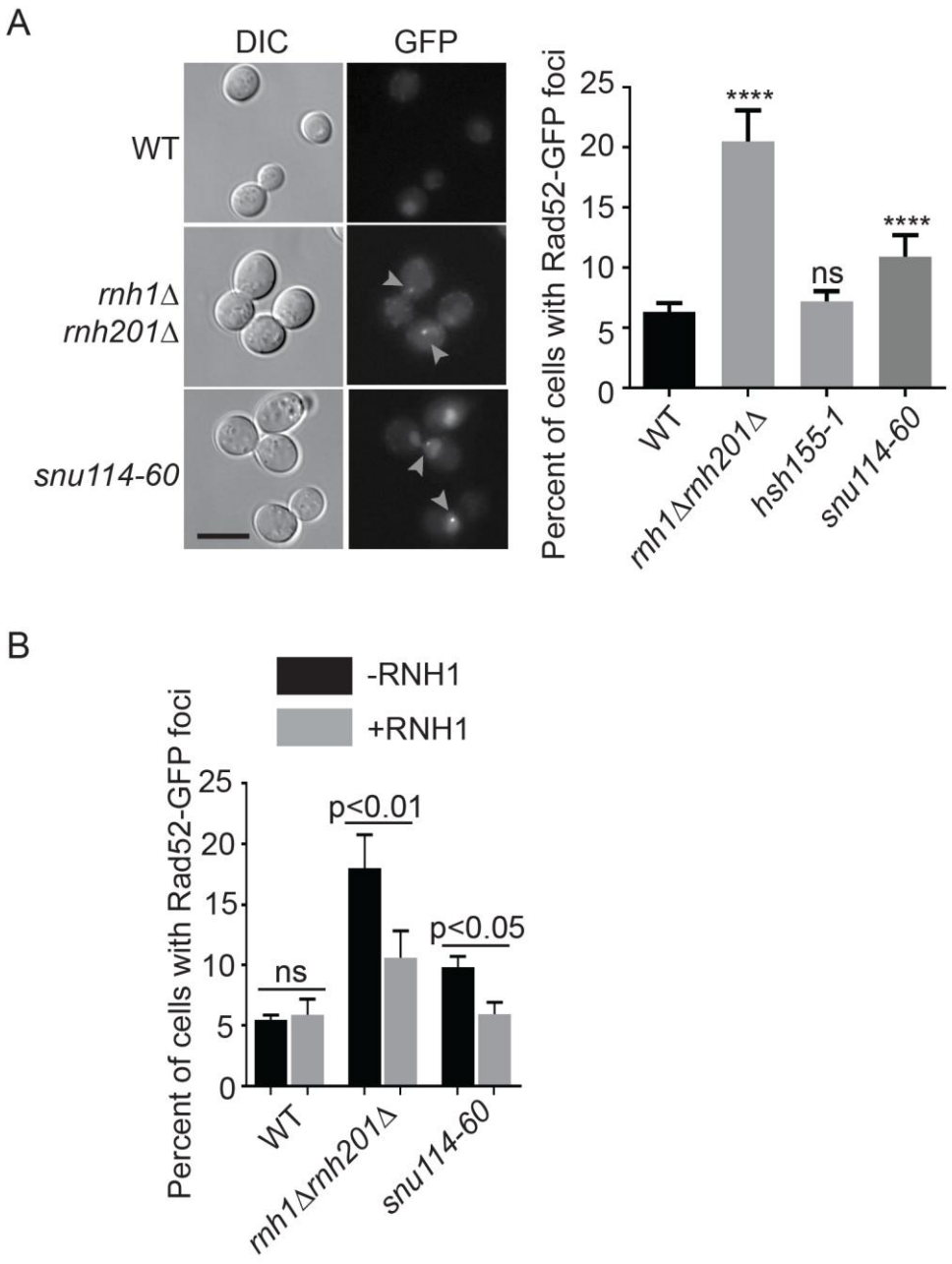

Figure 2. 4. R-loop associated genome instability in *snu114-60*.

(A) Percent of cells with Rad52-GFP foci (highlighted with grey arrows in representative images on left), scale bar = 5μm; Fisher's exact test ****p<0.0001. (B) Suppression of Rad52-GFP foci by ectopic yeast *RNH1* expression; one-way ANOVA. Mean values with S.E.M. error bars are shown, *n = 3*.

To further test phenotypes known to correlate with aberrant R-loop levels, I used a plasmid-based direct repeat recombination system to test for hyper-recombination. These systems monitor the frequency of DNA breaks occurring in a transcribed locus that is repaired by recombination to reconstitute a selectable marker. As expected, only *snu114-60* showed a significant increase in recombination compared to WT (**Figure 2.5A**, *mft1*Δ is a hyper-recombination positive control [125]). I also tested the impact of *snu114-60* on recombination at an integrated recombination substrate flanking a replication origin, to assess the potential for replication-transcription conflicts that give rise to DNA damage and promote recombination. *snu114-60* also causes hyper-recombination in this chromosomal context in a manner that was suppressed by *RNH1* (**Figure 2.5B**). These data show that, while splicing alleles like *snu114-60* contribute to R-loop accumulation and associated DNA damage, R-loops are not a unifying mechanism of genome instability across all of the various spliceosomal snRNP mutants [54].

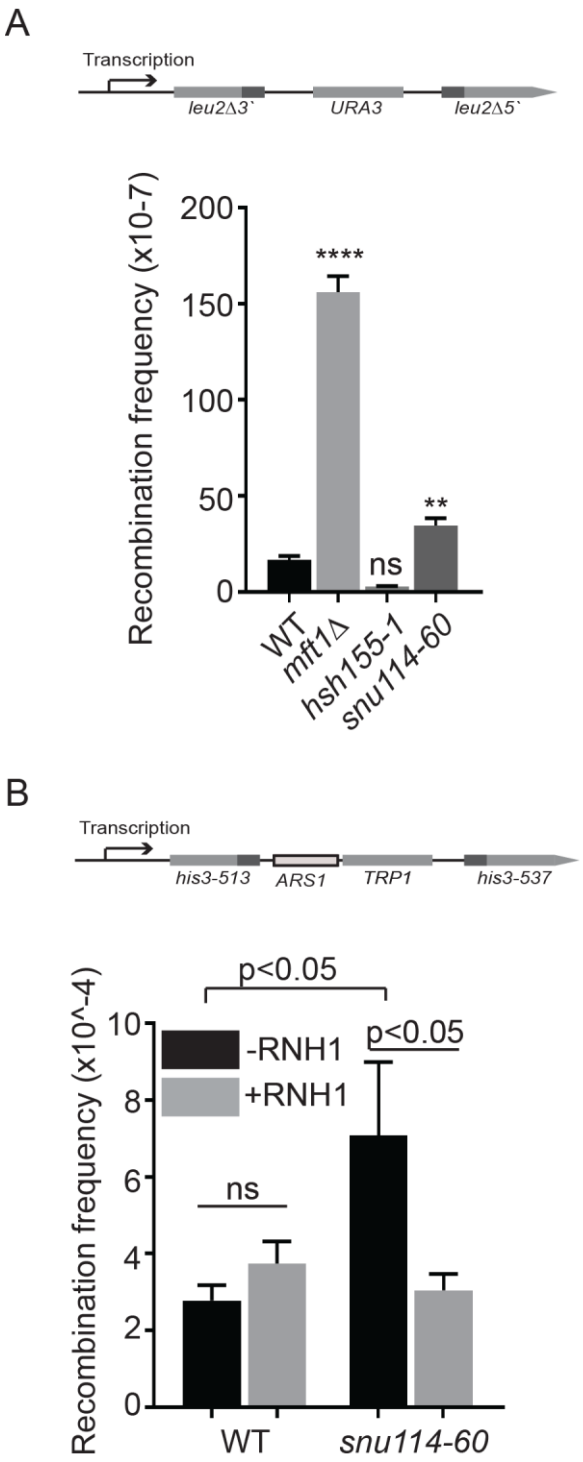

Figure 2. 5. R-loop associated hyper-recombination in *snu114-60*.

(A) Frequency of direct repeat recombination on the indicated plasmid; Student's t-test **p<0.01; ****p<0.0001. (B) Frequency of direct repeat recombination on the integrated genomic reporter; one-way ANOVA. Reporter construct schematics are presented above each panel. Mean values with S.E.M. error bars are shown, *n = 3*.

It was puzzling that *snu114-60* led to increased recombination in reporters that do not obviously contain canonical introns, and thus presumably do not recruit the spliceosome in either WT or *snu114-60* cells, although we do not directly measure this here. This raised the possibility that the expression of another R-loop regulator was sensitive to disruption of Snu114 activity and that the *snu114-60* mutation selectively depleted this factor. One candidate gene for this phenotype is *YRA1* which encodes an RNA export protein whose function has been linked to transcription-associated recombination and R-loop formation [17, 138, 139]. Measuring mRNA expression levels and splicing of *YRA1* by RT-qPCR indicated that *YRA1* is overexpressed in both *hsh155-1* and *snu114-60* relative to WT control, and that both alleles cause intron retention of *YRA1* transcript (**Figure 2.6A**). Both splicing mutants lead to lower levels of Yra1 protein by western blot relative to WT control (**Figure 2.6B**), consistent with previous work showing *YRA1* intron downregulating Yra1 expression in a splicing-dependent manner [140]. While Yra1 loss could have caused R-loop accumulation, the similar amount of Yra1 depletion in *hsh155-1* and *snu114-60* suggests this is not likely to be the case. mRNA export is known to regulate R-loop-associated genome instability [19] and it is notable that several RNA export proteins are encoded by transcripts with complex splicing behaviour. For example, *DBP2,* which encodes an RNA helicase and binding partner of Yra1, has the largest intron and its levels are also autoregulated by its intron [141]. Mtr2, another mRNA transport regulatory protein, is encoded by one of only a few *S. cerevisiae* genes to exhibit alternative splice isoforms [142]. Thus, while Yra1 protein levels alone are insufficient to explain R-loop driven instability in *snu114-60*, it is possible that these other factors play a role. This awaits a more systematic study of proteome changes amongst R-loop regulators in splicing mutants.

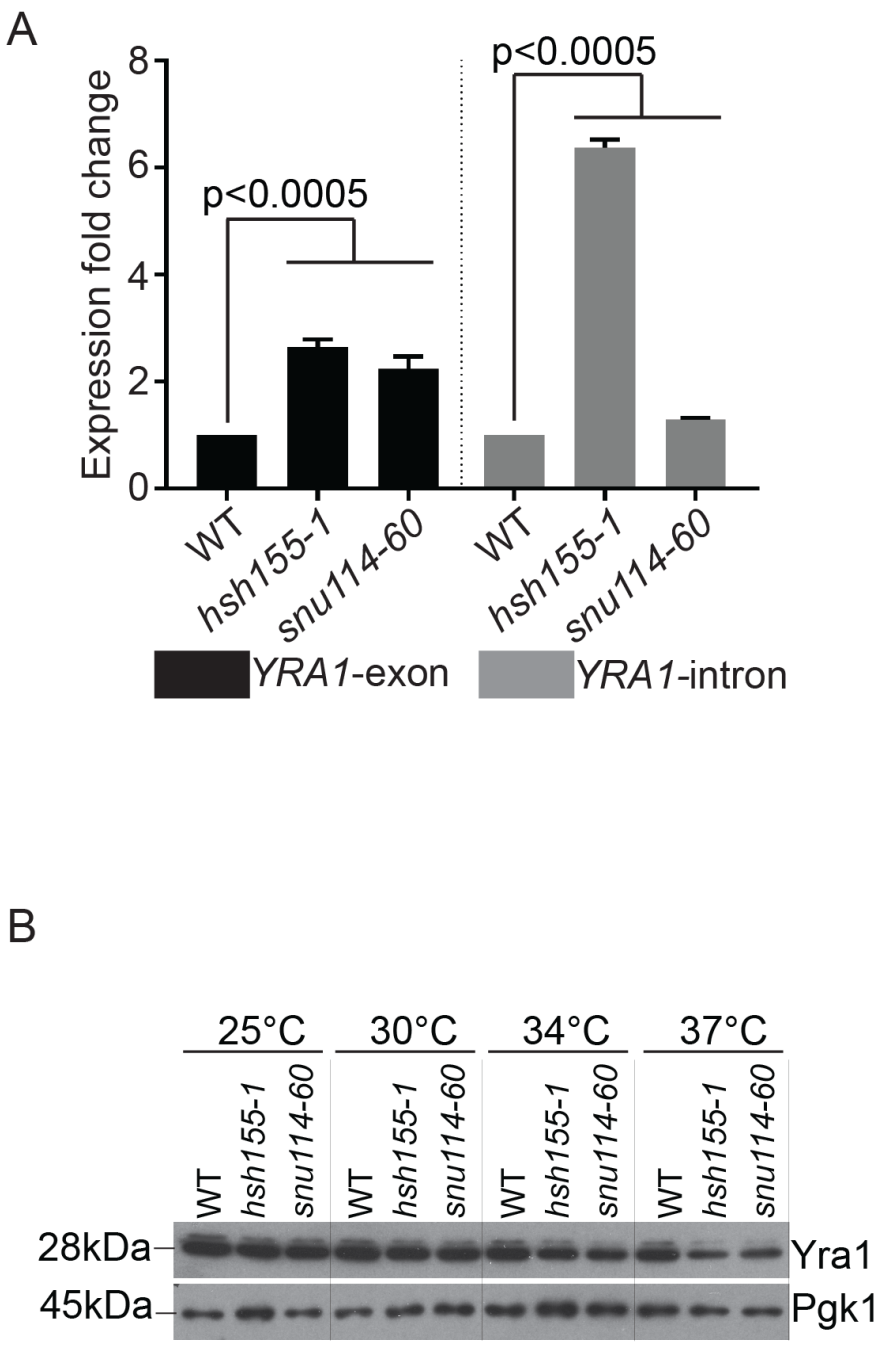

Figure 2. 6. Expression of RNA export protein YRA1 in splicing factor mutants.

(A) Quantification of *YRA1* mRNA transcript levels from exon region in WT, *hsh155-1* and *snu114-60* (left) and *YRA1* intron region (right) by RT-qPCR normalized to *SPT15* and relative to WT. p-values calculated from ΔΔCt levels (ANOVA). Mean values with S.E.M. error bars are shown, *n = 3*. (B) Representative western blot of Yra1 protein level in the indicated strains and temperatures.

2.3.3 Genetic interaction profiling reveals mitotic defects in hsh155-1

Knowing that R-loops likely do not account for genome instability at least in mutants of Hsh155 [54], I sought to understand common mechanisms. Mutations in *HSH155* exhibited strong CIN phenotypes, but showed no evidence of increased DNA damage or R-loops (**Section 2.3.2**). I hypothesized that if a common mechanism of CIN existed for splicing mutants, it would be at play in *hsh155-1* alleles. To determine this function, a synthetic genetic array (SGA) screen using *hsh155-1* as a query strain was performed to identify positive or negative genetic interactors of this mutant. This screen identified 102 negative and 103 positive genetic interaction candidates (**Appendix 1**). A greater than expected number of essential intron-containing genes were negative interactors, consistent with a splicing defect enhancing phenotypes of these mutants (**Appendix 1**). The observed positive interactions with proteasome subunits or translational apparatus could reflect stabilization of mutant Hsh155 protein leading to healthier cells (**Appendix 1**) [143].

Analysis of gene ontology (GO) terms among negative genetic interactions highlighted expected groups such as *spliceosomal complex assembly* (13.6 fold enriched) and *mRNA processing* (5.5 fold) (**Appendix 2**). Other potentially surprising GO terms such as *retrograde vesicle-mediated transport, Golgi to ER* (11.6 fold) and *Golgi-associated vesicle* (7.4 fold) (**Appendix 2**) can be explained by the preponderance of intron-containing genes in this pathway (e.g. *SNC1*, *BET1*, *SEC27*, *SFT1*, *SAR1, YIP3* all encode introns which could cause a vesicle trafficking defect in splicing mutants). Enrichment for cytoskeletal processes among negative genetic interaction partners of *hsh155-1* was also seen (**Appendices 1 and 2**) and was expected based on the enrichment of intron-containing genes in this pathway (e.g. *ACT1*, *TUB1*, *MCM21*, *COF1*, *GIM5*, *CIN2*, *TUB3*, *DYN2* encode introns) [124]. The clearest direct connection to chromosome segregation also came from this analysis of GO terms in this group. Terms like *attachment of spindle microtubules to kinetochore* were highly enriched (19.5 fold) (**Appendix 2**). GO term enrichments were visualized using REViGO [129], which is a Web server that can summarize long lists of gene ontology terms by removing redundant terms in the submitted list. REViGO highlighted *attachment of spindle microtubules to kinetochore* with processes like *microtubule polymerization*, and *mitotic sister chromatid biorientation* and components like the *Mis12/MIND type complex* of the kinetochore (**Figure 2.7**).

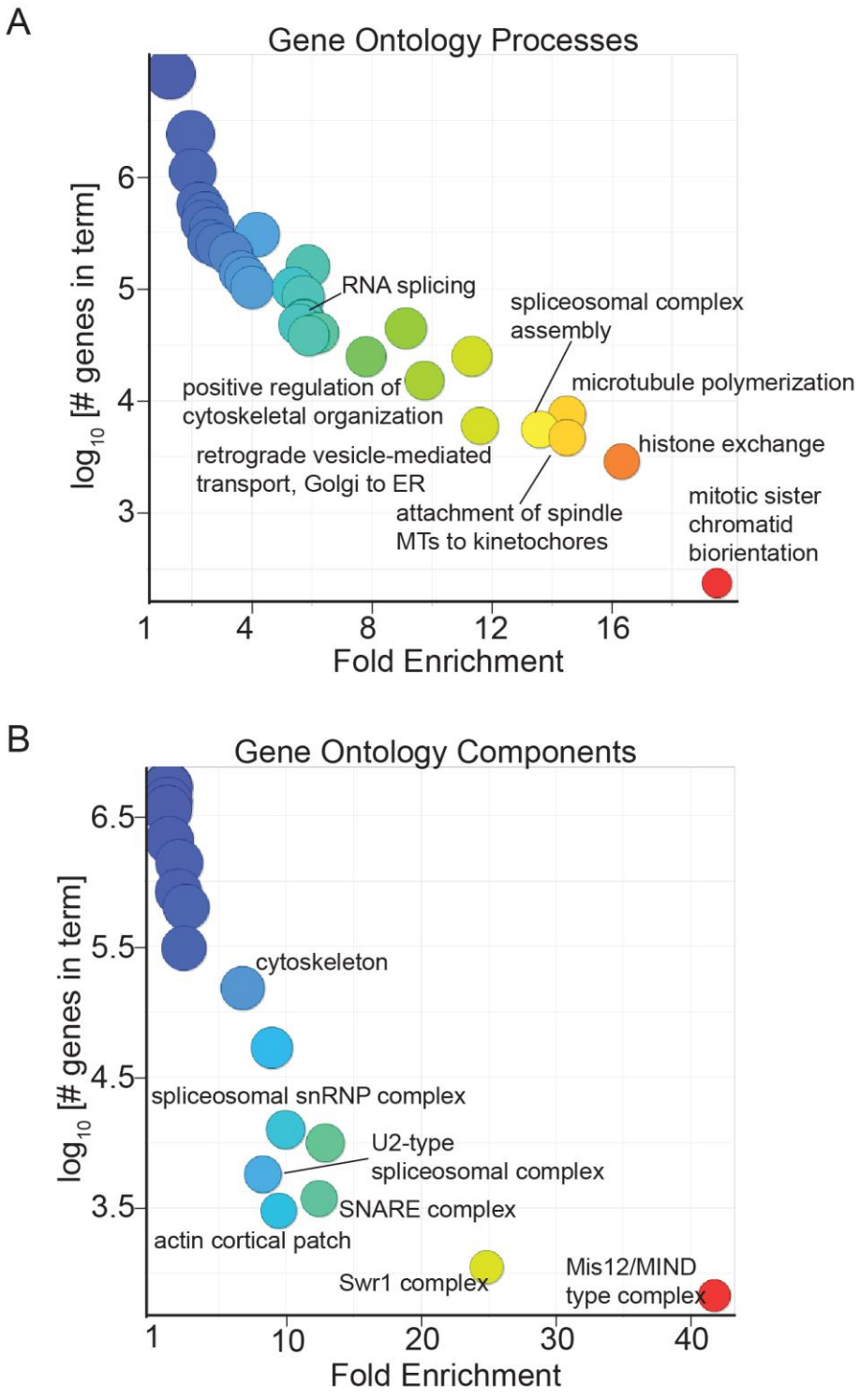

Figure 2. 7. Genetic interaction network of *hsh155-1*.

(A) GO biological process and (B) cellular component enrichments for *hsh155-1* negative interactions (**Appendices 1 and 2**). Warmer colors associate with higher enrichment. SGA, GO biological process and cellular component enrichment analysis were performed by P.C. Stirling.

I used spot dilutions of the various strains to validate the hits from the SGA screen to ensure these were not false positives. This assay validated that *hsh155-1* has negative interactions with mitotic genes like cohesin (*MCD1*), core kinetochore subunits (*MIF2*) and spindle regulators (*STU1*) (**Figure 2.8A**). I further used growth curves to identify subtle changes in growth in the *hsh155-1* double mutants (**Figure 2.8B**). Growth curves were performed at 30°C, the temperature used for the SGA screen. Analysis of published SGA profiles for *snu114-60* and other splicing factors (www.thecellmap.org) also revealed genetic dependence on a functional mitotic apparatus [127], which was consistent with the *hsh155-1* SGA results.

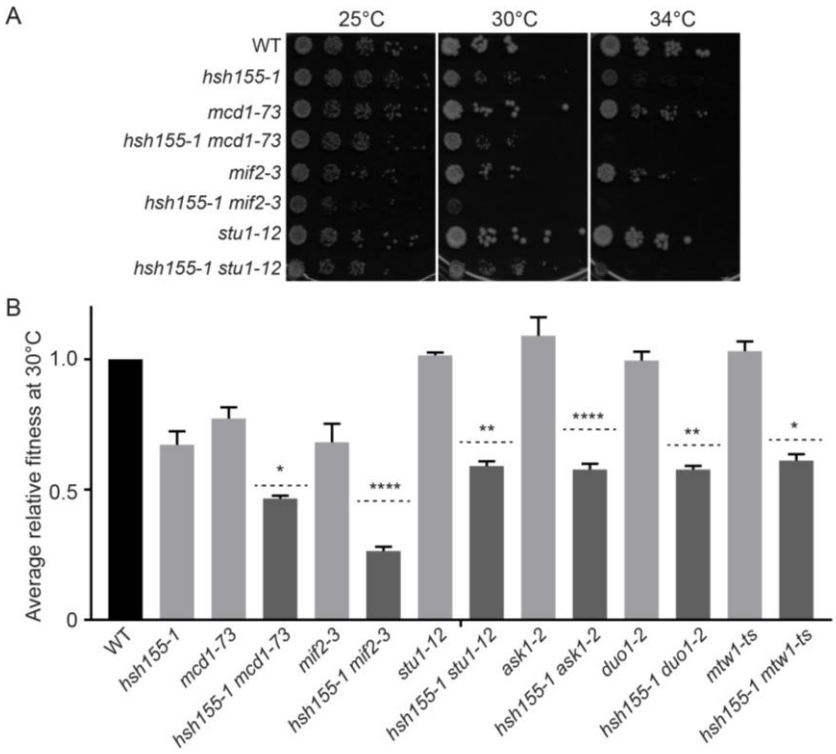

Figure 2. 8. Validation of SGA negative genetic interaction hits.

(A) Validation of negative genetic interactions by spot dilution assays on YPD plates at the indicated temperatures. (B) Validation of negative genetic interactions by growth curves. Light grey bars: fitness of single mutants; dashed lines: calculated expected fitness of double mutants using multiplicative model; dark grey bars: observed fitness of double mutants. Two-way ANOVA was used to calculate statistical significance between observed and expected fitness of double mutants. *p<0.05 **p<0.01 ****p<0.0001. *hsh155-1* double mutant strains were made by A. Zhang and V. Mathew.

2.3.4 *SNU114* and *HSH155* mutants have mitotic defects

To test potential mitotic defects, I first measured the cell cycle distribution of cells by budding index. Budding index refers to the fraction of budded cells in log phase yeast cultures, and is determined by manual counting of the size of the daughter cell (bud). After a shift to a non-permissive temperature of 37°C, a small but significant proportion of the splicing mutant cells accumulated as large-budded G2/M cells, indicating a potential mitotic delay and defect (**Figure 2.9**).

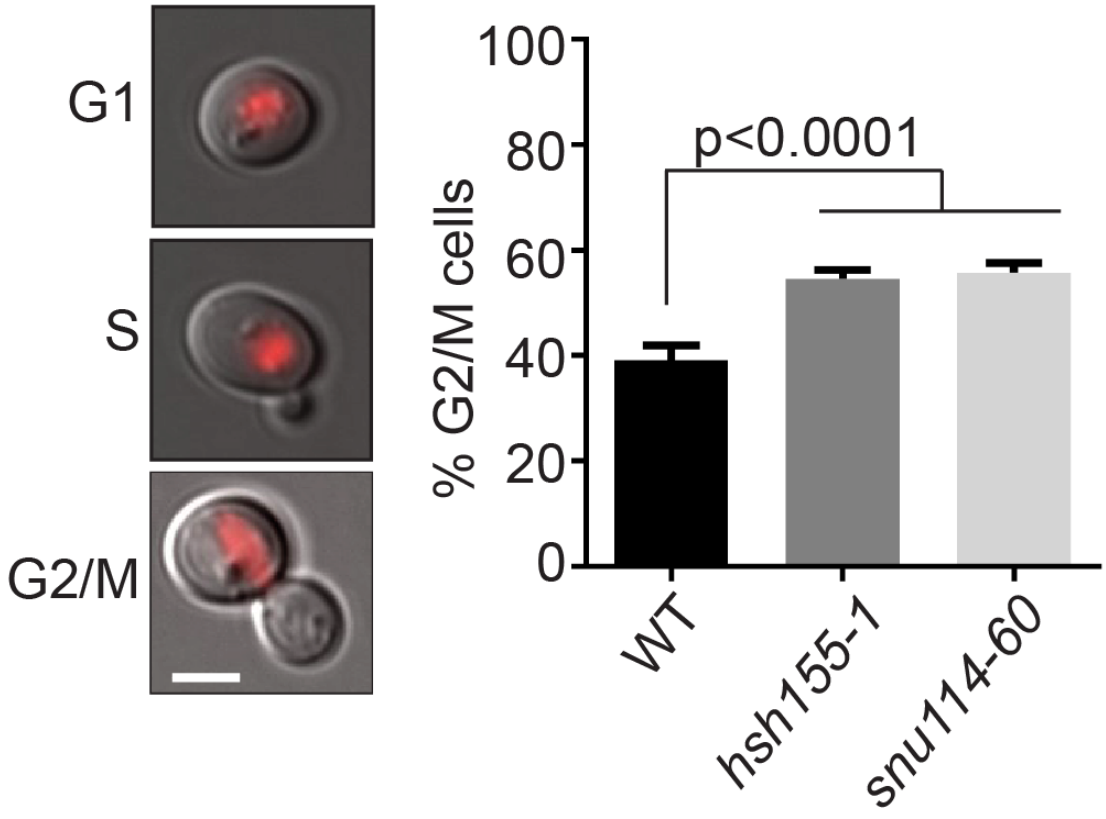

Figure 2. 9. Budding indices of splicing factor mutants.
Proportion of G2/M cells determined by budding relative to nuclear state determined by Hta2-mCherry fluorescence (left panels, scale bar = 2μm); Fisher's exact test. Mean values and S.E.M are shown, *n = 3*.

To confirm these observations, I used α-factor to arrest cells in G1, and collected samples at 30min, 90min and 150min after release for measurement of DNA content using FACS analysis. These time points were chosen to allow cells to complete one cell cycle post-release. The results

confirm a clear but subtle increase in 2N cells at 150minutes post release (**Figure 2.10**), compared to the positive control *ask1-2* that completely arrests in mitosis. This is consistent with the budding index data where the proportion of G2/M cells in mutants compared to wildtype is only slightly elevated.

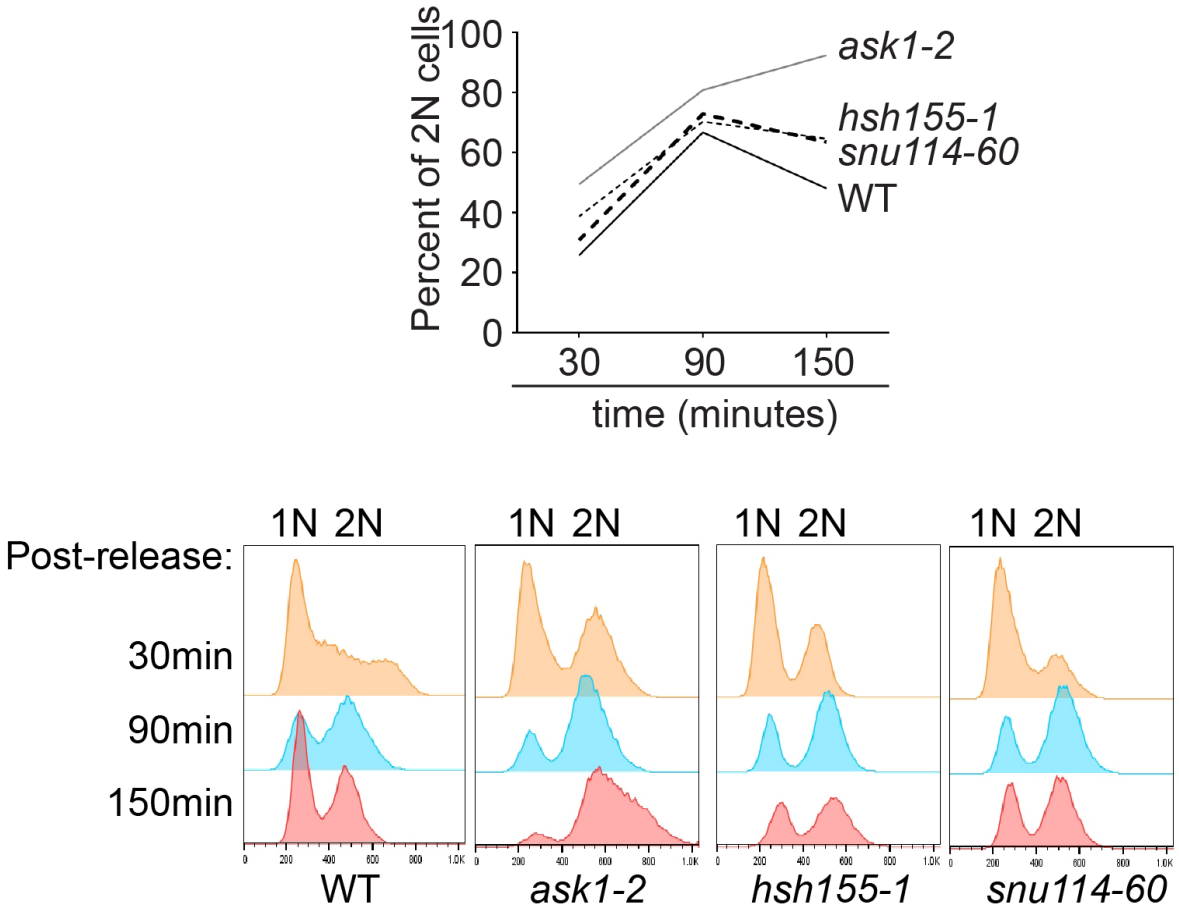

Figure 2. 10. **Cell cycle dynamics in splicing mutants.**

FACS analysis of DNA content (x axis, arbitrary units) in yeast cells 30min, 90min, and 150min after release from α-factor G1 arrest. Peaks indicate 1N and 2N cells. Top graph depicts percent of 2N cells in each strain at 30min, 90min, and 150min post-release.

To further confirm that cell cycle progress is abnormal, I used western blot analysis to measure levels of Clb2, a B-type cyclin that accumulates during G2 and M phases of the cell cycle [144]. At the restrictive temperature, there is an increase of Clb2 protein levels in the

splicing mutants, indicating more cells are indeed in G2/M phases of the cell cycle (**Figure 2.11**).

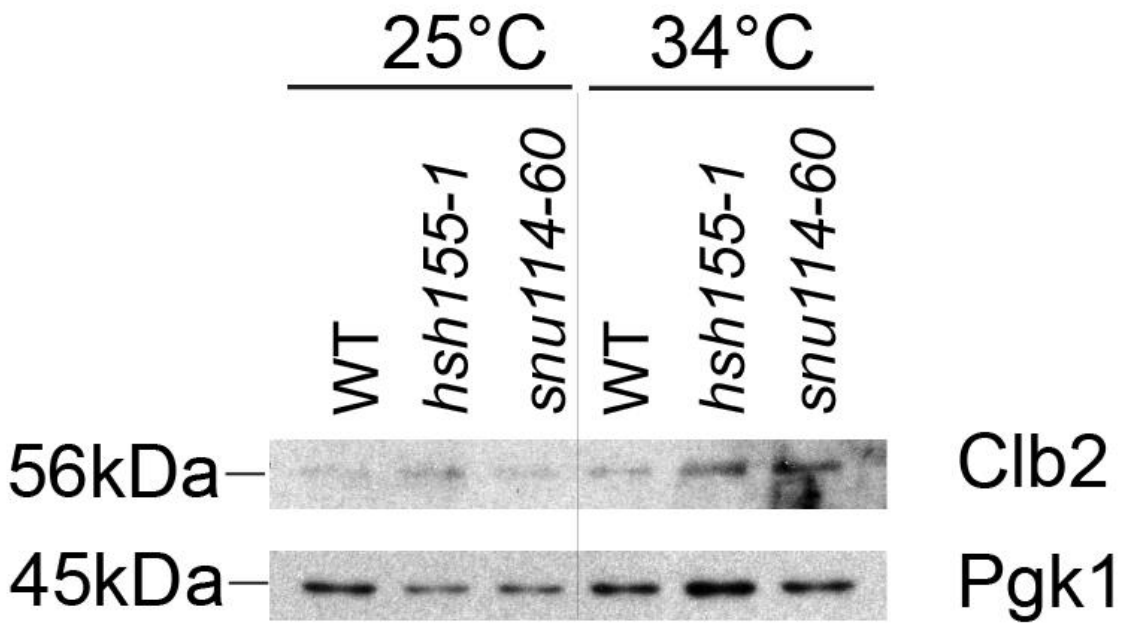

Figure 2. 11. G2/M delays indicated by accumulation of Clb2 protein levels.
Western blot of Clb2 protein levels in the indicated strains and temperatures.

These results complement the observed negative genetic interactions between *hsh155-1* and genes with functions in spindle and kinetochore subunits (**Figure 2.7**). Previously, it was shown in budding yeast that mutations in splicing factors *CEF1*, *PRP17*, and *PRP22* may have important roles in splicing transcripts critical for G2/M progression, therefore mutations in these genes cause cell cycle arrest phenotypes [69]. Based on those findings and also my data thus far, I hypothesized that cell cycle defects in *hsh155-1* and *snu114-60* were due to spindle defects activating the spindle assembly checkpoint (SAC). The SAC is a regulatory mechanism that maintains genomic stability by delaying cell division until chromosomes are accurately bi-oriented on the mitotic spindle [145]. During mitosis, unattached kinetochores trigger the SAC, which in turn inhibits the anaphase promoting complex, or cyclosome (APC/C), an E3 ubiquitin ligase that targets cell cycle regulators for degradation. As a consequence, inhibition of APC/C

prevents sister chromatid separation. In yeast, key components of the SAC include Mad1, Mad2, Mad3, Bub1, and Bub3, amongst others [146].

To test if the cell cycle defects observed in the mutant strains were due to activation of the SAC, I deleted the SAC regulator *MAD1* in each splicing mutant. Loss of *MAD1* further sensitized *hsh155-1* and *snu114-60* alleles to the microtubule depolymerizing drug benomyl at semi-permissive temperatures (**Figure 2.12A**). Mutants with defective microtubules often display hypersensitivity to benomyl [147], therefore in strains with existing microtubule defects, loss of a SAC regulator could lead to loss of viability when assessing a population of cells like in the case of a spot dilution assay, since chromosome missegregation and the resulting fluctuations in genetic material leads to abrupt changes in gene expression [148]. To further understand the consequences of defective microtubules in the splicing mutants, I used the A-like faker (ALF) assay to quantify the extent of chromosome instability [126]. Briefly, in the ALF assay, α-mating type haploid mutant yeast carrying the *MATα* locus are crossed to a wildtype α-mating type haploid strain, which normally leads to very few colonies as α-mating types do not mate with each other, and only mate with a-mating type haploid yeast. However, upon a chromosome instability event leading to loss of the *MATα* locus, this leads to de-differentiation to a-mating type. As a result, it is simple to qualitatively assess chromosome instability frequency using this assay by looking at the number of colonies formed. I performed this assay in the single (*hsh155-1* and *snu114-60*) and double (*hsh155-1 mad1*Δ and *snu114-60 mad1*Δ*)* mutants at permissive temperature of 25°C, and found a dramatic synergy between disruption of *SNU114* or *HSH155* and loss of the SAC through *MAD1* deletion (**Figure 2.12B**). Overall, these data are consistent with a mitotic defect in splicing mutants that requires the activity of the SAC for genome maintenance.

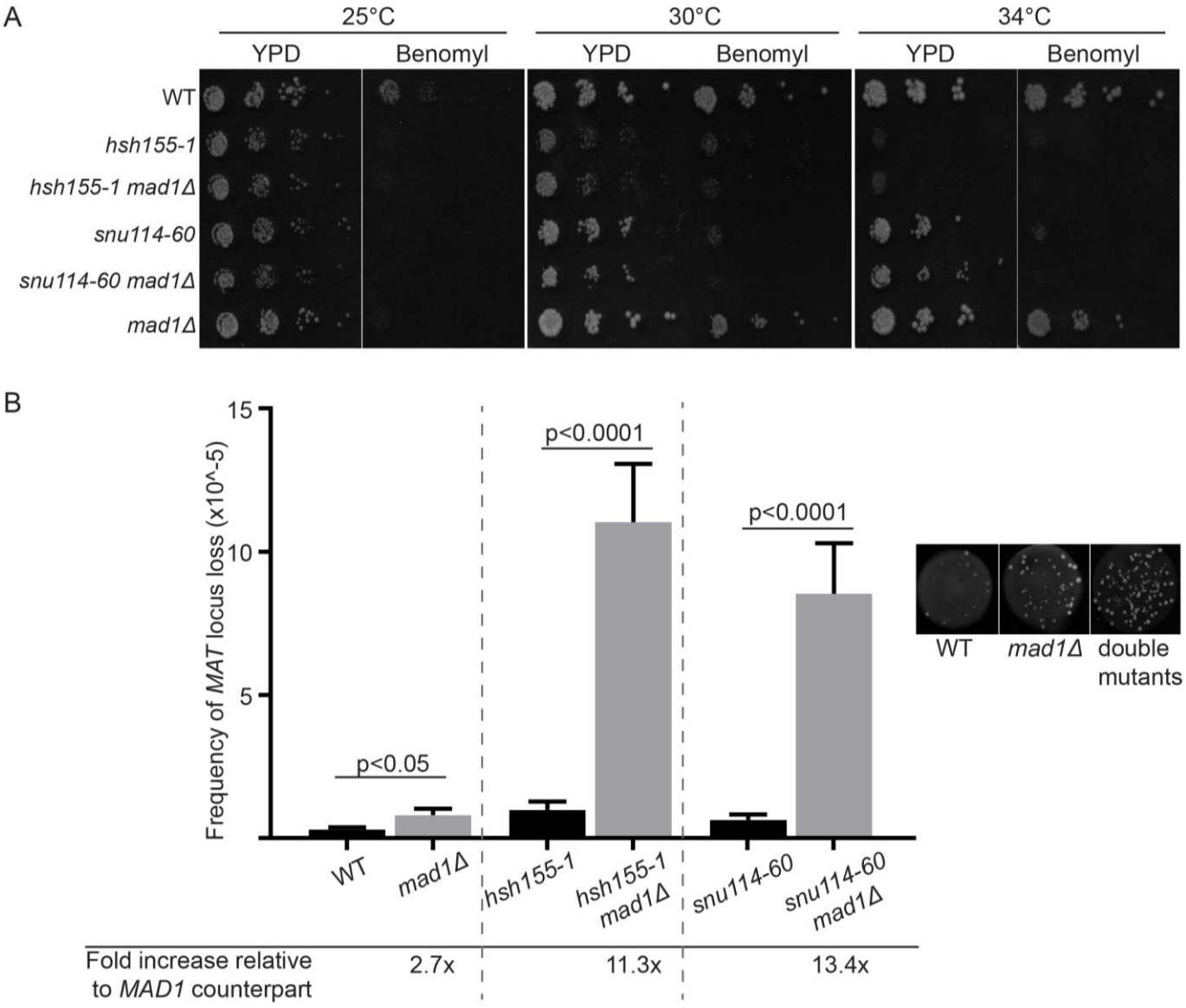

Figure 2. 12. Mitotic defects in spliceosome mutants.

(A) Spot dilution assay of benomyl sensitivity at the indicated temperatures (15μg/mL benomyl YPD plates). (B) Frequency of *MAT* loss in single (black bars) and *mad1Δ* double mutants (grey bars), grown at 25°C. Right, representative images of WT, *mad1Δ* and double mutant spots. Student's t-test was used to calculate significance. Mean values and S.E.M are shown, *n = 3*.

2.3.5 Tubulin levels control genome integrity in splicing mutants

Only about 5% of yeast genes contain introns, however the majority of introns are found in the most highly transcribed mRNAs encoding ribosomal protein genes, therefore more than 70%

of bulk transcripts in yeast cells come from intron containing genes [41, 124, 149]. Splicing of the *TUB1* transcript, encoding α-tubulin, has previously been implicated in cell cycle delays in other splicing mutants [14, 69]. Moreover, mutant alleles of *TUB1* such as the cold-sensitive *tub1-1* have been shown to increase chromosome missegration at low temperatures [150] and lead to decreased Tub1 protein (**Figure 2.13A**). Tub1 protein levels decreased in both *hsh155-1* and *snu114-60* mutants as temperature increased (**Figure 2.13B**).

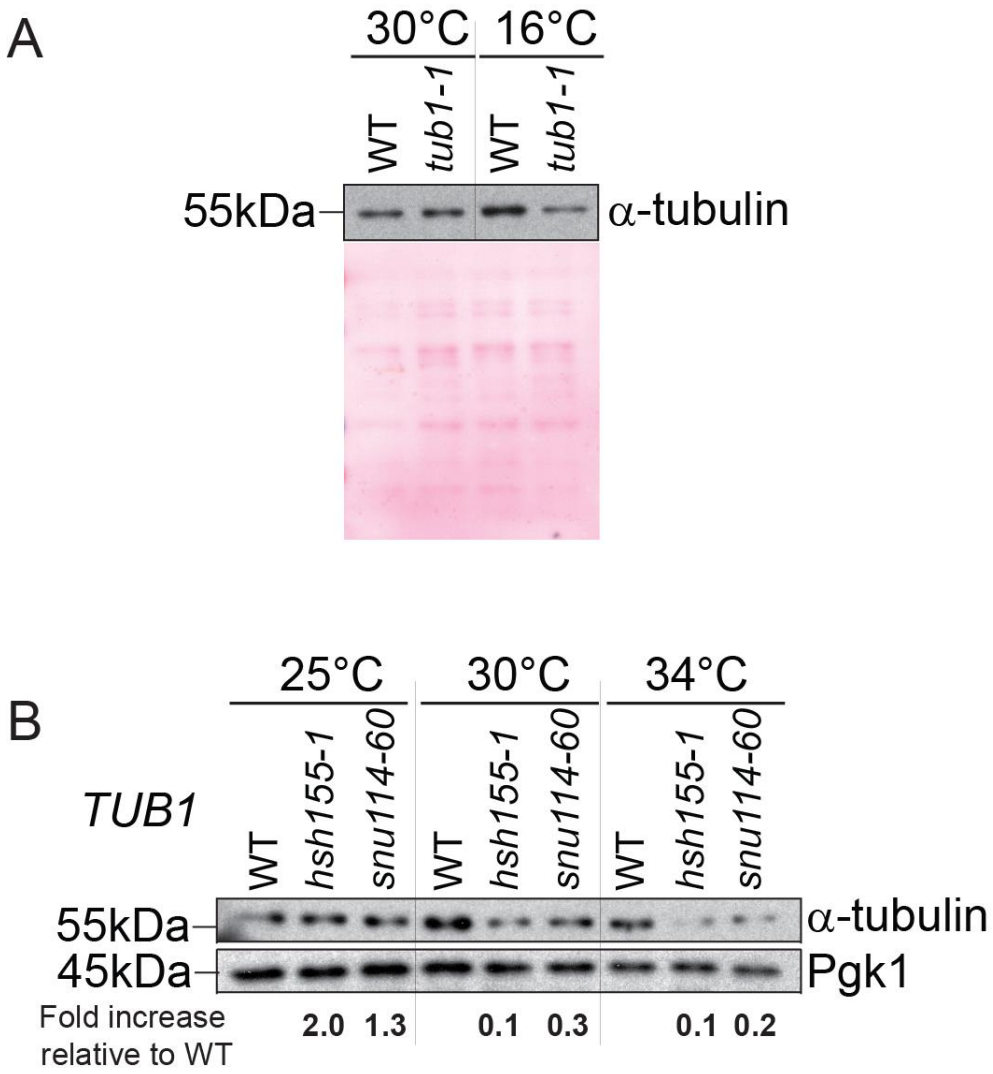

Figure 2. 13. Fluctuating Tub1 protein levels in the indicated strains.

(A) Cold sensitive *tub1-1* allele at permissive (30°C) and restrictive (16°C) temperatures. Ponceau stain was used as loading control. (B) Western blot of relative α-tubulin protein levels in *hsh155-1 and snu114-60* at increasing temperatures. Protein bands were quantified using ImageJ software.

RT-qPCR of *TUB1* mRNA under the same conditions indicated a mild decrease in *TUB1* mRNA expression in *hsh155-1* and *snu114-60*, which was accompanied by considerable intron retention (**Figure 2.14A**), supporting the idea that defective *TUB1* mRNA splicing drives loss of protein. Indeed, Tub1 protein level decreases of variable penetrance in a panel of splicing factor mutants tested (*i.e.* alleles of *YHC1, SNU13, SYF1, CWC2, PRP4, PRP31*, and *PRP6)* (**Figure 2.14B**), supporting that this is a general phenomenon.

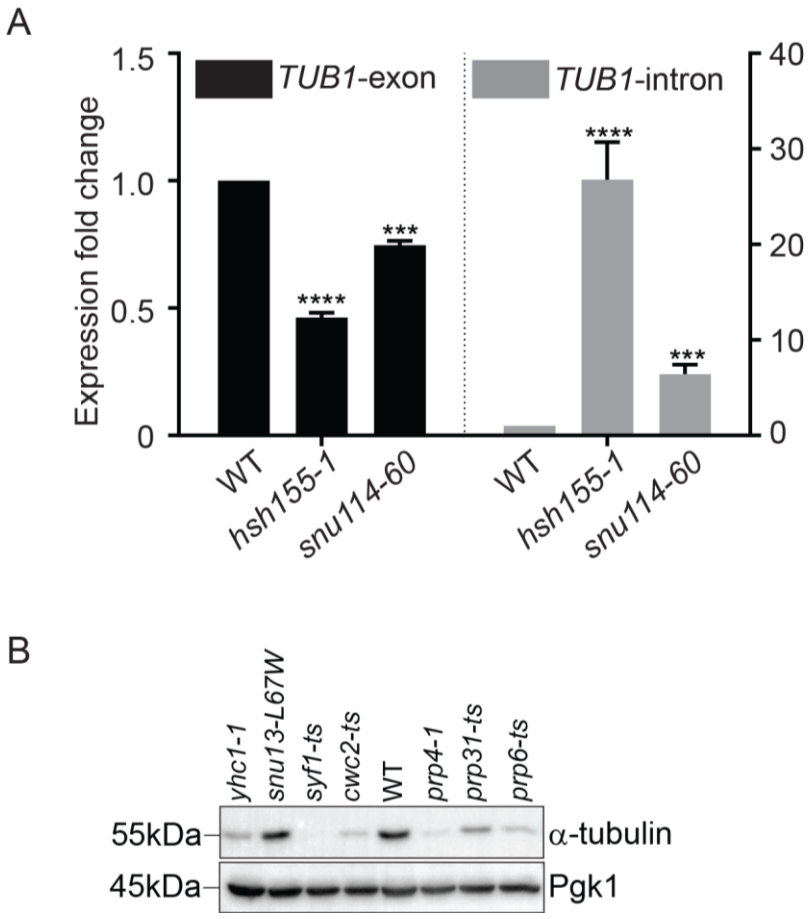

Figure 2. 14. TUB1 is suboptimally expressed in splicing factor mutants.

(A) Quantification of *TUB1* mRNA transcript levels from exon region (left) and *TUB1* intron region (right) by RT-qPCR in *hsh155-1* and *snu114-60*. Asterisks show p-values of ΔΔCt – ***p=0.0002; ****p<0.0001. Mean values with S.E.M. error bars are shown, *n = 3*. (B) Western blot of α-tubulin protein levels in the indicated strains at 37°C. Western blots were performed by K.L. Milbury.

To directly connect defective *TUB1* splicing to genome maintenance, I retested CIN in *hsh155-1* and *snu114-60* encoding an intronless *TUB1* gene, *tub1*Δ*i*. The intronless *TUB1* strain used to generate the *hsh155-1*, *tub1*Δ*i* and *snu114-60*, *tub1*Δ*i* double mutants had its single intron deleted genomically, and was marker free to ensure no other transcriptional perturbations affected expression. The presence of intronless *TUB1* was verified by colony PCR. As expected, *tub1*Δ*i* increased the amount of Tub1 protein expressed in each splicing mutant relative to WT (**Figure 2.15A**). More importantly, intronless *TUB1* partially suppressed the CTF phenotype observed in *snu114-60* and *hsh155-1* (**Figure 2.15B**). I also tested the effect of *tub1*Δ*i* on the rate of endogenous chromosome III mis-segregation using the LacO-LacI-GFP system (explained in section 2.3.1) and found suppression of aneuploidy in both splicing mutants (**Figure 2.15C**).

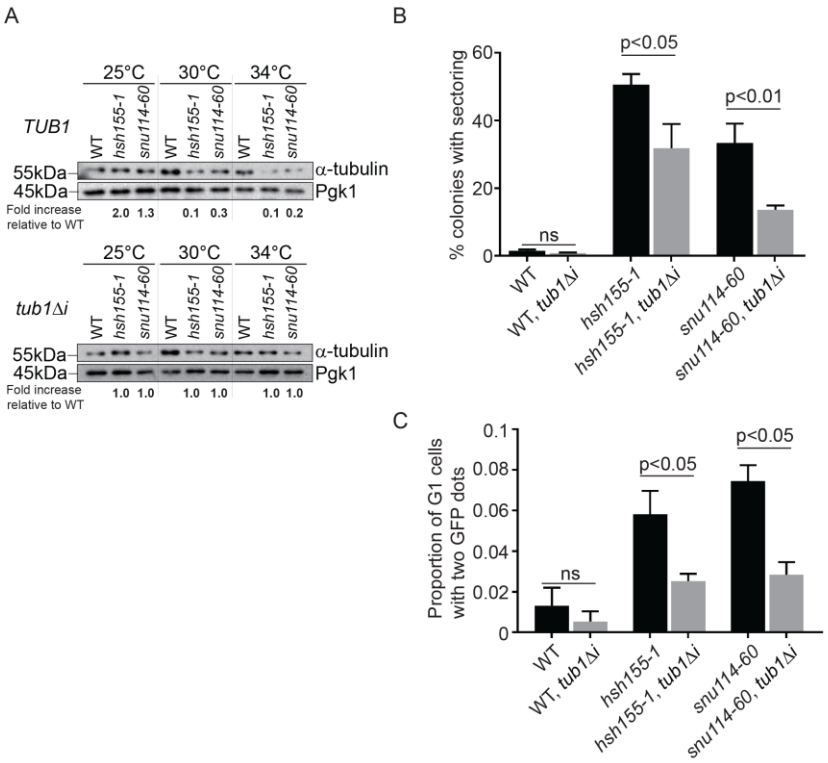

Figure 2. 15. Tubulin stability contributes to genome maintenance in splicing mutants. (A) Western blot of relative α-tubulin protein levels. Top: *TUB1* (Note, this is a copy of Figure 2.13B); Bottom: intronless *TUB1*. *TUB1* and intronless *TUB1* samples across different temperatures were collected at the same time per replicate. (B) CTF phenotypes in *TUB1* or *tub1*Δ*i* strains. Protein bands were quantified using ImageJ software. (C) Endogenous Chr III stability in *hsh155-1* and *snu114-60* with intronless *TUB1* (grey bars). (B,C) Fisher's exact test. Mean values with S.E.M. error bars are shown, *n = 3*.

As expected, *tub1Δi* has no effect on R-loop levels or recombination in *snu114-60*, demonstrating that there are multiple mechanisms, operating concurrently, that create genome instability in this strain (**Figure 2.16**).

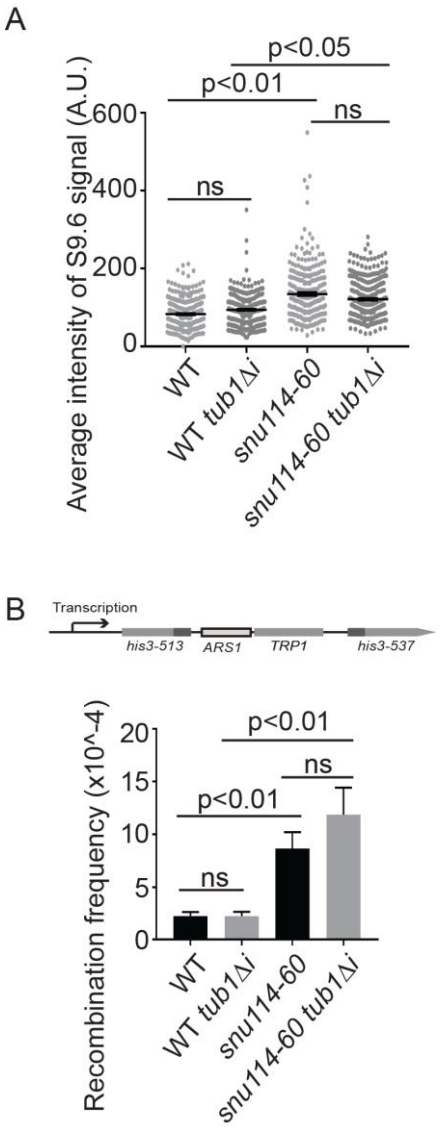

Figure 2. 16. R-loop-mediated recombination is not caused by suboptimal splicing of TUB1. (A) s9.6 antibody staining intensities in chromosome spreads. ANOVA was used to determine statistical significance. Mean values and S.E.M are shown, *n = 3*. (B) Frequency of direct repeat recombination on the indicated integrated genomic reporter. Reporter construct schematic is presented above panel. ANOVA was used to determine statistical significance. Mean values and S.E.M are shown, *n = 5*.

Tubulin levels must be tightly regulated, and maintaining an equimolar ratio of α- to β-tubulin is known to be critical for functional spindles [151]. The results presented in this data chapter indicate that splicing defects lead to intron retention and downregulation of α-tubulin levels, ultimately leading to chromosome instability. If this is true, then simple reduction of *TUB1* transcription should also cause genome instability even with normal splicing. To test this, I engineered strains where the *TUB1* gene was under the control of a galactose-regulated promoter in a WT or *tub1*Δ background. Shifting the strains to dextrose repressed Tub1 expression in the *tub1*Δ background (**Figure 2.17A**) and led to a significant increase in chromosome instability measured by the ALF assay (**Figure 2.17B**).

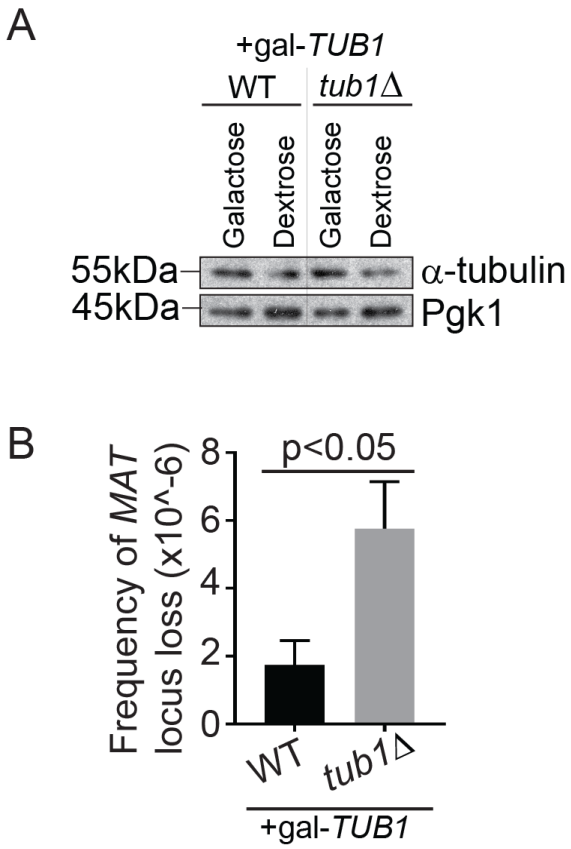

Figure 2. 17. Reduction of *TUB1* expression causes genome instability.

(A) Western blot of Tub1 protein levels in the indicated strains, in the presence of galactose or dextrose-containing media to induce or repress expression of *TUB1*. (B) Frequency of MAT locus loss in the indicated strains in dextrose. Student's t-test was used to calculate significance. Mean values and S.E.M are shown, n = 5.

The results presented in this data chapter indicates altered α-tubulin levels cause CIN, whether through splicing defects or decreased transcription. My data support a model (**Figure 2.18**) where intron retention and associated decreases in α-tubulin lead to sporadic chromosome missegregation, buffered by the SAC.

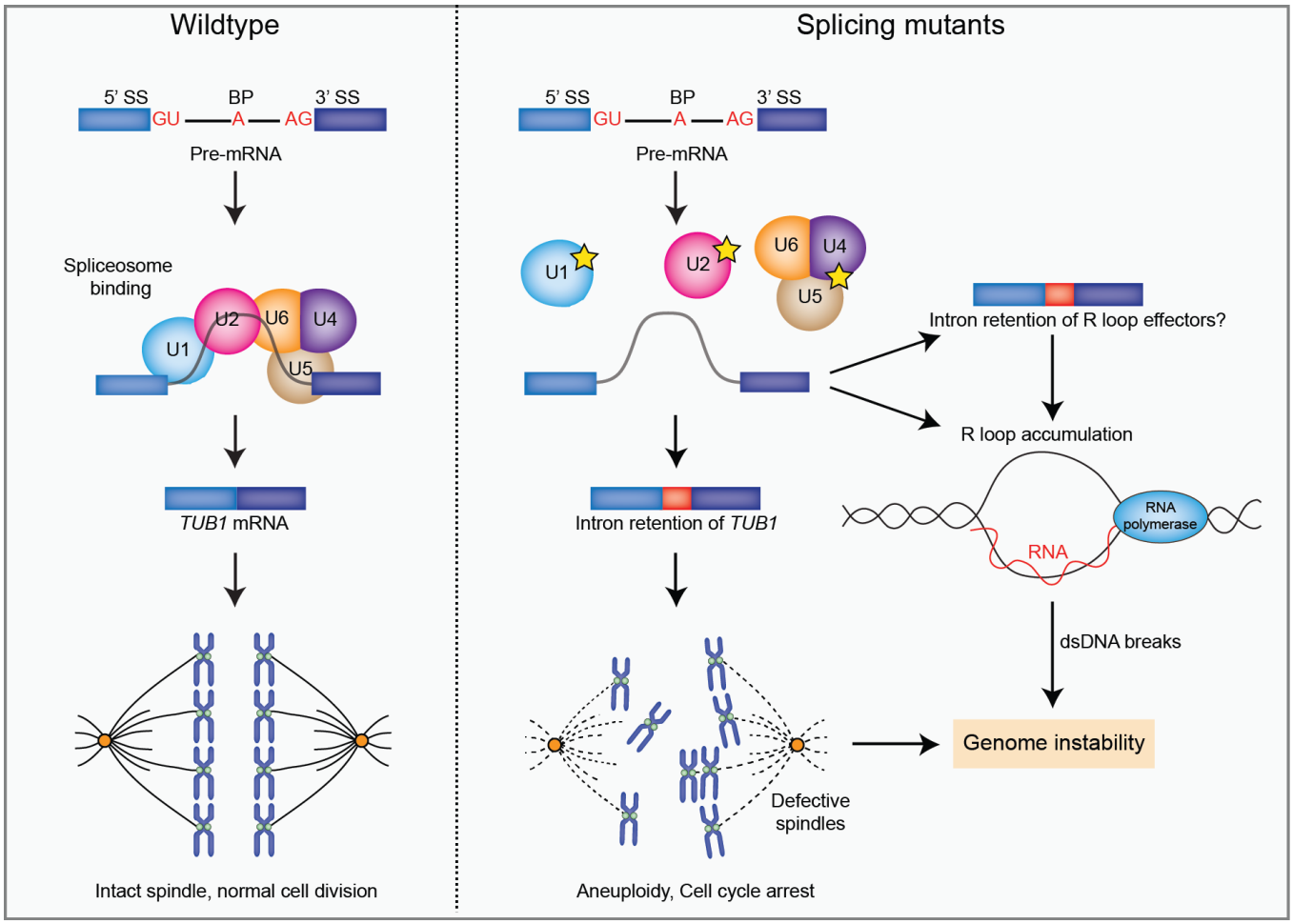

Figure 2. 18. Model of defective splicing-induced genome instability in yeast.
Model schematic designed and illustrated by V. Mathew.

2.4 Conclusion

Determining sources of genome instability is important to understand the accumulation of mutations during cellular adaptation and in human disease. While much is known about DNA replication, repair and mitosis, considerably less is known about the effects of other cellular pathways on the genome. Nonetheless, these non-canonical pathways account for a large

proportion of reported genome maintenance factors [53]. RNA processing has emerged as a major contributor to genome maintenance and various mechanisms have been described. Direct roles for some RNA processing factors have been found in DNA repair, such as the moonlighting function of Prp19 in ATR activation [87], or the role of the spliceosome in R-loop mediated ATM activation [152]. More recently, dominant cancer-associated splicing mutations have been linked to RNA polymerase pausing and R-loop accumulation [64]. Still other studies have suggested a role for the aberrant gene expression landscapes produced in RNA processing mutants as drivers of genome instability [83, 86].

Overall, my data reveal how changes in information flux due to allele-specific effects on RNA maturation can influence specific mechanisms of genome instability. This is important because more than 150 yeast genes with functions in transcription, RNA processing and translation can be mutated to cause a genome instability phenotype [53]. In principle my data suggest that many of these mutants with broad impacts on gene expression could selectively impair a specific aspect of genome maintenance that can be identified. Indeed, cancer-associated mutations in core splicing factors like SF3B1, the orthologue of yeast Hsh155, are invariably non-synonymous coding variants suggesting that a specific change in the transcriptome is required for cancer formation and maintenance [77]. Mechanistic studies of how specific mutations in the spliceosome alter the proteome in a specific cellular context could elucidate roles for spliceosomal mutations in cancer. Building on the data presented in Chapter 2, in Chapter 3, I will expand this research into studying *SF3B1* mutations in the context of cancer biology and genome maintenance. I will describe how a specific point mutation prevalent in human cancers can cause replication stress and DNA damage through both R-loop accumulation and potentially missplicing of a key transcript important for genome maintenance. Implications of these findings in the context of the rest of the thesis are discussed in Chapter 4.

Chapter 3: Cancer-associated SF3B1 hotspot mutations influence genome instability

3.1 Background

SF3B1 is the most frequently mutated splicing factor in cancer. Mutations in *SF3B1* occur in a broad range of cancers and blood disorders including MDS [153, 154], CLL [155, 156], AML [157], breast cancer [158] and uveal melanoma [159], amongst others [29], implicating spliceosome dysfunction as a driver of disease. Although many transcriptome analysis studies have characterized aberrant splicing events induced by *SF3B1* mutations [78, 79, 120, 159-164], the biological consequences of these missplicing events remains to be determined.

The hotspot mutations identified in *SF3B1* are located in the HEAT repeat domain, and are heterozygous missense mutations, suggesting that these mutations likely confer a gain of function phenotype rather than a loss of SF3B1 function. Prior studies have linked SF3B1 to genome maintenance, either through direct interactions with DNA damage response proteins like BRCA1 and BCLAF1 upon DNA damage [85], changes in expression of select genes like *ATM* and *TP53* which has been suggested to alter ATM/p53 transcriptional and apoptotic responses to DNA damaging agents [165], or global deregulation of splicing in factors that are important for cell cycle progression and DNA damage response [166].

Splicing is a transcription coupled process, therefore splicing factors that affect transcription may lead to accumulation of DNA:RNA hybrids in genomic DNA. These three-stranded R-loop structures contribute to genome instability by exposing ssDNA and by blocking replication forks, causing replication stress induced genome instability [119, 120]. Previous work has suggested loss of splicing factors like ASF/SF2 [51], or treatment with splicing inhibitors [55] induce aberrant R-loops. In yeast, splicing has been shown to be an important factor in genome maintenance [59]. Recently, two studies have demonstrated that cancer-associated mutations in splicing factors *U2AF1* and *SRSF2* elevated R-loops in leukemia cell lines, resulting in replication stress, activation of the ATR-CHK1 signaling pathway, and increased DNA damage [63, 64]. Although it has long been known that splicing defects can cause R-loop mediated DNA damage, and treatment with splicing inhibitors of SF3B1 like spliceostatin A and pladienolide B

induce R-loop mediated damage [55, 167], it has been unclear whether *SF3B1* cancer-associated mutations also lead to aberrant R-loop accumulation and genetic instability.

In Chapter 2, I used a conditional mutant of *HSH155*, the yeast homolog of *SF3B1*, to identify that suboptimal expression of a protein important for mitotic spindle assembly contributed to genetic instability. In an effort to gain functional insights to how *HSH155* could influence genetic instability in the context of cancer progression, I extended my analysis to include five cancer-associated *SF3B1* mutations in the equivalent Hsh155 protein residues. While I found that the splicing activity in SF3B1 and Hsh155 was conserved, yeast *HSH155* mutations at corresponding amino acids to those found mutated in cancer caused only subtle defects in splicing reporters, and lacked obvious reproducible phenotypes, precluding the further use of yeast as an experimental system. Instead, I shifted my attention to human cells, and used one isogenic pair of cell lines to account for complexities in the mammalian splicing machinery. Using human cells, my goal was to determine if a specific *SF3B1* hotspot mutation contributed to genetic instability, as well as to identify the mechanism by which this occurred. The results of my study indicate that the common cancer-associated SF3B1 H662Q mutation induces R-loop-mediated replication stress, as well as aberrant expression of DYNLL1, a small protein that binds to and regulates many proteins with functions that include the DNA damage response and apoptosis, amongst others. My data suggests that higher DYNLL1 protein may account for an observed potential increase in the activity of nonhomologous end joining, and downregulation of RAD51 loading in S/G2 phase. I propose a model in which this dysregulation of repair proteins by high levels of DYNLL1 could be a driver of replication stress in SF3B1 mutated cancers. Overall, I have identified a cancer-associated *SF3B1* mutation that may contribute to genetic instability both through R-loop accumulation and changes in gene expression.

3.2 Materials and methods

3.2.1 Yeast strains, growth, and CIN assays

All yeast strains were in the s288c background, and were grown under standard conditions in the indicated media and growth temperature. Spot assays were performed by diluting log phase cultures to OD 0.1 and serially diluting each condition in 1:10 dilutions in YPD. Cultures were spotted on to YPD plates using a pin tool, and incubated at 30°C. Growth curves were conducted

in YPD using a Tecan M200 plate reader and were compared using the area under the curve as previously described [125]. Briefly, logarithmic phase cultures were diluted to OD 0.05 in a 96-well plate and grown for 48 hours, with OD600 readings taken every 30 minutes. For all transformations, yeast cells were transformed with lithium acetate transformation and incubated on selective media. For plasmid loss assay to measure frequency of chromosome instability, strains carrying *pRS313::HIS3* were grown overnight in SC-histidine, then plated on YPD and allowed to form colonies without selection before replica plating onto SC-histidine. The frequency of colonies that could not grow on SC-histidine is reported as the plasmid loss rate. For all experiments statistical significance was determined using Prism7 (GraphPad Software). For all experiments, sample means were compared with Fisher's exact test, Student's t-test or ANOVA for multiple comparisons as indicated.

3.2.2 *HSH155* mutagenesis

To generate the cancer-associated *SF3B1* mutations in the *HSH155* homolog, first the wildtype *HSH155* gene was cloned into the integrating vector SB221 using primers designed for constructing the temperature sensitive allele *hsh155-1* [168, 169]. Site-directed mutagenesis to produce the cancer-associated mutations was performed using PCR and QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies, cat# 200518). Sequence verified clones were digested with NotI and transformed into the heterozygous *hsh155*Δ*::KanMX/HSH155* diploid strain. SB221 integrates into the *KanMX* locus and integrants were selected to be *URA3* positive and G418 sensitive. Sporulation and tetrad dissection of the confirmed heterozygotes was used to isolate the haploid yeast used in this chapter.

3.2.3 Yeast splicing efficiency assay

Splicing assay protocol was performed as described [133]. All measurements were taken with individual transformants in triplicate. Cells were struck as a patch on SC-leucine and then replica plated to glycerol-lactate–containing SC medium without leucine (GGL-leu). Cells from each patch were inoculated in liquid GGL-leu media for 2 h at 30°C and then were induced with final 2% galactose for 4 h. Cells carrying reporters were lysed and assayed for β-galactosidase assay using a Gal-Screen β-galactosidase reporter gene assay system for yeast or mammalian cells (Applied Biosystems) as per the manufacturer's instructions and read with a SpectraMax i3

(Molecular Devices). Relative light units were normalized to cell concentration as estimated by measuring OD600.

3.2.4 Imaging and microscope settings

Fluorescent imaging was conducted on a Leica DMi8 microscope using an HCX plan apochromat 1.4 NA oil immersion 100x lens. The images were captured at room temperature by an ORCA Flash 4.0 V2 camera (Hamamatsu Photonics), using MetaMorph Premier acquisition software (Molecular Devices). Scoring was done in ImageJ (National Institutes of Health).

3.2.5 Yeast imaging and cell cycle analysis

Yeast chromosome spreads were performed exactly as described; primary DNA-RNA Hybrid [S9.6] (Kerafast cat#ENH001); secondary Alexa Fluor® 568 goat anti-Mouse IgG (Invitrogen cat#A-11004) [132]. Flow cytometry was done using the BD FACSCalibur™ platform. For cell cycle analysis, cells were fixed with 70% ethanol overnight at 4°C, washed with sodium citrate and stained with 16 μg/mL propidium iodide (Sigma-Aldrich cat#287075) in 0.25 mg/mL Ribonuclease A (Sigma-Aldrich cat#R-5000). The proportion of cells in G1, S and G2/M cell cycles was quantified using FlowJo.

3.2.6 Mammalian cell culture and transfection

NALM6 wildtype and *SF3B1-H662Q* cells were purchased from Horizon Discovery and grown in RPMI-1640 medium (STEMCELL Technologies) supplemented with 10% fetal bovine serum (Life Technologies) in 5% CO_2 at 37°C. Transfections were performed using the Neon™ Transfection System (Invitrogen) following manufacturer instructions. Briefly, 80-90% confluent cells were harvested and washed twice with PBS, then 2×10^5 cells were transfected with a 10μL Neon Tip via electroporation using the following optimized settings: 1350 V, pulse length 10ms, total of 4 pulses. Cells were harvested 48 hours after transfection. For experiments with overexpression of GFP or nuclear-targeting GFP-RNaseH1 (gift from R. Crouch), 1 μg of plasmid per reaction was used. For RNA interference, siRNA concentrations of 20, 50, and 100 μM si-Control or si-DYNLL1 were used (siGENOME-SMARTpool siRNAs from Dharmacon).

3.2.7 Western blot analysis

Whole-cell lysates were prepared using 1 x RIPA buffer containing protease inhibitor (Sigma) and phosphatase inhibitor PhosSTOP™ (Roche Applied Science) cocktail tablets. 10X RIPA buffer: 0.5M Tris-HCl, pH 7.4, 1.5M NaCl, 2.5% deoxycholic acid, 10% NP-40, 10mM EDTA.

Protein concentrations were assessed with Bio-Rad Protein assay (Bio-Rad). Lysates were separated on a 15%, 10%, or 8%/15% gradient SDS-PAGE gel, transferred to 0.45μm PVDF membranes (Millipore), blocked with 5% skim milk in TBS containing 0.1% Tween-20 (TBS-T), and probed with the following antibodies: γH2AX (1:1000, abcam cat#ab81299), p-ATR (Ser428) (1:1000, Cell Signaling cat#2853), p-CHK1 (Ser345) (1:1000, Cell Signaling cat#133D3), p-ATM (1:500, Santa Cruz cat#sc-47739), p-CHK2 (Th468) (1:1000, Cell Signaling cat#C13C1), GAPDH (1:3000, ThermoFisher Scientific cat# MA5-15738), vinculin (1:2000, Santa Cruz cat#sc-73614), p-RPA32 (1:1000, Bethyl cat# A300-246A), RAD51 (1:500, Santa Cruz sc-398587), CDC7 (1:500, Santa Cruz, sc-56274), p-MCM2 (1:2000, Abcam, cat# ab70371), MCM2 total (1:1000, Abcam cat# ab108935), DYNLL1 (1:5000, Abcam, cat# ab51603). Secondary antibodies were conjugated to horseradish peroxidase (HRP) and peroxidase activity was visualized using SuperSignal™ West Pico PLUS Chemiluminescent Substrate (Thermo Scientific cat# 34577). ImageJ software was used to quantify protein bands [136].

3.2.8 Immunofluorescence EdU staining kit for imaging

Suspension cells were harvested, washed twice with PBS and incubated on poly-L-lysine coated coverslips for 15 mins at room temperature before fixation. To make coverslips, 0.1% poly-L-lysine solution (Sigma-Aldrich cat# P8920) was added to coverslips for 20 mins, removed, and slides were dried at 37°C for 15 mins. For S9.6 staining, cells were fixed with ice-cold methanol for 15 mins and permeabilized with ice-cold acetone for 1 min. After PBS wash, cells were blocked in 3% bovine serum albumin (BSA), 0.1% Tween-20 in 4X saline sodium citrate buffer (SSC) for 30 mins at room temperature in a humid chamber. Cells were incubated with primary antibody DNA-RNA Hybrid [S9.6] (1:500, Kerafast cat#ENH001) overnight at 4°C. Next day, cells were washed three times with PBS, and stained with secondary Alexa Fluor® 568 goat anti-Mouse IgG (1:1000, Invitrogen cat#A-11004) for 1 hr at room temperature, washed twice and PBS and one time with PBS+0.1% Tween-20 (PBS-T), and stained with DAPI for 5 mins. Cells were imaged on a LeicaDMI8 microscope at 100X and ImageJ was used for processing and quantifying nuclear S9.6 intensity in images. For experiments with GFP overexpression, only GFP-positive cells were quantified. For γH2AX (1:1000), MCM2 (1:1000), RPA32 (1:1000), 53BP1 (1:1000), and RAD51 (1:500), immunostaining was performed as above, except for the following: fixation with 4% paraformaldehyde for 15 mins and permeabilization with 0.2% Triton X-100 for 5 mins; secondary antibody was rabbit or mouse

Alexa-Fluoro-488 or 568-conjugated antibody (1:1000, Invitrogen). For EdU incorporation, Click-iT™ EdU Imaging Kit with Alexa Fluor™ 647 Azides (Invitrogen cat# C10086) was used following manufacturer instructions. Cells were labeled with 10 μM EdU for 15 mins before fixation. After fixation and permeabilization, Click-iT® reaction cocktail with Alexa Fluor™ 647 was added to slides for 30 mins, wash twice with 3% BSA in PBS, and blocked with 3% bovine serum albumin (BSA), 0.1% Tween-20 in 4X saline sodium citrate buffer (SSC) for 30 mins at room temperature in a humid chamber. Antibody staining proceeded as described above.

3.2.9 Neutral and alkaline comet assay

The neutral and alkaline comet assays were performed using the CometAssay Reagent Kit for Single Cell Gel Electrophoresis Assay (Trevigen cat#4250-050-K) following manufacturer instructions. Cells at a concentration of 1.5 x 10^5 cells/mL were combined with low melt agarose (LMAgarose) at 37 °C at a ratio of 1:10 (cell:LMAgarose) and spread onto CometSlide. After gelling in 4 °C in the dark, the slides were then immersed in Lysis Solution overnight at 4°C. For the neutral comet assay, slides were removed from Lysis Solution and immersed in 4°C Neutral Electrophoresis Buffer for 30 min. Electrophoresis was then performed at 4°C at 21 Volts for 40 min in Neutral Electrophoresis Buffer. Slides were immersed in DNA Precipitation Solution for 30 min followed by 70% ethanol for 30 min at room temperature, and dried at 37°C for 15 min. Slides were then stained with 16 μg/mL propidium iodide (Sigma-Aldrich cat#287075) for 30 mins, gently washed with water, and allowed to dry completely at 37°C for 15 min. Slides were imaged on LeicaDMI8 microscope at 10X. Comet tail moments (tail length x fraction of total DNA in the tail) were obtained using an ImageJ plugin as previously described [170], and at least 50 cells per sample were analyzed for each independent replicate. For the alkaline comet assay, the protocol is the same except for the following: after lysis, slides were immersed in Alkaline Unwinding Solution for 20 mins at room temperature, electrophoresis was done in Alkaline Electrophoresis Solution, after which the slides were immersed twice in dH2O for 5 mins each, followed by 70% ethanol for 5 mins.

3.2.10 Mammalian cell viability and doubling capacity

Cell numbers and viability were measured using the Advanced Image Cytometer NucleoCounter® NC-250™ (ChemoMetec). 20 μL of cell suspension was incubated with 1 μL of Solution 18 (ChemoMetec cat#910-3018) in Eppendorf tubes, and 10 μL of cell:dye mixture was added to NC-Slide A8™ chamber slides (ChemoMetec cat# 942-0003) for analysis. Solution

18 contains acridine orange (AO) which is used to counterstain living and dead cells, and DAPI, which is used to stain dead cells. For cell viability analysis, 500 µL of cell cultures were incubated for 72 hours, and measurements were taken every 24 hours. For cell viability and doubling capacity post-drug treatment, cells were incubated at the respective drug concentrations for 72 hours, except in the case of olaparib and mirin treatments where cells were initially treated for 48 hours, then media and drugs were replenished for another 72 hours. For viability measurements post si-DYNLL1, 500 µL of cell cultures were treated with 0, 20, 50, or 100 µM si-Control or si-DYNLL1 for 48 hours.

3.2.11 Mammalian cell cycle progression

Cell cycle progression of NALM6 cells was assessed using Click-iT™ EdU Alexa Fluor™ 488 Flow Cytometry Assay Kit (Invitrogen cat # C10425), following manufacturer instructions. Cells were incubated with 10 µM EdU for 45 mins, harvested, and washed twice with PBS. Cells were washed once with 1% BSA in PBS, fixed with 100 µL of Click-iT® fixative for 15 mins at room temperature, washed with 1% BSA in PBS, and permeabilized with 100 µL of 1X Click-iT® saponin-based permeabilization and wash reagent for 15 mins at room temperature. Click-iT® reaction cocktail (PBS, CuSO4, Alexafluor-488-azide, 1x Reaction Buffer Additive) was added to cells for 30 mins at room temperature protected from light. Lastly, cells were washed with 1X Click-iT® saponin-based permeabilization and wash reagent, and stained with 16 µg/mL propidium iodide (Sigma-Aldrich cat#287075) with 0.25 mg/mL Ribonuclease A for 30 mins at room temperature. Cells were washed and resuspended in PBS and analyzed with a BD LSRFortessa™ flow cytometer using CellQuestPro software. Cell cycle plots were generated using FlowJo Version 9.3.2.

3.2.12 Endpoint RT-PCR

Total RNA was isolated using RNeasy Plus Mini Kit (Qiagen), and reverse transcribed to cDNA using anchored-oligo(dT)18 primer and Transcriptor Reverse transcription (Roche). DYNLL1 was tested for splicing defects, specifically cryptic 3′ splice site usage as identified in Dolatshad, H. et al, 2016 [171]. Primer sequences are DYNLL1-forward (5'-GTTTCGGTAGCGACGGTATCT-3') and DYNLL1-reverse (5'-TCCGCATTTTTGATCACGGC-3'), taken from Dolatshad, H. et al, 2016 [171]. Primers were designed to obtain a PCR product that spans the spliced exon junctions in DYNLL1. The canonical splicing products results in a 133bp PCR product while the aberrant splicing product

results in a 147bp PCR product. PCR reaction was carried out as follows: Initial denaturation 98°C for 30 sec, 98°C 10 sec, 62°C 30 sec, 72°C 15 sec for 40 cycles, final extension of 72°C for 10 mins. PCR products were subsequently run on a 2% agarose gel at 100V for 1.5 hours to visualize the splice isoforms.

3.2.13 Drug treatments

NALM6 cells were plated at a concentration of 6 x 10^5 cells/mL, 500 µL per well in a 24-well plate. Cells were cultured with the following concentrations of drugs: CDC7 inhibitor PHA-767491 (Selleckchem cat#S2742) 0 µM, 2.5 µM, 5 µM, and 10 µM; olaparib (Selleckchem cat#S1060) 0 µM, 1 µM, 5 µM, and 10 µM; mirin (Selleckchem cat#S8096) 0 µM, 25 µM, 50 µM, and 100 µM. All 0 µM cultures were treated with final 0.01% DMSO. PHA-767491(CDC7i) was added in a single time point, and cell viability and proliferation was assessed at 0, 24, 48, and 72 hrs after drug exposure. Olaparib and mirin were initially added for 48 hrs, then 250 µL of the cell suspension from each well was discarded, and 250 µL of fresh media and the appropriate drug concentrations were added back for an additional 72 hrs before cell viability counts were taken.

3.3 Results and discussion

3.3.1 Conserved splicing function of SF3B1 cancer-associated mutations in yeast homolog Hsh155

SF3B1 is a large protein that consists of an intrinsically unstructured N-terminal domain and a C-terminal domain made up of 20 HEAT repeat residues [172]. SF3B1 mutations identified in cancers cluster in the HEAT repeat domain, and although over 40 residues were reported to be mutated in this domain, 33 clustered in HEAT repeats four to nine [172]. This is a region that is highly conserved between human SF3B1 and the yeast ortholog Hsh155 (>50% identical) [161].

In Chapter 2, I used a temperature sensitive (ts) allele of *HSH155* to identify that suboptimal expression of α-tubulin contributed to genetic instability in this mutant. To gain functional insights into how *HSH155* may influence genetic instability in the context of cancer, I used five cancer-associated *HSH155* mutants (E291D, R294L, H331Q, K335N, and P369E) for further analysis. These mutants were generated by a postdoctoral fellow in the lab, Dr. Veena Mathew

using site-directed mutagenesis. To account for possible differences in expression as a result of *URA3* marker integration (please refer to Section 3.2.2 for mutagenesis procedure), a wildtype strain carrying the same marker was used for comparison. First, I assessed if the splicing function in these mutated residues were conserved between yeast and humans. To do this, I monitored splicing efficiency using a *LacZ* reporter construct system that was previously described in Section 2.3.1. Briefly, this assay measures efficiency of splicing out an intron contained in *LacZ*. As a positive control, I used a temperature sensitive allele of *HSH155*, *hsh155-1* for comparison of splicing activity. The results revealed two mutants, K335N and P369E had splicing defects when compared to wildtype control (**Figure 3.1**). The H331Q mutant appeared to have decreased splicing efficiency compared to wildtype but the difference was not statistically significant, due to variability between replicates. The data indicates that although these are point mutations, mutations in the K335 and P369 residues results in changes in splicing, pointing to conservation of splicing function between yeast *HSH155* and human *SF3B1*. Indeed, it has now been shown that the HEAT repeats where the *SF3B1* cancer-associated mutations are found in humans can functionally complement the equivalent region in yeast Hsh155 [173].

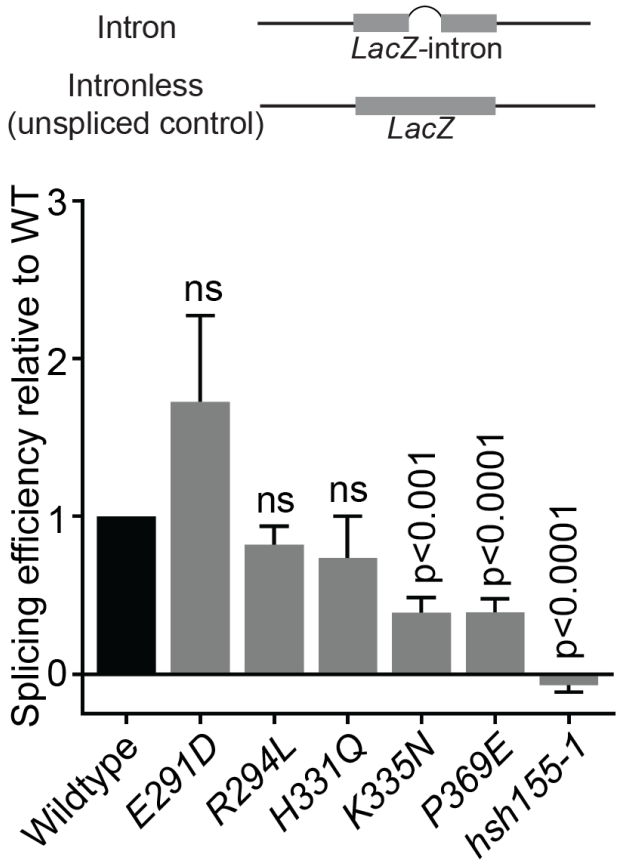

Figure 3. 1. Level of splicing deficiency in yeast *HSH155* cancer-associated mutants. Splicing efficiency of a *LacZ* reporter, values are relative to splicing levels in a wildtype strain carrying the same *URA3* marker. For each replicate, 3 technical replicates per strain were used to account for variability. Schematic of reporter constructs are presented above panel; Student's t-test was used to calculate significance. Mean values and S.E.M are shown, *n = 5*. Yeast *HSH155* mutants were generated by V. Mathew using site-directed mutagenesis.

3.3.2 Lack of observable phenotypes in Hsh155 cancer-associated mutations

Moving forward, I decided to include H331Q, K335N and P369E for further phenotype assessment. I chose to include H331Q because although the splicing defect was not statistically significant compared to wildtype, when assessing this mutant splicing efficiency appeared to be reduced in 4 out of 5 replicates. To characterize these mutants, I first assessed viability by serial dilution spot assays (**Figure 3.2A**) and 48 hour growth curves (**Figure 3.2B**) to check for subtle differences. The spot assay shows the mutants do not exhibit growth defects, and the growth

curves indicate that while the growth of H331Q appears to be reduced relative to wildtype, the difference is not statistically significant. Therefore, using these assays, the mutants do not appear to have growth defects compared to wildtype.

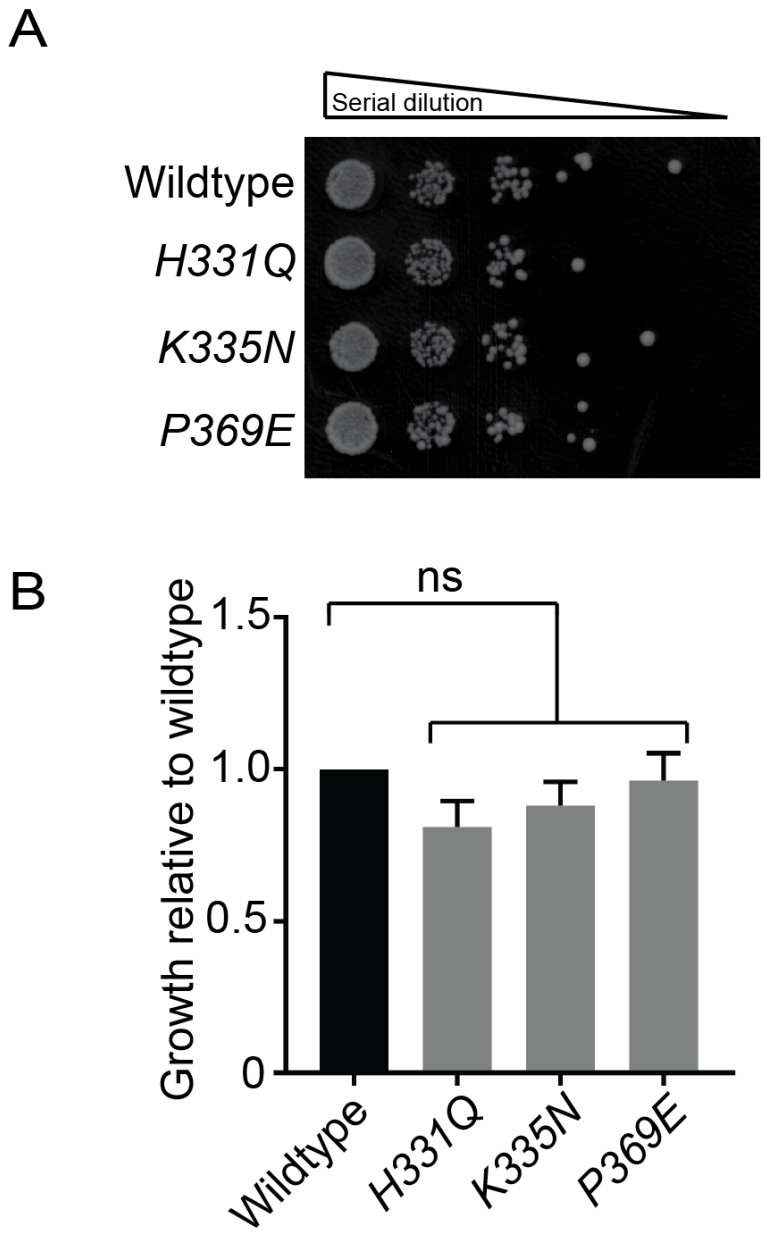

Figure 3. 2. Assessing viability in *HSH155* mutant strains.

Viability of strains were measured by (A) spot dilution assays on YPD plates at 30°C and (B) growth curves; For each replicate, 5 technical replicates were used to account for variability; Student's t-test, mean values with S.E.M. error bars are shown, *n = 3*.

Despite not showing obvious loss of viability, I wondered if the splicing defects observed in some of the mutants could impact chromosome instability phenotypes and R-loop accumulation, which I observed in Chapter 2 when using a temperature sensitive allele of HSH155. I used plasmid loss assay to measure chromosome instability and immunofluorescence probing for S9.6 antibody to detect levels of R-loops. The plasmid loss assay measures stable transmission of a CEN plasmid, one that contains a yeast centromere and therefore replicates like endogenous chromosomes. Therefore tracking the presence of the plasmid can be used as a proxy for assessing chromosome instability. The results show that the H331Q mutant strain has a higher frequency of plasmid loss when compared to wildtype (**Figure 3.3A**). However, when I used a qualitative A-like faker (ALF) assay to confirm these findings, the results indicate there does not appear to be a difference in chromosome instability frequency (data not shown). The ALF assay measures the stability of the *MAT* locus on chromosome III and loss of this locus leads to quantifiable aberrant mating events [126]. Detailed description of this assay is outlined in Section 2.3.4. The discrepancy in results could be due to the inherently high frequency of plasmid loss in yeast, as seen in the wildtype strain (~20% loss) (**Figure 3.3A**). In addition, the plasmid loss assay depends on the transformation process, although measures were taken to control for variations in plasmid copy number by using a CEN plasmid. CEN plasmids are maintained in the host cell with about one copy per haploid cell, however it has been shown that the plasmid copy number can vary across the cell population, where cells could either not carry any plasmid or carry more than one copy [174]. To measure R-loop accumulation, I used chromosome spreads to probe for S9.6 antibody (**Figure 3.3B**). The results indicate that, consistent with the *hsh155-1* R-loop data seen in Chapter 2, there does not appear to be an accumulation of R-loops in the *HSH155* cancer associated mutant strains. In Figure 3.3B, *rnh1*Δ*rnh201*Δ was used as a positive control.

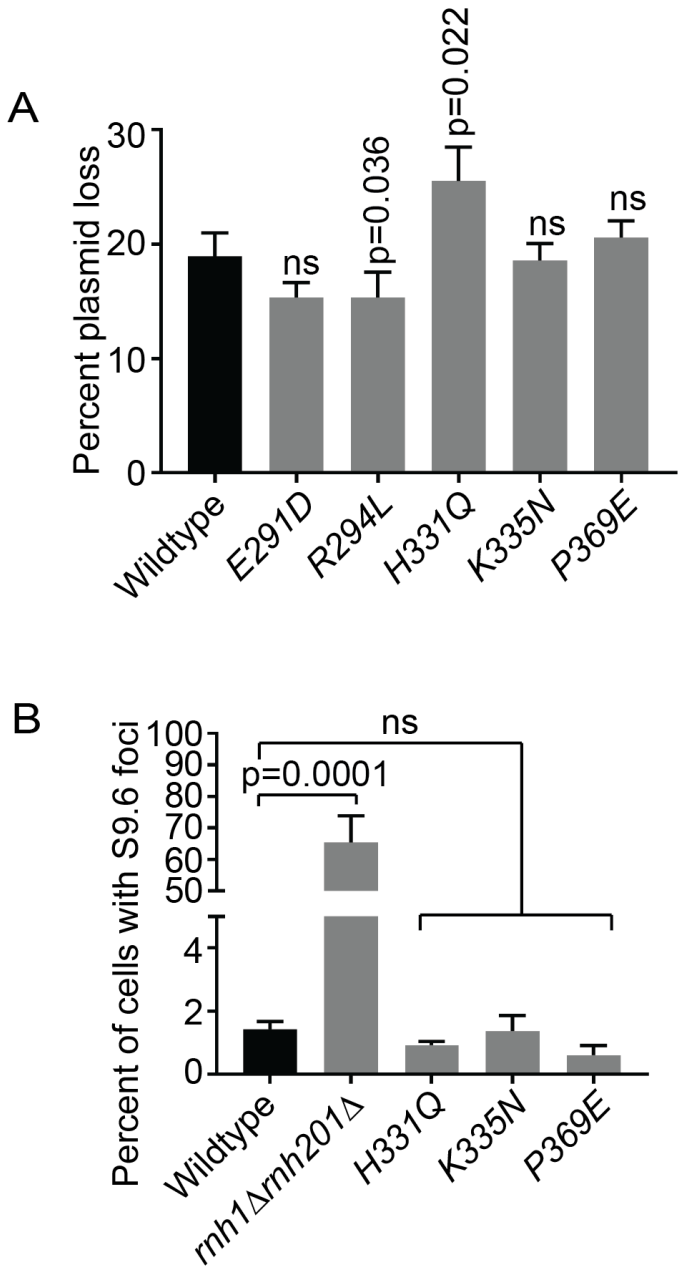

Figure 3. 3. Genome instability and R-loop detection in splicing mutants.

(A) Plasmid loss frequency of *pRS313::HIS3.* Approximately 150 colonies were scored for each replicate; Student's t-test was used to calculate statistical significance, *n = 9*. (B) s9.6 antibody staining intensities in chromosome spreads. >100 nuclei were measured per replicate; Fisher's exact test was used to calculate statistical significance, *n = 3*. For both panels, mean values with S.E.M. error bars are shown.

Despite the lack of strong observable phenotypes, I was curious as to why there were fluctuations in splicing ability as shown in Figure 3.1. Although RNA splicing is an essential process, previous work has shown that in yeast, deletion of introns from one third of intron-containing genes does not impact cell growth in unstressed conditions [175]. In addition, the same group showed that only three genes (*MTR2*, *YRA1*, and *TAD3*) required introns for complete gene function, and five introns in the entire yeast genome were essential for viability (four in *MTR2* and one in *ERV46*) [176], although many intron deletions cause minor phenotypes under different growth conditions using selective media [175]. If mutations in H331Q, K335N and P369E indeed cause splicing defects, what are the consequences of decreased splicing efficiency if not changes in cell growth or chromosome instability? Early studies have linked splicing to cell cycle progression [177, 178]. In these studies, conditional mutations in several yeast splicing factors indirectly cause cell cycle defects at the restrictive temperature. For example, mutations in the splicing factors *CEF1*, *PRP17*, *PRP22*, *ISY1* and *SYF2* lead to suboptimal expression of α-tubulin, resulting in cell cycle arrest phenotypes [70, 179]. Taken together with my data presented in Chapter 2, this indicates that in general splicing factor mutations impact cell cycle indirectly by disrupting expression of one or more genes required for cell cycle progression. However, the difference between these alleles and the three *HSH155* cancer-associated mutants studied thus far (H331Q, K335N and P369E) is that the mutants previously used are conditional temperature sensitive mutants that deplete protein levels or otherwise inactivate the respective splicing factor, while H331Q, K335N and P369E are point mutants of *HSH155* that do not deplete splicing efficiency to the same level as *hsh155-1*, suggesting that most of the Hsh155 protein activity in these strains is functional (**Figure 3.1**). Nevertheless, I wondered if the splicing defects observed in H331Q, K335N and P369E could lead to cell cycle arrest phenotypes since this appears to be a general consequence of splicing factor mutations in yeast. To assess cell cycle dynamics, I measured DNA content of asynchronous cultures using FACS analysis. The results show a small increase of 2N cells in the H331Q mutant strain relative to wildtype, potentially indicating a G2/M delay phenotype (**Figure 3.4**). Previously, the plasmid loss assay showed that H331Q mutants appear to have a higher frequency of chromosome instability relative to wildtype (**Figure 3.3A**). It is possible that the *HSH155* H331Q mutation exhibits a weak phenotype that is associated with subtle changes in splicing.

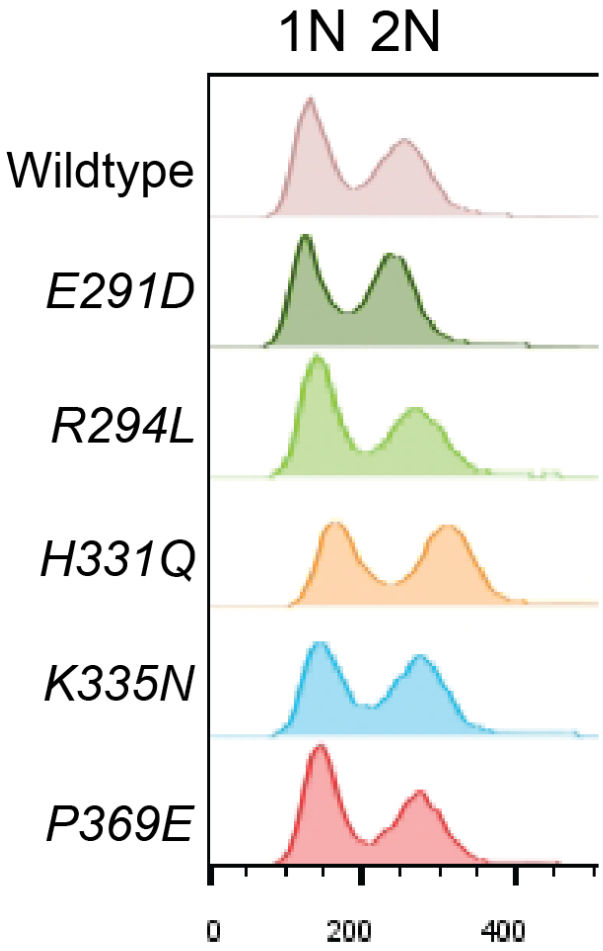

Figure 3. 4. Cell cycle dynamics in *HSH155*-cancer-associated mutants.
Representative image of plots generated from FACS analysis of DNA content (x axis, arbitrary units) in asynchronous populations of yeast cells. Peaks indicate 1N and 2N cells. 50,000 cells were collected per replicate; *n = 3* replicates showed similar trends.

3.3.3 Summary of findings in yeast Hsh155 cancer-associated mutants

Why do mutations in different *HSH155* residues result in different phenotypes in yeast when the mutations found in cancer cluster in these same HEAT repeats? To answer this question, it is important to first discuss the splicing changes that occur when these amino acids are altered. Other groups have utilized a copper reporter assay to monitor splicing activity in yeast to understand the consequences of cancer-associated *HSH155* mutations [161, 173, 180]. Briefly,

this reporter assay fuses an intron-containing fragment from yeast *ACT1* gene to *CUP1*, the protein product of which binds copper and confers resistance to high copper concentrations [181, 182]. By transforming this plasmid into the cancer-associated *HSH155* mutants, splicing efficiency can be correlated to levels of copper resistance in these strains. In addition, this reporter system can be modulated by mutating consensus splice sites within the *ACT1* intron, the 5' splice site (donor site), branch site, and 3' splice site (acceptor site) to produce suboptimal introns that are defective at different stages in the splicing pathway. Interestingly, using this system it was found that the cancer-associated *HSH155* mutations in yeast did not change splicing efficiency in the wildtype reporter system when the introns contained the consensus splice site, and instead the results indicate the *HSH155* mutations either improved or reduced splicing activity when the −2, −1 and +1 positions relative to the branchpoint adenosine was defective due to usage of an upstream cryptic branch site sequence [161, 180]. These findings indicate that the *HSH155* cancer-associated alleles might result in missplicing in response to specific intron features within the branch site sequence. It is interesting to note that while I used a different splicing efficiency reporter system, my findings are consistent with these two papers: the E291D variant increased splicing efficiency while the H331Q, K335N and P369E variants decreased splicing efficiency with no consequences to cell viability (**Figure 3.1**, **Figure 3.2**). It was also shown that when a single position in Hsh155 was mutated to different amino acids, this could also result in opposite splicing efficiency phenotypes [161]. Therefore it appears that while the *HSH155* cancer-associated alleles in yeast could affect branch point usage in general, the efficiency of splicing is dependent on the amino acid change. The goal of Chapter 3 was to understand how cancer-associated mutations in *SF3B1* impact genetic instability and tumourigenesis. However, my results indicate that while some of the *HSH155* cancer-associated alleles have splicing defects and the H331Q variant may exhibit weak cell cycle defects, this did not translate to a reproducible chromosome instability phenotype across assays. Transcriptome sequencing of mammalian cell lines and cancer patient samples revealed that *SF3B1* hotspot mutations in the HEAT repeats most frequently induce cryptic 3' splice site usage due to defects in recognizing the branch site sequence during splicing [78, 160, 162, 183, 184]. In yeast, it appears that the necessary sequence context is not present for cryptic 3' splice site usage to occur despite defects in branch site selection. Yeast transcriptome analyses indicate that the *HSH155* cancer-associated alleles likely do not affect 5' or 3' consensus or nonconsensus splice site usage, as the changes in 3' splice site usage in human *SF3B1*-mutant cells may be due to a defect

of the spliceosome in recognizing weak branch site sequences in human genes [161]. It has been proposed that the function of Hsh155/SF3B1 may be to allow the spliceosome to be less stringent in branch site sequence recognition when the sequence deviates from the consensus sequence, and is thus necessary for human introns which have poorly conserved splice sites [161]. Yeast introns on the other hand tend to have strong consensus sequences at the 5'/3' splice sites and branch site sequences and very few nonconsensus branch site sequences, potentially explaining why changes in alternative 3' splice site usage is not prevalent [185, 186]. Therefore, since the phenotypes observed in the *HSH155* cancer-associated yeast mutants were weak, and splicing changes between yeast and human cells appear to be different when *HSH155*/*SF3B1* is mutated, I opted to move on to characterize how cancer-associated variants in SF3B1 could affect cancer progression in human cells.

3.3.4 Genetic instability phenotypes in SF3B1-H662Q mutant human cells

Previous groups have used RNA sequencing data from different cancer types (MDS, CLL, breast cancer, skin melanoma, and uveal melanoma) to understand how hotspot mutations in the various SF3B1 HEAT repeat residues lead to aberrant splicing [159, 160, 163, 183]. These groups observed that the predominant splice aberrations resulting from *SF3B1* hotspot mutations were alternative 3' splice site usage, and additionally the sets of aberrantly spliced junctions appeared to cluster within each cancer type, although many overlapping events were also observed across cancer types. This suggests that tissue specific splicing regulation may contribute to these aberrant splicing patterns, although the functional consequences of these patterns are not known. To study the consequences of *SF3B1* mutations *in vivo*, pancreatic (Panc05.04) and leukemia (NALM6, a B-cell precursor leukemia cell line derived from a patient with acute lymphoblastic leukemia) cell lines carrying *SF3B1* hotspot mutations were generated by adeno-associated virus (AAV)-mediated homology to generate isogenic wildtype and mutant cell lines, and subsequent RNA sequencing analyses revealed the incidence of aberrant splice junctions is consistent with the data using *SF3B1* mutated cancer patient samples, confirming that both of these cell lines are representative models to study splicing activity [160]. To begin to understand if and how *SF3B1* mutations could contribute to tumourigenesis from the perspective of genome stability, I acquired two isogenic NALM6 cell lines from a commercial source for comparison, one with wildtype *SF3B1* while the other had a heterozygous knock-in of *SF3B1*

H662Q mutation. The reason for choosing this cell line and *SF3B1* mutant variant for my studies is twofold: first, as previously mentioned NALM6 *SF3B1*-mutated cell lines carrying hotspot mutations including H662Q was determined to be a representative model to study splicing activity, and second, my yeast *HSH155* data indicates the H662Q equivalent mutation, H331Q exhibits a weak phenotype that is associated with subtle changes in splicing.

To determine if mutations in *SF3B1* cause genetic instability phenotypes in these cell lines, I measured levels of DNA damage by monitoring phosphorylation of the histone H2A variant H2AX, resulting in the γH2AX marker. Upon DNA double strand breaks (DSBs), phosphorylation of H2AX at Serine-139 to γH2AX provides a signal for DNA repair proteins to accumulate, and serves as an early marker of DNA damage [187, 188]. Using immunofluorescence microscopy, regions of γH2AX foci localization can be used as a proxy for DNA DSBs, and the data indicates that the H662Q mutant had constitutively higher levels of γH2AX signal when compared to the isogenic wildtype cell line (**Figure 3.5A**). To confirm these results, I used the neutral and alkaline comet assays, a single cell gel electrophoresis method, to measure DNA strand breaks. The principle of the comet assay is based on the ability of denatured, broken DNA fragments to migrate more easily under electrophoresis, while undamaged DNA migrates more slower due to DNA supercoiling [189]. Therefore the length and shape of the comet tail relative to the intensity of the comet head (i.e. the remaining undamaged nuclear material) is used to evaluate levels of DNA damage. The neutral comet assay is used to detect DNA double strand breaks, and the alkaline comet assay is used to detect smaller amounts of DNA damage that includes both single and double strand breaks [190]. The results indicate that the H662Q mutant cell line had significant DNA damage compared to wildtype SF3B1 cells, as measured by the neutral (**Figure 3.5B**) and alkaline (**Figure 3.5C**) comet assays. Thus, multiple assays indicate H662Q mutation in NALM6 cells lead to a significant increase in DNA damage.

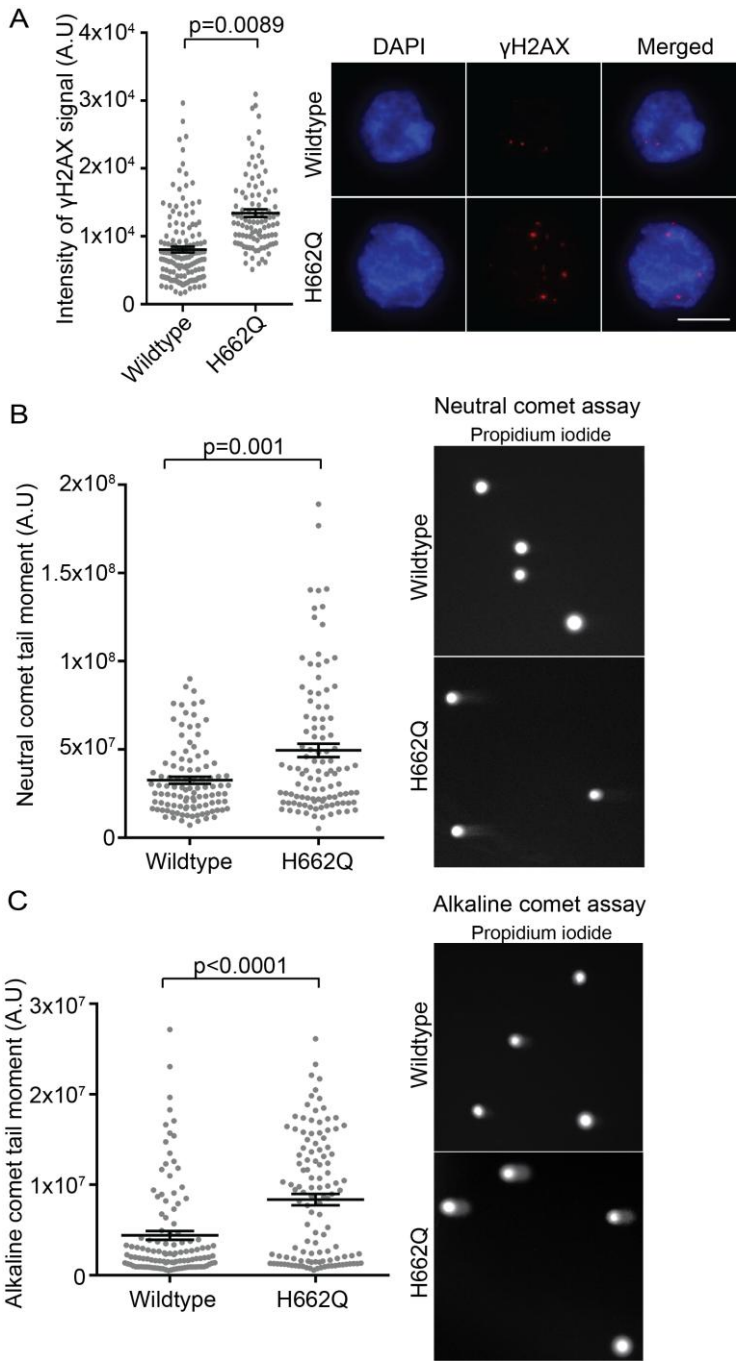

Figure 3. 5. Genetic instability phenotypes in SF3B1-H662Q mutant NALM6 cells.

(A) Immunofluorescence staining for γH2AX intensities. >100 nuclei were scored per replicate. (B) Neutral comet assay; propidium iodide was used as DNA stain. >50 cells were scored per replicate (C) Alkaline comet assay; propidium iodide was used as DNA stain. >50 cells were scored per replicate. (A-C) Student's t-test was used to calculate statistical significance. Mean values with S.E.M. error bars are shown, *n = 3*. Representative images are shown on the right. Scale bar = 5 μm.

3.3.5 Aberrant R-loop accumulation in SF3B1-H331Q mutant human cells

The DNA damage phenotypes observed in the H662Q mutant could arise by several mechanisms including the formation of R-loops [51, 59]. To test this model, I used the S9.6 antibody to detect and quantify the DNA:RNA hybrid part of the R-loops. The S9.6 antibody was developed by immunization with a DNA:RNA hybrid produced by RNA polymerase transcription of single stranded DNA from a bacteriophage, and was found to bind to DNA:RNA hybrids and not single or double stranded DNA, with lower affinities for structured RNA [191, 192]. By quantifying staining intensity of S9.6 in individual cells, I observed the H662Q mutant cell line had increased R-loop accumulation relative to the wildtype control (**Figure 3.6A**). To validate that the signal observed was indeed DNA:RNA hybrid accumulation, I transfected the cells with a plasmid overexpressing GFP or GFP-RNaseH1, a nuclease that specifically targets the RNA moiety of DNA:RNA hybrids [193]. S9.6 staining in cells with GFP or GFP-RNaseH1 showed significant decreases in hybrid levels when RNaseH1 was overexpressed (**Figure 3.6B**), indicating the observed signal is likely R-loop hybrids and that R-loop levels can be efficiently modulated with RNaseH1 overexpression in this cell line.

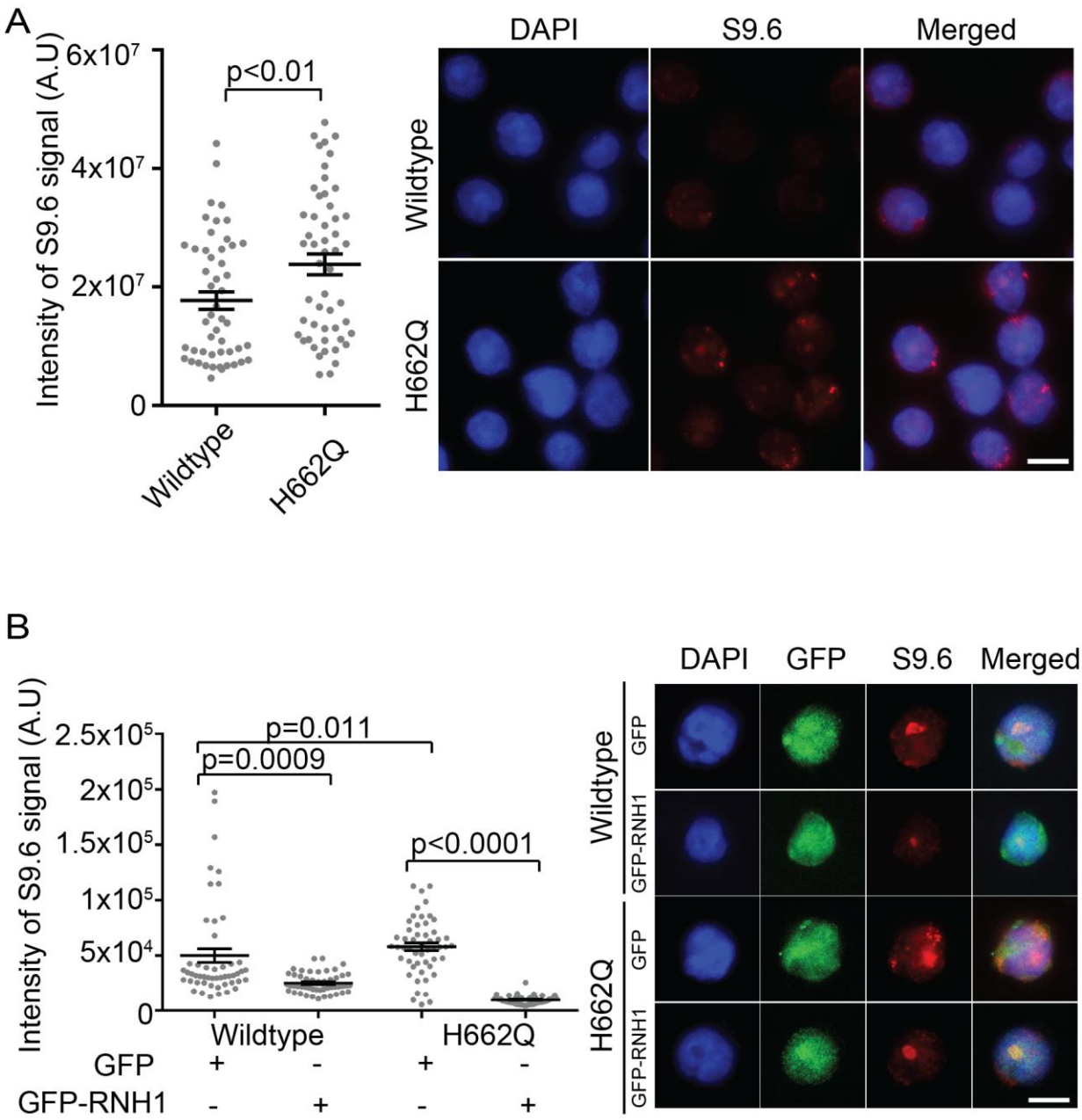

Figure 3. 6. R-loop accumulation in SF3B1-H662Q mutant NALM6 cells.

(A) Immunofluorescence staining for S9.6 intensity per nucleus. Student's t-test was used to calculate statistical significance. >100 nuclei were scored per replicate. (B) Immunofluorescence staining for S9.6 intensities with either a control vector (GFP) or one expressing GFP-RNaseH1 (GFP-RNH1). ANOVA was used to determine statistical significance. >100 GFP positive nuclei were scored per replicate. (A-B) Mean values with S.E.M. error bars are shown, *n = 3*. Representative images are shown on the right. Scale bar = 5 μm.

3.3.6 Testing for R-loop-mediated DNA damage and replication stress

To determine whether R-loop accumulation contributes to the observed DNA damage phenotypes, I overexpressed RNaseH1 to modulate R-loop levels, and quantified γH2AX signal and DNA double strand breaks using the neutral comet assay in asynchronous cells. The results indicate that when R-loops are resolved by RNaseH1 overexpression, the DNA damage phenotypes measured by γH2AX foci number (**Figure 3.7A**) and DNA double strand breaks using the neutral comet assay (**Figure 3.7B**) are slightly reduced in the H662Q mutant, but not in a statistically significant way compared to cells bearing the control GFP plasmid. This was somewhat surprising given the higher damage and S9.6 staining in the H662Q mutant cells and spurred me to consider alternative explanations.

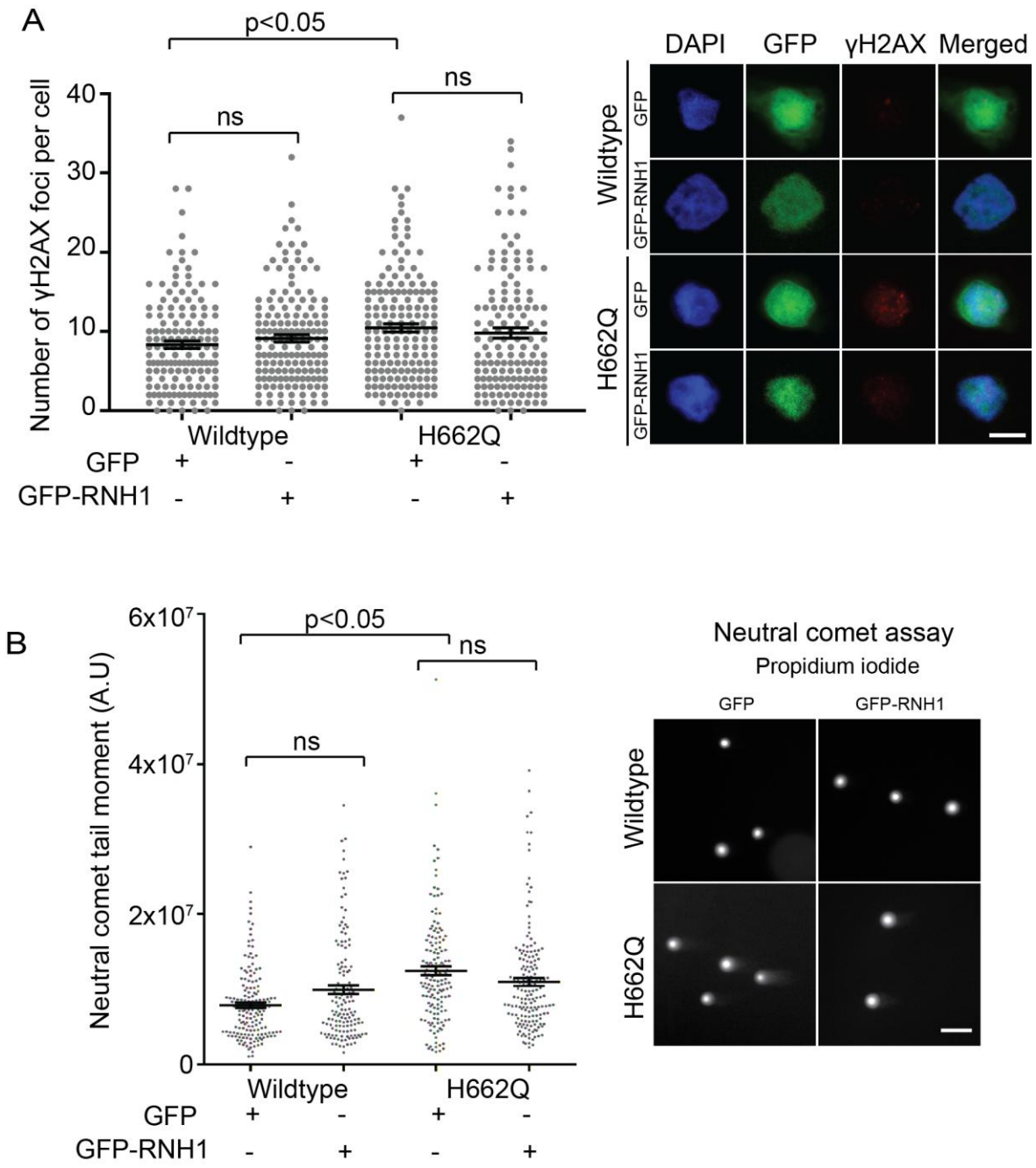

Figure 3. 7. R-loop removal partially suppresses genetic instability.

(A) Immunofluorescence staining for γH2AX foci number per cell with either a control vector (GFP) or one expressing GFP-RNaseH1 (GFP-RNH1). >100 GFP positive nuclei were scored per replicate. (B) Neutral comet assay of cells transfected with GFP control vector or GFP-RNaseH1; propidium iodide was used as DNA stain. >50 cells were scored per replicate. (A-B) ANOVA was used to determine statistical significance. Mean values with S.E.M. error bars are shown, *n = 3*. Representative images are shown on the right. Scale bar = 5 μm.

Previous studies in yeast and mammalian cells have correlated DNA replication with R-loop mediated genetic instability [194-196]. The current model in these studies states that aberrant co-transcriptional R-loop formation blocks replication fork progression during S phase, and if not resolved, could in turn result in genetic instability due to DNA breaks, fork collapse or incomplete DNA replication [197]. Additionally, it has been shown that active replication is required for R-loop mediated damage in HeLa cells depleted for the splicing factor SRSF1 [198]. Therefore to test if the observed R-loop accumulation in the H662Q mutant might be causing DNA damage in S phase, I conducted 5-ethynyl-2'-deoxyuridine (EdU) labeling and immunofluorescence to directly visualize cells undergoing DNA synthesis during the experiment. EdU is a nucleoside analog to thymidine, which is transported into dividing cells and incorporated into newly synthesized DNA, with subsequent detection by a fluorescent azide through copper-mediated click chemistry to an alkyne group on EdU [199, 200]. To verify that EdU negative and positive cells can be used as a marker for G1 and S/G2 cells, respectively, I co-stained with phosphorylated minichromosome maintenance protein 2 (MCM2) antibody to mark phases of the cell cycle (**Figure 3.8A**). DNA is replicated by a multi-protein machinery called the replisome, and the MCM2-7 helicase complex is necessary for ensuring initiation of replication [201]. MCM proteins are maintained throughout the cell cycle, but during transition from G1 to S phase, the MCM complex is activated through phosphorylation by DBF4/CDC7 kinase (DDK) and cyclin dependent kinases (Cdks) [202, 203]. Thus phosphorylation of MCM2 can be used as a marker of S phase entry and ongoing replication. The results indicate that EdU negative and positive cells indeed track with phosphorylated MCM2 levels (low levels – G1, high levels – S/G2), indicating this scoring method can be used to determine phases of the cell cycle (**Figure 3.8A**). Using this method, I co-stained EdU treated cells with γH2AX antibody to quantify γH2AX signal at G1 and S/G2 phases of the cell cycle. The results show that DNA damage, as measured by number of γH2AX foci, appears to predominantly occur during S/G2 phases of the cell cycle, in both wildtype and H662Q mutant cell lines, indicating that the baseline DNA damage levels in the NALM6 cell line occurs during S/G2 phase, and that the dramatic increase in DNA damage caused by H662Q mutation is associated with active replication (**Figure 3.8B**). Next, to determine if the observed increase in R-loop staining contributed to DNA damage during S phase, I overexpressed RNaseH1 to modulate R-loop levels, and quantified γH2AX signal after EdU treatment to determine cell cycle phase. The results indicate that, similar to data in Figure 3.7A, while DNA damage foci was reduced in

RNaseH-expressing H662Q mutant cells, the difference in damage was not statistically significant from cells transfected with the control GFP plasmid (**Figure 3.8C**). The results indicate that while R-loops may accumulate and account for some DNA damage in H662Q mutant cells, there must be other factors at play driving genetic instability.

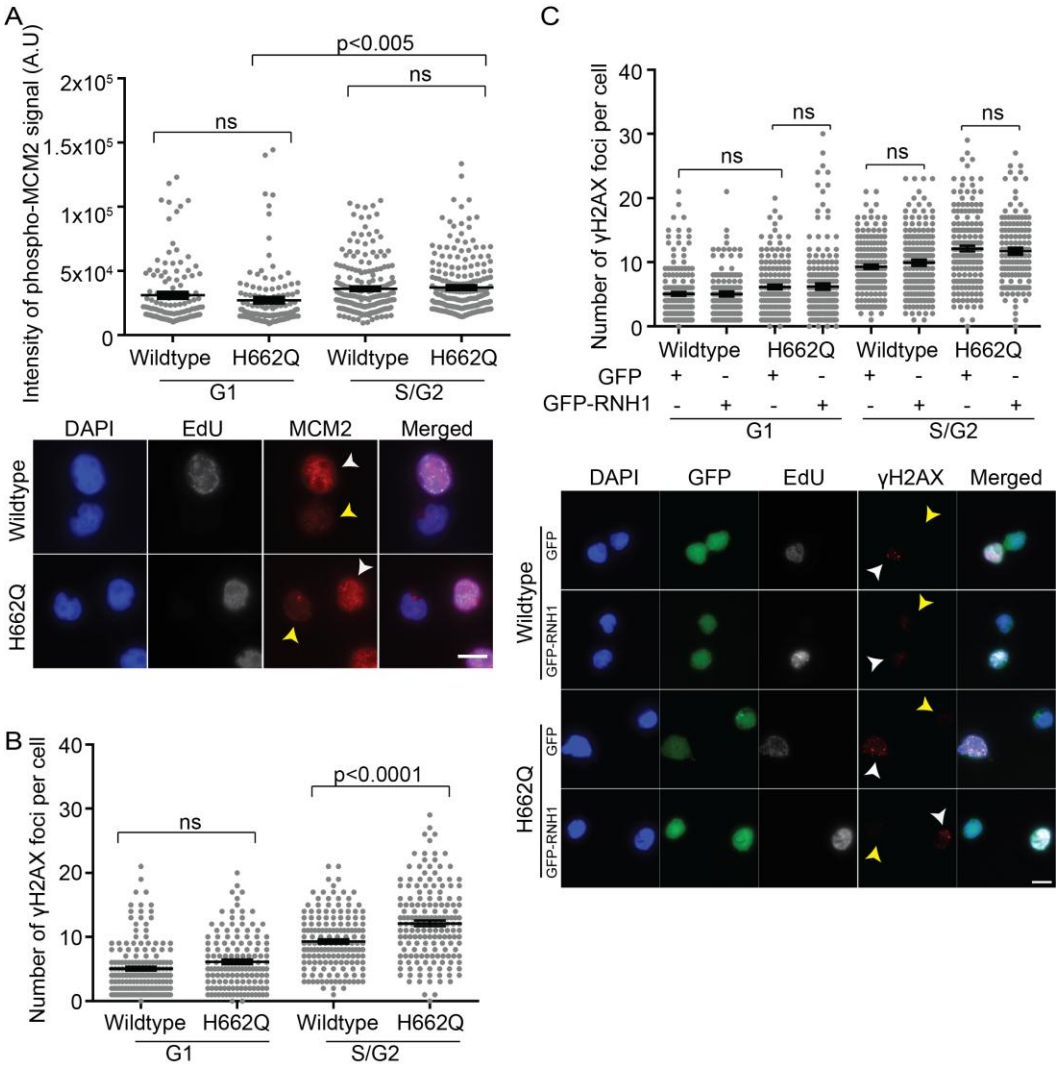

Figure 3. 8. DNA damage is S phase specific in SF3B1-H662Q mutant cells.
Immunofluorescence staining for (A) phosphorylated-MCM2 intensity per cell after EdU incorporation to determine cell cycle phase. >100 nuclei were scored per replicate. (B) γH2AX foci number per cell after EdU incorporation. >100 nuclei were scored per replicate. (C) γH2AX foci number per cell after EdU incorporation, with cells transfected with GFP control vector or GFP-RNaseH1. >100 GFP positive nuclei were scored per replicate. (A-C) ANOVA was used to determine statistical significance. Mean values with S.E.M. error bars are shown, *n = 3*. Representative images are shown below graphs. Scale bar = 5 μm. Yellow arrows indicate EdU negative (G1 phase) cells; White arrows indicate EdU positive (S/G2 phase) cells.

As discussed previously, R-loop formation could interfere with replication fork progression by causing transcription replication conflicts [198]. It has been shown that replication stress due to R-loop accumulation is largely caused by collisions between the transcription and replication machinery, particularly when these processes occur head-on [204]. Based on the observed S phase associated DNA damage in Figure 3.8, I decided to determine if replication stress occurs in H662Q mutant cells. I performed Western blots probing for phosphorylated -ATR, -Chk1, -ATM, and -Chk2 to measure activation of the DNA damage response. The results indicate that in the H662Q mutant cells, both ATR and ATM are activated, as assessed by increased levels of phosphorylated-ATR and ATM, however only Chk1 was activated downstream, therefore I can only conclude from these results that the ATR-CHK1 signaling pathway is activated in H662Q cells, consistent with increased replication stress (**Figure 3.9**). To confirm these results, I used Western blots to measure levels of phosphorylated RPA32 at the Serine-33 residue, which is specifically activated by ATR in response to replication stress and fork stalling [205, 206]. The results show that phosphorylation of RPA32-s33 is increased in H662Q mutant cells, suggesting replication stress and fork stalling, which is consistent with aberrant co-transcriptional R-loop formation blocking replication fork progression during S phase.

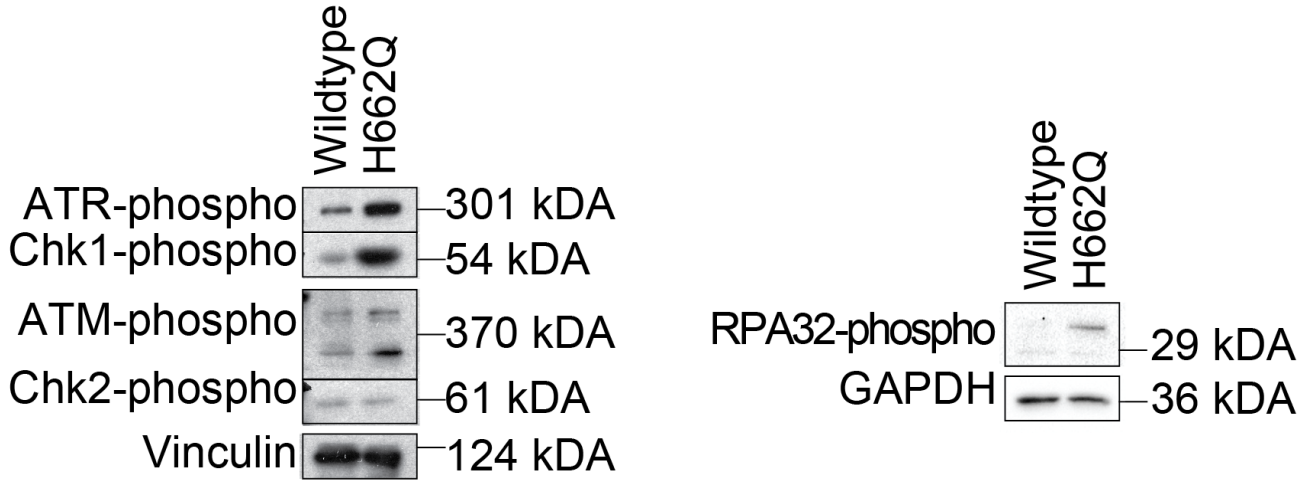

Figure 3. 9. Western blot probing for DNA damage response proteins.
Representative Western blots probing for phosphorylated-ATR, phosphorylated-Chk1, phosphorylated-ATM, phosphorylated-Chk2, and phosphorylated-RPA32. Vinculin or GAPDH were used as loading controls. *n = 3*.

Next, I wondered if the phosphorylated RPA32-s33 signal is enhanced only during S phase. Previous work examining cell cycle dependence of RPA32-s33 phosphorylation has shown that RPA32-s33 levels rise after cells enter S phase [206]. This group showed that cells in G1 have low levels of phosphorylated RPA32-s33, and higher levels in S and G2 phases of the cell cycle, suggesting that phosphorylation is an indicator of stalled DNA replication forks during normal S phase progression in unstressed cells. Indeed, when I performed immunofluorescence experiments co-staining EdU treated cells with phosphorylated RPA32-s33 antibody to quantify RPA32 signal at G1 and S/G2 phases of the cell cycle, wildtype cells had a small but statistically significant increase of RPA32 foci number in S/G2 phase (**Figure 3.10A**). This increase in RPA32 signal between G1 and S/G2 cells was also observed in the H662Q mutant cells, however this increase was much greater, suggesting enhanced replication stress and fork stalling in these cells, which is consistent with the Western blot data (**Figure 3.9**). The data also shows that phosphorylated RPA32-s33 accumulates in the H662Q mutant cells relative to wildtype, regardless of cell cycle phase (**Figure 3.9**). This is surprising, because RPA32 signal is thought to accumulate during S phase when single stranded DNA is more readily available to be bound. Previously, two studies using DT40 and human cell lines observed RPA foci formation in G1 phase cells in response to ionizing radiation (IR) [207, 208]. It is not immediately clear how single stranded DNA is generated for RPA32 to bind during G1 phase, however there is speculation that single stranded DNA could result from actively transcribed sites or during DNA repair [208, 209]. During mitosis, many transcription associated proteins (including RNA polymerase II) are excluded from chromatin and transcription is low, therefore upon G1 entry, transcription is rapidly increased to make proteins necessary for S phase and DNA replication [210, 211]. This increase could result in transcriptional stress, and in cells that have defects in two processes that occur co-transcriptionally – RNA splicing and R-loop formation – perhaps there is too much transcriptional stress for the cell to process properly. More likely, the RPA signal is a result of R-loop accumulation. RPA coated single stranded DNA has been shown to be present at both R-loops and replication forks [212], and it has been recently suggested that RPA is a sensor of R-loops in all cell cycle phases [213]. Indeed, when I overexpressed RNaseH1 to modulate R-loop levels and quantified phosphorylated RPA32-s33 signal after EdU treatment, I saw a dramatic decrease in foci number in the H662Q mutant cells, in all phases of the cell cycle (**Figure 3.10B**). Taken together the data indicates that aberrant R-loops formed in

the H662Q mutant cells cause replication stress, the exposure of ssDNA marked by phosphorylated-RPA32, and potentially transcription replication conflicts, ultimately driving genetic instability.

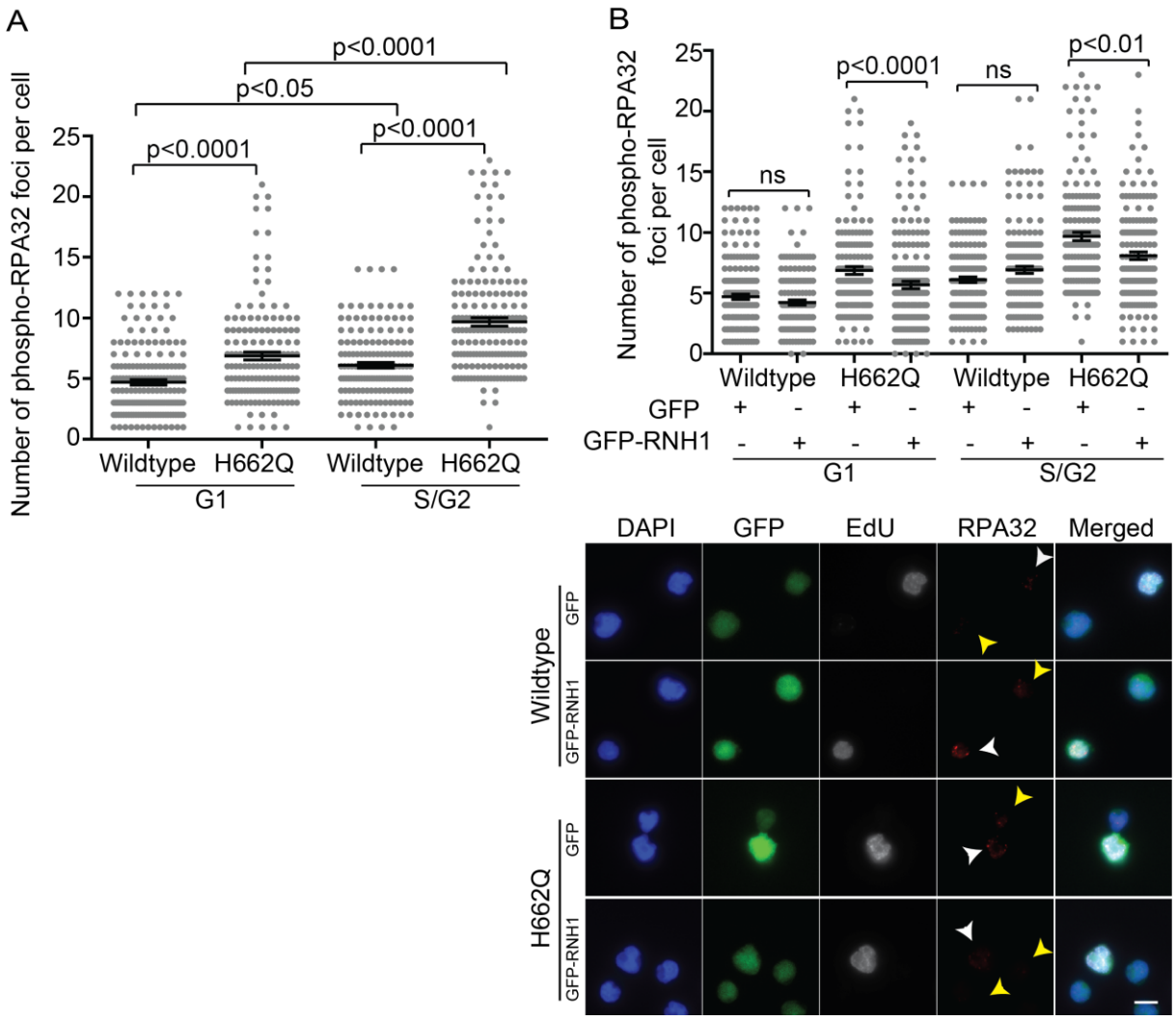

Figure 3. 10. Replication stress is suppressed by R-loop removal.

Immunofluorescence staining for (A) phosphorylated-RPA32 foci number per cell after EdU incorporation. >100 nuclei were scored per replicate. (B) phosphorylated-RPA32 signal after EdU incorporation, with GFP or GFP-RNH1 overexpression, with cells transfected with GFP control vector or GFP-RNaseH1. >100 GFP positive nuclei were scored per replicate. Representative images are shown below graph. Scale bar = 5 μm. Yellow arrows indicate EdU negative (G1 phase) cells; White arrows indicate EdU positive (S/G2 phase) cells. (A-B) ANOVA was used to determine statistical significance. Mean values with S.E.M. error bars are shown, *n = 3*.

3.3.7 Replication defects in SF3B1 mutant cells

The data thus far indicate that mutations in H662Q lead to increased R-loops and associated DNA replication stress. To gain further insight into the consequences of persistent replication stress in these cells, I analyzed cell proliferation over 72 hours and cell cycle progression using propidium iodide to stain DNA and EdU incorporation to monitor S phase cells. The results show that H662Q mutant cells have problems in cell proliferation when compared to wildtype (**Figure 3.11A**), and FACs analysis revealed there was a statistically significant increase of S phase cells when H662Q is mutated (**Figure 3.11B**). The results indicate that the H662Q mutation leads to defective cell proliferation and S phase progression, which is consistent with replication stress phenotypes.

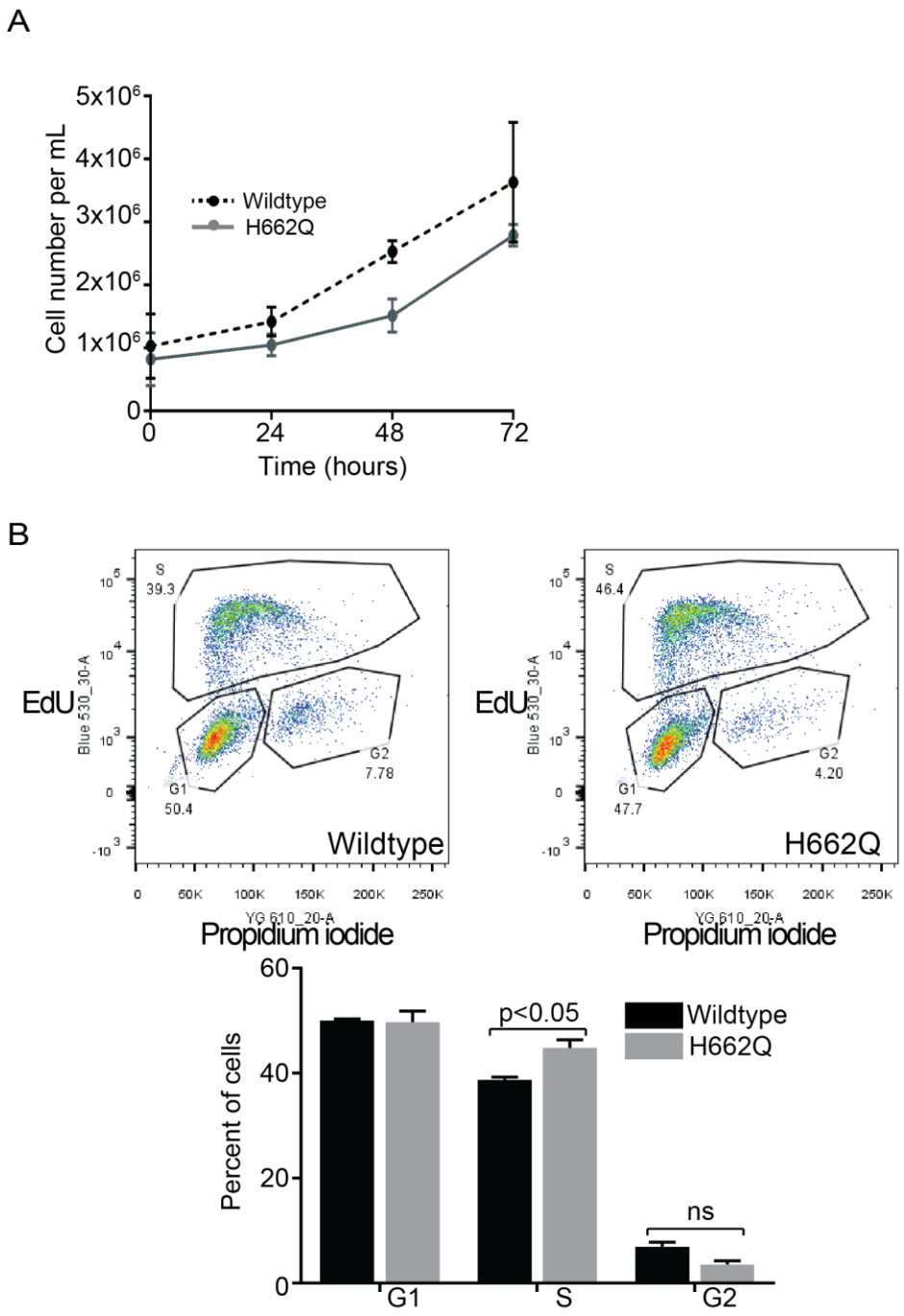

Figure 3. 11. Replication defects in SF3B1 H662Q cells.

(A) Cell doubling capacity measured over 72 hours. Mean values with S.E.M. error bars are shown, *n = 3*. (B) Cell cycle progression using FACS measuring propidium iodide staining for DNA and EdU incorporation for replicating cells. Top: Representative FACs plots of wildtype (left) and SF3B1 H662Q mutant (right). Bottom: Bar graph showing quantification of FACs plots. Student's t-test was used to calculate statistical significance. Mean values with S.E.M. error bars are shown, *n = 3*. 10,000 cells were analyzed per replicate.

3.3.8 Replication inhibition suppresses DNA damage and replication stress

Replication defects appear to play a central role in the phenotypes observed when H662Q is mutated. First, I observed that H662Q mutant cells accumulate DNA damage specifically in S phase. Second, H662Q mutant cells have elevated R-loops, replication stress, and ATR-CHK1 activation. Lastly, cell proliferation is affected and cells appear to be stalled or slowed in S phase. Therefore, I wondered if further inhibiting replication could suppress genetic instability observed in the mutant cells. I inhibited replication with the CDC7 inhibitor PHA-767491, which has been shown to be a potent inhibitor of CDC7, with minor selectivity for other kinases [214]. As previously mentioned in Section 3.3.4, CDC7 kinase is activated by binding to its regulatory protein DBF4, which then phosphorylates MCM2 proteins to initiate DNA synthesis [215]. Although the IC50 values for CDC7 inhibitor PHA-767491 (subsequently referred to as "CDC7i") has been determined for a variety of cell lines [214], I wanted to use a dose that could temporarily prevent S phase entry without manifesting acute toxic effects on viability, since the goal was to test if DNA damage could be suppressed. By treating cells with increasing doses of CDC7i treatment, I found that at a concentration of up to 5 μM CDC7i could inhibit cell proliferation (**Figure 3.12A**), while maintaining cell viability (**Figure 3.12B**).

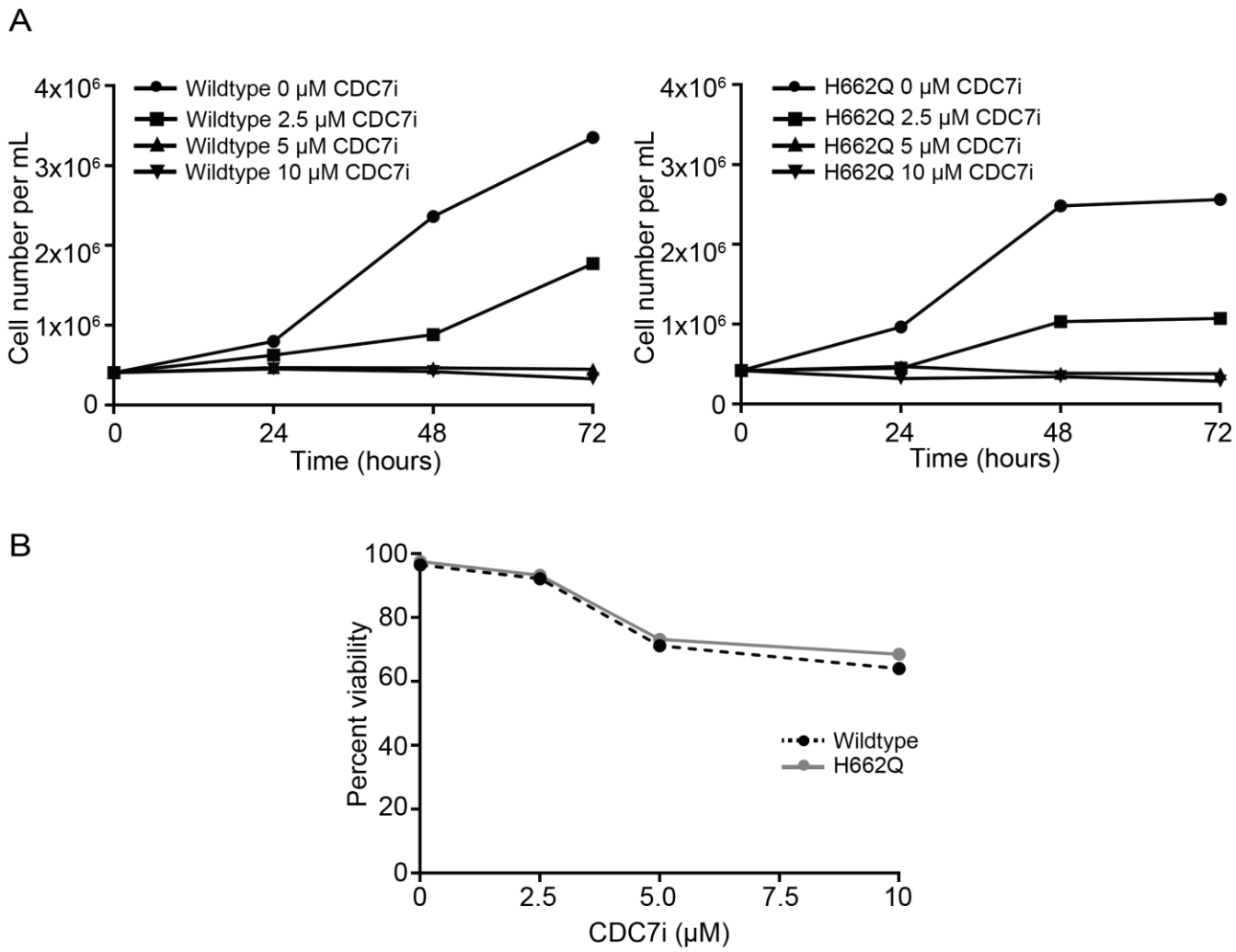

Figure 3. 12. Optimization of CDC7 inhibitor dose.

(A) Cell doubling capacity of NALM6 wildtype (left) and NALM6 SF3B1-H662Q (right) cells at the indicated CDC7i doses across 72 hours. (B) Cell viability at 0, 2.5, 5, and 10 μM of CDC7i treatment. Only 1 replicate was done to ensure replication inhibition and to establish dose for subsequent experiments.

To verify that CDC7 is indeed inhibited, I performed Western blot analysis probing for CDC7 and phosphorylated MCM2 at increasing doses of CDC7i. The data show that CDC7 and phosphorylated MCM2 levels decrease in a dose dependent manner, while the loading controls MCM2 total protein and GAPDH remain unchanged (**Figure 3.13A**). Therefore the decreased cell proliferation capacity observed in Figure 3.12A is due to inhibition of CDC7 activity leading to replication inhibition.

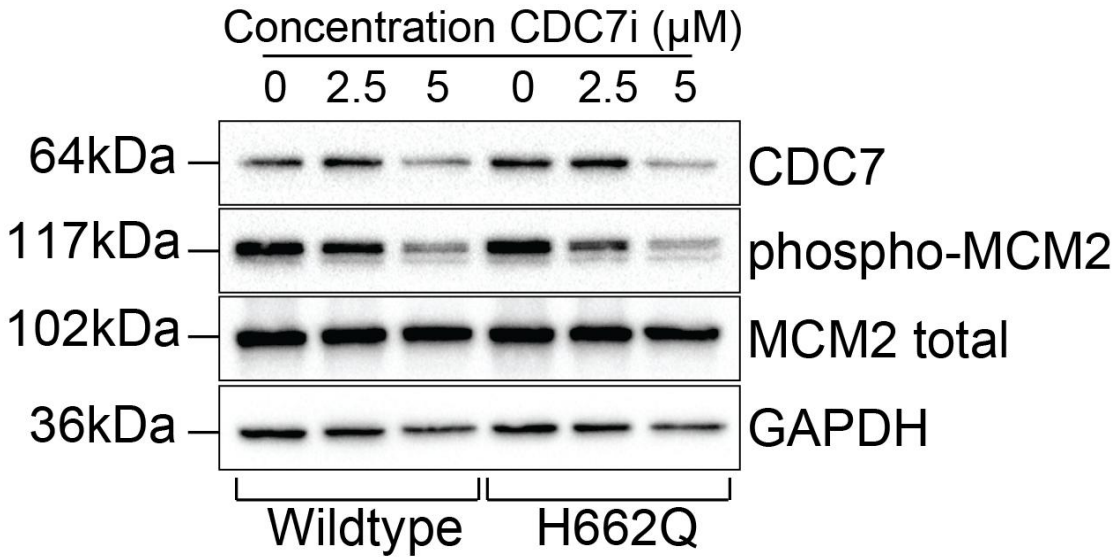

Figure 3. 13. Western blot verification of replication inhibition after CDC7i treatment. Representative Western blot probing for CDC7, phosphorylated-MCM2, total MCM2, after treatment with the indicated CDC7i doses. GAPDH served as a loading control.

To test if replication inhibition can suppress replication stress and DNA damage in H662Q mutant cells, I treated cells with increasing CDC7i doses and measured phosphorylated RPA32-s33 protein levels with Western blot and DNA damage using neutral comet assay. I observed a reduction in phosphorylated RPA32-s33 protein levels (**Figure 3.14A**), and reduction in DNA damage to wildtype levels (**Figure 3.14B**) in the H662Q mutant cells, indicating active replication is required for replication stress and DNA damage when H662Q is mutated. Taken together the data indicates that DNA damage is primarily driven by DNA replication stress in H662Q mutant cells.

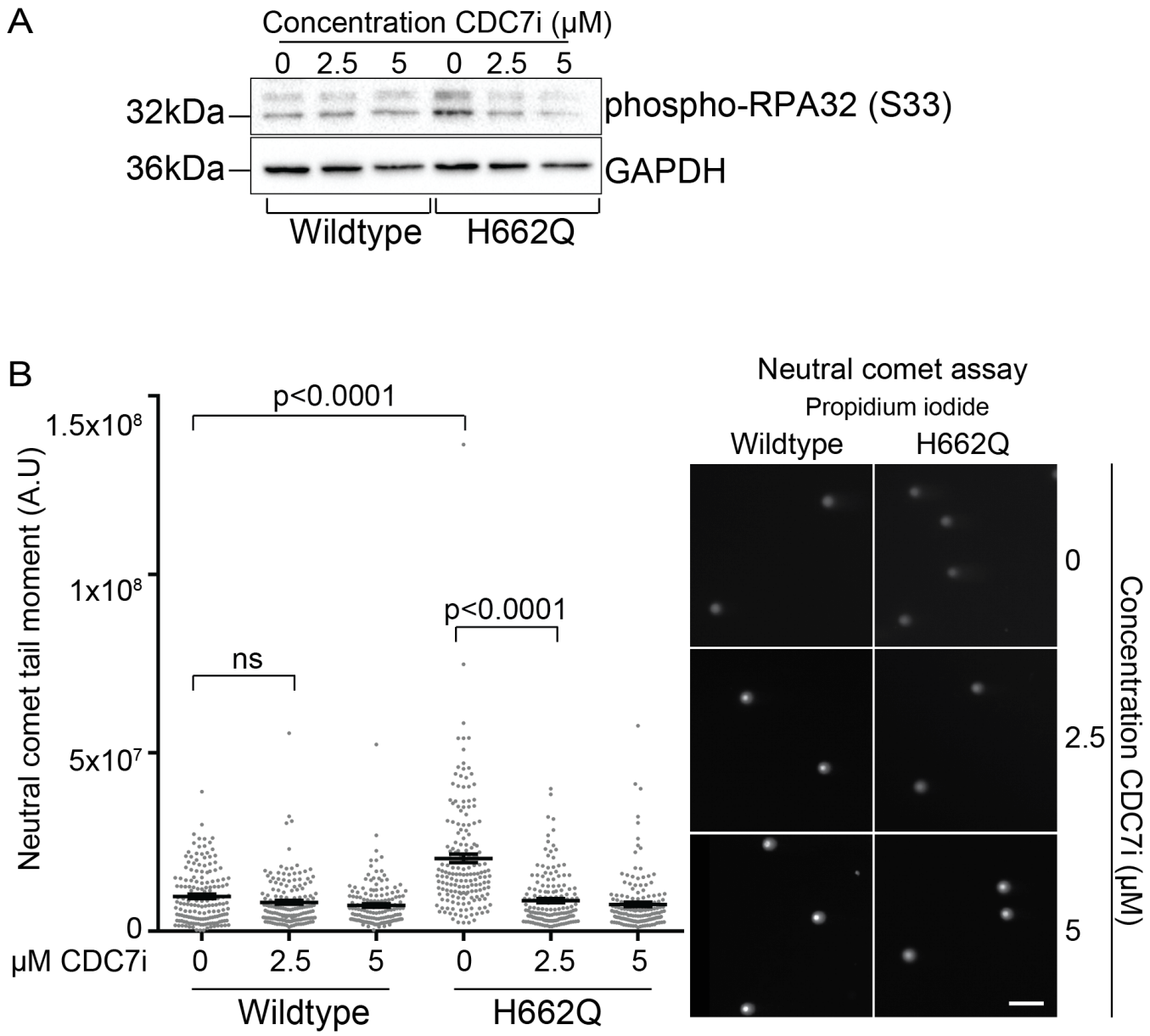

Figure 3. 14. Replication inhibition suppresses replication stress and DNA damage.
(A) Representative Western blots probing for phosphorylated-RPA32 after treatment with the indicated CDC7i doses. GAPDH served as a loading control. (B) Neutral comet assay of cells after treatment with the indicated CDC7i doses; propidium iodide was used as DNA stain. >50 cells were scored per replicate. ANOVA was used to determine statistical significance. Mean values with S.E.M. error bars are shown, *n = 3*. Representative images are shown on the right. Scale bar = 5 μm.

Previously, modulating R-loop levels by RNaseH1 overexpression suppressed replication stress, but did not significantly suppress DNA damage as measured by γH2AX signal and neutral comet assay. However upon replication inhibition, both replication stress and DNA damage phenotypes were suppressed in the H662Q mutant cells. Current models of R-loop mediated DNA damage focuses on how transcription replication conflicts lead to stalled or collapsed forks, resulting in genetic instability [197]. However genetic instability results from complex interplay between DNA damage and the subsequent DNA repair processes. Indeed, in fission yeast it has been shown that when the splicing factor Nrl1 is depleted, aberrant R-loop accumulation or their associated DNA damage sequesters DNA repair proteins from exogenous DNA damage lesions, leading to overall DNA repair defects in these mutants [216]. In addition, work in budding yeast revealed that persistent R-loops mediated through RNaseH depletion causes DNA damage that is repaired by break-induced replication (BIR), a pathway of homologous recombination (HR) that repairs one ended DNA breaks in G2/M phase of the cell cycle [217]. BIR contributes to replication fork repair, and is associated with high levels of mutagenesis, for example the frequency of frameshift mutations associated with BIR is 1000 times higher than in normal S phase replication [218, 219]. Therefore multiple factors could contribute to genetic instability observed in cells that have aberrant R-loop accumulation. While my results indicate R-loops indeed cause replication stress and might partially contribute to overall DNA damage observed in H662Q mutant cells, it is likely that other factors, potentially changes in gene expression due to defective splicing, contribute to DNA damage and the repair of R-loop mediated damage.

3.3.9 DYNLL1 is commonly misspliced in SF3B1 mutant cells

Since R-loop removal by RNaseH1 overexpression failed to suppress DNA damage phenotypes in the H662Q mutant cells, I wondered if an R-loop-independent mechanism was driving genetic instability as in my yeast studies (Chapter 2). It is possible that specific or global missplicing of transcripts could be driving genetic instability in *SF3B1* mutant cells. Previous groups have used RNA sequencing data from different *SF3B1* mutated cancers to identify changes in splicing pattern [159, 160, 163, 183]. These groups observed that the predominant splice aberrations resulting from *SF3B1* hotspot mutations were alternative 3’ splice site usage when cryptic 3’ splice sites are found between the branch point and canonical 3’ splice site.

While these aberrant splice events appear to be common amongst the different cancer types, the exact mechanism by which these hotspot mutations alter SF3B1 interactions with RNA, components of the U2 snRNP and other proteins remains unclear. In addition, it is not well understood what the functional consequences are when these gene transcripts are misspliced, and further studies are required to understand how cryptic 3' splice site selection drive oncogenesis in *SF3B1* mutated cancers. Although it is not entirely clear how aberrant splicing contributes to cancer progression, many groups performing transcriptome analyses have revealed a wealth of information identifying genes that are aberrantly spliced when *SF3B1* is mutated. Mining the literature, I manually curated a list of common aberrantly spliced genes from these papers (**Table 3.1**). Although many genes were identified as misspliced when *SF3B1* is mutated in each model, there appears to be common hits regardless of the cell type tested.

Cell type	Misspliced genes						References
	ABCB7	*DYNLL1*	*TMEM14C*	*SEPT6*	*ABCC5*	Others	
Primary MDS, CD34+ bone marrow	+	+		+	+	*LARP4* *METTL5* *SEPT2* *DDX24* *ERCC3* *FANCI* *SETX* *ATR* *FOXRED1*	[220]
Myeloid cell lines: K562, HEL, TF1, and SKM1 +siRNA knockdown	+					*CDC7* *SRSF11* *TP53*	[221]
Primary MDS, CD34+ bone marrow	+	+	+	+	+	*ENOSF1* *BRD9* *HINT2*	[171, 221]
Primary uveal melanoma					+	*GUSBP11* *UQCC* *ANKHD1* *GAS8* *F8* *CRNDE*	[159]
Primary breast cancer		+	+		+	*RPL31* *ICA1* *RPL24* *CRNDE* *MTERFD3* *UQCC* *F8* *GUSBP11*	[164]
Primary samples: CLL, breast cancer, skin melanoma, uveal melanoma, MDS RARS Cell lines: Panc05.04, NALM6	+	+	+	+	+	Many	[160]
Primary CLL		+		+		*DVL2* *DNAJC3* *TRIP12* *HDAC7* *CHD1L* *RAD9A*	[79]

Table 3. 1 Summary of commonly misspliced genes in *SF3B1* mutant cells
+ indicates positive identification of aberrant splicing in these genes

The data summarized in Table 3.1 shows that commonly misspliced transcripts are: *ABCB7, ABCC5, DYNLL1, TMEM14C* and *SEPT6*. ABCB7 and ABCC5 are members of the superfamily of ATP-binding cassette (ABC) transporters, that have roles in iron and cyclic nucleotide transport, respectively [222]. DYNLL1 (Dynein light chain 1) was originally identified as a light chain of the dynein motor complex with a role in dynein assembly [223], but has subsequently been found to have a regulatory role in protein dimerization of many targets that could affect transcription, DNA damage response, apoptosis, and cell migration [224-227]. TMEM14C (transmembrane protein 14C) is required for mitochondrial heme biosynthesis [228], and SEPT6 (septin 6) is a GTPase septin subunit that has roles in cytokinesis [229]. I was particularly interested in DYNLL1, not only because it was commonly misspliced in a range of *SF3B1* mutated cancers and cell lines, but also because previous work has shown that *DYNL11* appears to be aberrantly spliced in different cell populations derived from the same samples [220]. The authors determined overlapping aberrant splicing events in granulocytic, monocytic, and erythroid precursors and hematopoietic stem (CD34+) cells in *SF3B1* mutant MDS patient cells, and while many aberrant splicing events were common between CD34+ cells and one or more of the other cell populations, *DYNLL1* was one of the few aberrantly spliced genes that was in common to all four cell populations [220]. Most importantly, DYNLL1 protein has tangible links to DNA repair and genome stability (see 3.39 below). To first determine if *DYNLL1* is aberrantly spliced in the NALM6 H662Q mutant cells, I used end point reverse transcription polymerase chain reaction (RT-PCR) to obtain amplification products that spanned the predicted spliced-exon junctions identified through transcriptome analysis by Dolatshad *et al* [171]. I ran the PCR products on a 2% agarose gel to visualize aberrantly spliced isoforms in relation to the canonical isoform, and the results indicate that *DYNLL1* is indeed aberrantly spliced in NALM6 H662Q mutant cells (longer 5'UTR) (**Figure 3.15**). Each of the lanes represent cells extracted from individual cultures, that were subsequently extracted for RNA and converted to cDNA before PCR amplification using primers designed by Dolatshad *et al* [171]. PCR products were subsequently sequence verified for the presence of the canonical and aberrant splice isoforms, confirming the presence of the longer 5'UTR in the SF3B1 mutant cells.

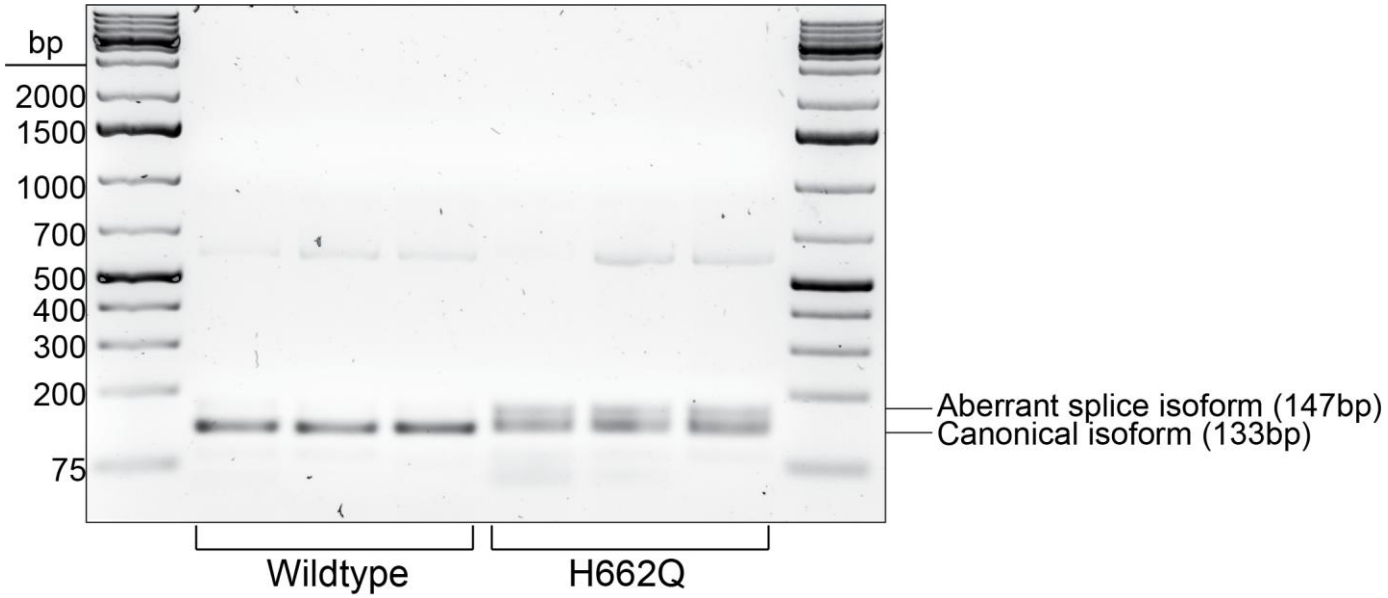

Figure 3. 15. *DYNLL1* is misspliced in SF3B1-H662Q mutant cells.
End point RT-PCR amplication products visualized on 2% agarose gel with SYBR™ Safe (Invitrogen). Three independent replicates are visualized per cell line.

Consistent with the literature, the results indicate that *DYNLL1* is aberrantly spliced, confirming this is a commonly misspliced transcript. Next I wondered if aberrant splicing of *DYNLL1* could lead to up or downregulation of the protein product. Bioinformatics analysis using NALM6 SF3B1-K700E mutant cell lines predicted that 44% of all aberrant mRNAs result in a transcript that is sensitive to nonsense mediated decay (NMD), and additional RNA sequencing analysis revealed many NMD-sensitive genes are downregulated, suggesting these transcripts are actively degraded [160, 230]. SILAC experiments with mass spectrometry used to quantify protein levels also indicate that many misspliced transcripts for which peptides could be detected were found to be decreased in the mutant cells, including ABCB7, and SEPT6, although DYNLL1 was not identified in the analysis [160]. To determine if DYNLL1 protein expression is modulated upon aberrant splicing, I used Western blots to assess changes in protein levels. The results show that when H662Q is mutated, DYNLL1 protein levels are upregulated relative to the wildtype levels (**Figure 3.16**). The RNA and protein expression analysis (end point RT-PCR and Western blots) indicate *DYNLL1* is aberrantly spliced, leading to upregulation of protein in H662Q mutant cells.

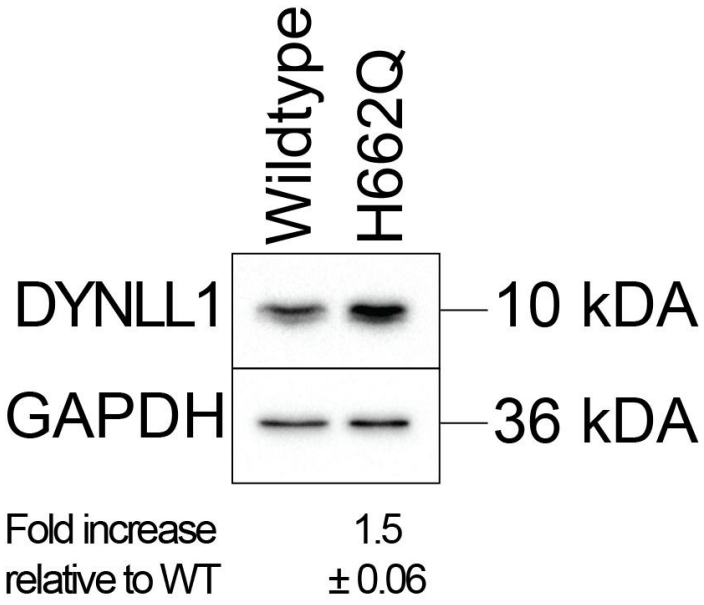

Figure 3. 16. DYNLL1 protein level is upregulated in SF3B1-H662Q mutant cells. Representative Western blot probing for DYNLL1. GAPDH served as a loading control. Band intensities were quantified using ImageJ.

3.3.10 Reducing DYNLL1 protein levels mitigates DNA damage in SF3B1 H662Q cells

DYNLL1 (dynein light chain 1) was originally identified as a light chain of the dynein motor complex with a role in dynein assembly [223]. Subsequently, many groups have discovered that DYNLL1 has key regulatory roles in protein dimerization of targets that include transcription factors, DNA damage response proteins, regulators of apoptosis, and many other processes [224-227, 231]. DYNLL1 is regulated by its transcriptional activator ASCIZ (ATM Substrate Chk2-interacting Zn2+ Finger), also known as ATMIN (ATM interactor), and high levels of DYNLL1 in turn inhibit the transcriptional activity of ASCIZ, leading to negative autoregulation of gene expression [232]. Studies have identified a role of ASCIZ in the DNA damage response, specifically base excision repair pathway [233], and oxidative stress response [234]. More recently, it was found that ASCIZ and DYNLL1 overexpression is required for MYC-driven development of lymphoma using a B cell leukemia/lymphoma mouse model [235]. To my knowledge nothing is known about the functional consequences of aberrant DYNLL1 splicing in

SF3B1 mutated cells, and I was curious if the upregulation of DYNLL1 protein contributes to the genetic instability phenotype I observed in the H662Q mutant cells. To assess the contribution of DYNLL1 overexpression on genetic instability, I used siRNA knockdown of DYNLL1 to decrease protein levels, and used the neutral comet assay to measure DNA damage. The results indicate that siRNA knockdown of DYNLL1 lead to suppression of DNA damage in the H662Q mutant cells to wildtype levels, while damage remained unchanged in wildtype cells (**Figure 3.17**). The data thus far reveals that when the SF3B1 H662Q residue is mutated, this leads to upregulation of DYNLL1 protein levels mediated by aberrant splicing of the transcript, and upon downregulation of DYNLL1, genetic instability is suppressed.

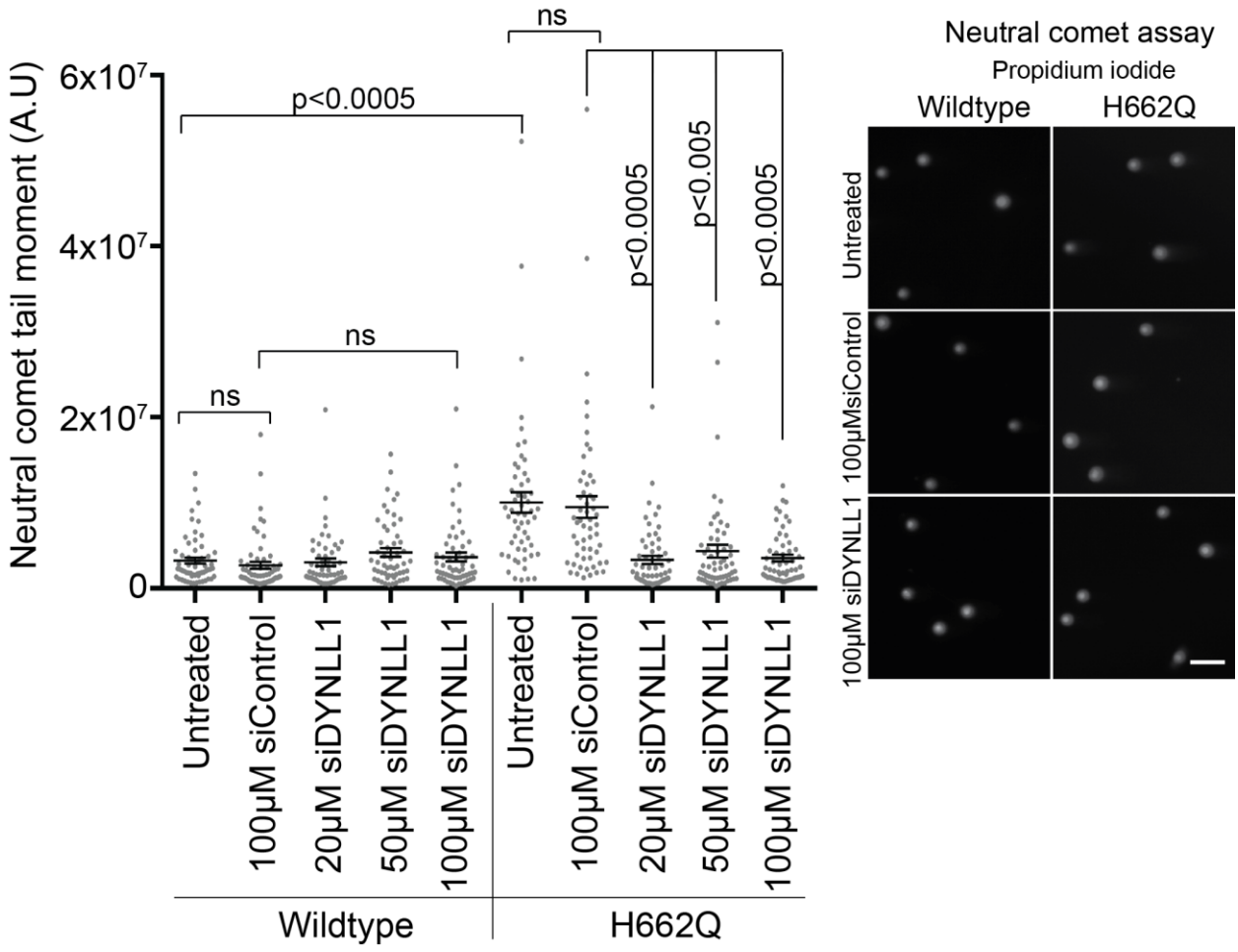

Figure 3. 17. DYNLL1 knock-down suppresses DNA damage.

Neutral comet assay following siRNA treatment with 100 μM si-Control and 20, 50, and 100 μM si-DYNLL1. >50 cells were scored per replicate. ANOVA was used to determine statistical significance. Mean values with S.E.M. error bars are shown, *n = 3*. Representative images are shown on the right. Scale bar = 5 μm.

Before further analysis, I wanted to verify that the suppression of DNA damage was not due to loss of viability and cell death upon si-DYNLL1 knockdown. I measured cell viability 48 hours after siRNA treatment with increasing concentrations of si-DYNLL1, and the results indicate that siRNA treatment did not affect viability of wildtype or H662Q mutant cells when compared to cells treated with the same concentration of si-Control (**Figure 3.18**). This is consistent with what is known about the siRNA depletion of DYNLL1 – that siRNA-mediated knockdown of DYNLL1 does not affect cell viability in a variety of cell lines whether *SF3B1* is mutated (in Panc05.0 and ESS-1 cells) or wildtype (in PANC-1, Capan-2, Capan-1, MFE-296, HEC1A, and HEC59 cells) [164].

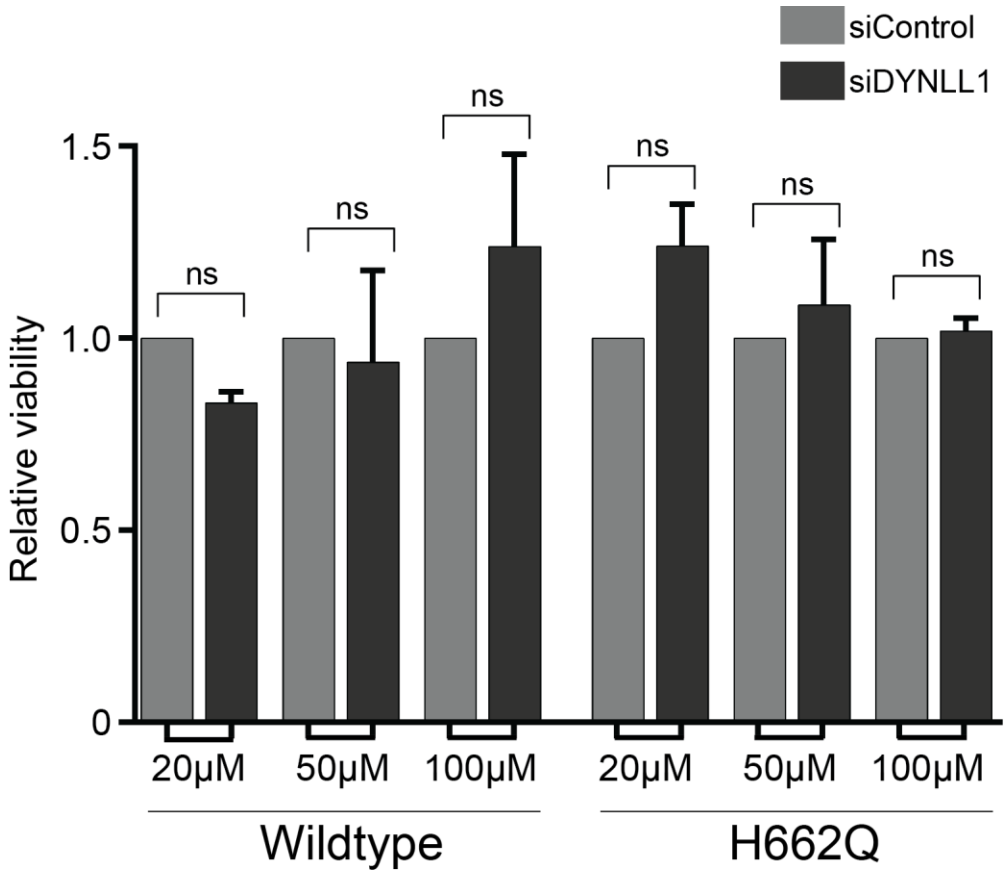

Figure 3. 18. DYNLL1 knock-down does not affect cell viability.
Cell viability 48 hours post-siRNA treatment with 20, 50, and 100 μM of si-Control or si-DYNLL1. ANOVA was used to determine statistical significance. Mean values with S.E.M. error bars are shown, *n = 3*. Viability count was performed as described in Section 3.2.10.

3.3.11 DYNLL1 overexpression may perturb regulation of repair choice

To briefly summarize the data thus far, I observed that mutations in H662Q lead to DNA damage in S phase that is partially caused by aberrant R-loops, R-loop mediated replication stress (i.e. activation of the ATR-CHK1 signaling pathway and phosphorylated RPA32-s33 accumulation), in which both DNA damage and replication stress are suppressed by replication inhibition and DYNLL1 siRNA depletion. Clearly active replication plays a key role in genetic instability in cells with SF3B1 H662Q, and the data suggests a potential cause may be transcription replication conflicts due to aberrant R-loop accumulation. The observation that DYNLL1 downregulation results in suppression of DNA damage seems to suggest that aberrant splicing of DYNLL1 in conjunction with R-loop accumulation caused by H662Q mutation might both be contributors to the overall genetic instability observed in these cells. While a model exists for how R-loop accumulation leads to genetic instability, how might excess DYNLL1 cause DNA damage in S phase cells? Recent studies suggest that DYNLL1 plays a role in the cellular response to DNA double strand breaks [236]. I wondered if aberrant splicing and subsequent upregulation of DYNLL1 in H662Q mutant cells could have defects in dealing with stalled or collapsed replication forks caused by aberrant R-loops. Before moving forward, I will briefly introduce work by groups that have discovered a role of DYNLL1 in the context of cellular response to double strand breaks.

As described in Chapter 1, regulation of homologous recombination (HR) or nonhomologous end joining (NHEJ) proteins, as well as DNA end resection is important for double strand break repair choice, and DYNLL1 has been shown to have at least two roles in regulating the balance between NHEJ and HR. First, it was found that DYNLL1 specifically binds to 53BP1 [237], and this interaction plays an essential role in promoting NHEJ, because depletion of DYNLL1 reduced the efficiency of 53BP1-mediated NHEJ [236]. Second, DYNLL1 directly interacts with MRE11 to limit DNA end resection, leading to reduced HR pathway choice [238]. Therefore I wondered if upregulation of DYNLL1 could increase efficiency of NHEJ and reduce HR repair choice in H662Q mutant cells.

I performed immunofluorescence experiments co-staining EdU treated cells with 53BP1 antibody as a marker for NHEJ and RAD51 antibody as a marker for HR to quantify 53BP1 and

RAD51 signal at G1 and S/G2 phases of the cell cycle. The results show H662Q mutant cells had a higher number of 53BP1 foci at G1 and S/G2 phases (**Figure 3.19A**), and a lower number of RAD51 foci at S/G2 phase relative to wildtype (**Figure 3.19B**), suggesting that NHEJ is upregulated while HR is downregulated. Since HR is generally active in S and G2 phases of the cell cycle when the sister chromatid is available as a repair template, the immunofluorescence data showing that RAD51 foci are predominantly observed in wildtype S/G2 phases is consistent with RAD51 as a good marker for HR, and suggests that H662Q cells have a defect in normal repair pathway choice (**Figure 3.19B**). Groups using ionizing radiation to study cell cycle dependent interactions between 53BP1 and BRCA1 showed that upon damage, 53BP1 foci increase in G1 phase, and progressively decline during early and late S phase, while BRCA1 levels gradually rise after S phase due to availability of the replicated chromatin for HR [104, 239]. My immunofluorescence experiments show that in the H662Q mutant cells, high 53BP1 levels persist in S/G2 phase, while RAD51 levels decline, potentially pointing to a skew towards NHEJ. Thus, consistent with the known role of DYNLL1 in regulating 53BP1-mediated NHEJ, it appears upregulation of DYNLL1 caused by aberrant splicing in H662Q mutant cells perturb the balance between HR and NHEJ-mediated response to replication stress and subsequent DNA damage.

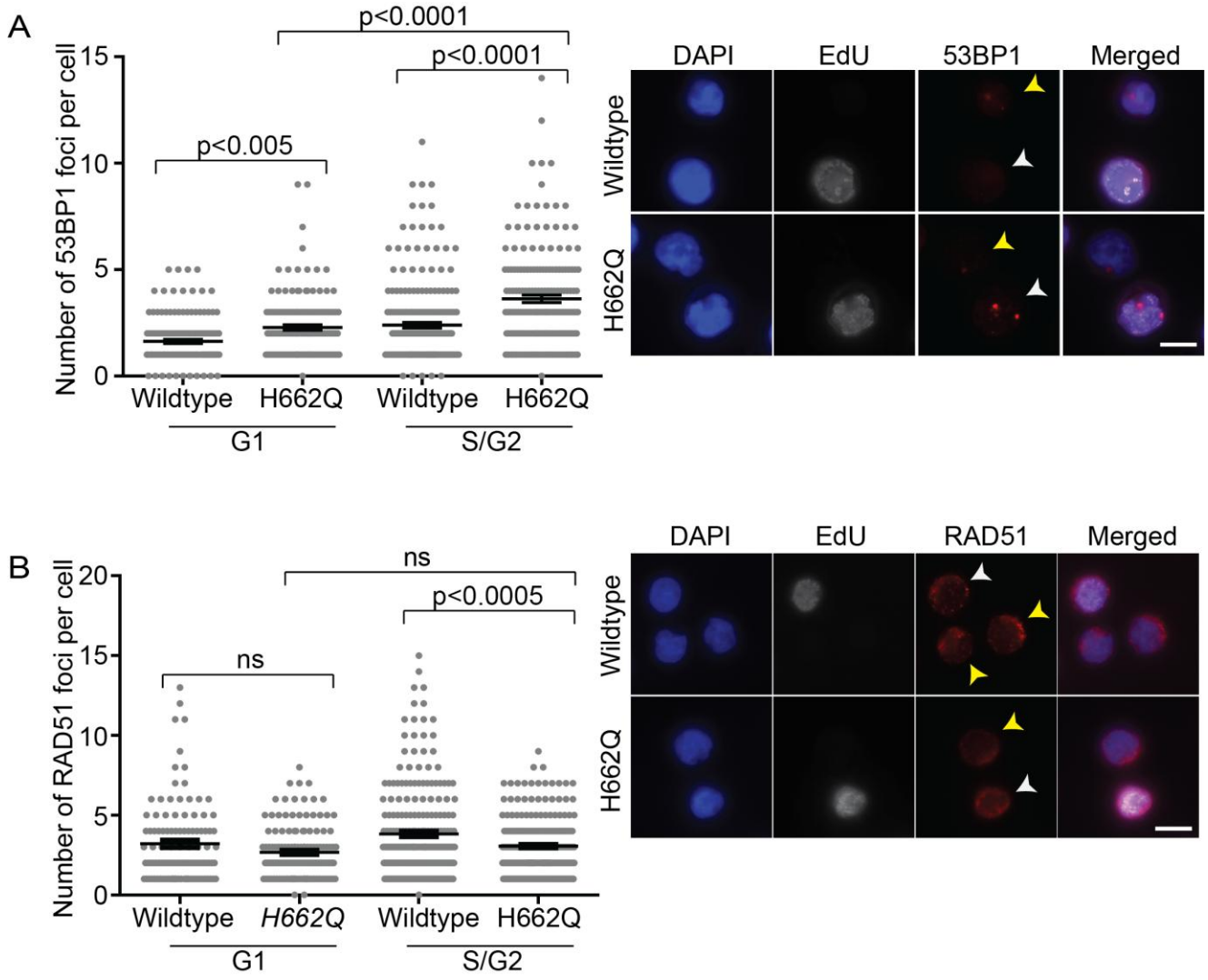

Figure 3. 19. DNA double strand break repair foci alterations in H662Q cells.
Immunofluorescence staining for (A) 53BP1 signal and (B) RAD51 signal after EdU incorporation. ANOVA was used to determine statistical significance. Mean values with S.E.M. error bars are shown, *n = 3*. >100 nuclei were scored per replicate. Representative images are shown on the right. Scale bar = 5 μm. Yellow arrows indicate EdU negative (G1 phase) cells; White arrows indicate EdU positive (S/G2 phase) cells.

3.3.12 H662Q mutant cells exhibit sensitivity to olaparib treatment

When DNA replication stalls, checkpoint proteins stabilize stalled forks, and fork restart is typically facilitated by HR proteins [240]. Indeed HR has been identified as a key pathway that allows cells to tolerate replication stress [241, 242]. While fork collapse can lead to double

strand breaks, stalled forks that are reversed resemble one ended double strand breaks. My data shows a correlation between DYNLL1 upregulation, increased 53BP1 foci, and decreased RAD51 foci, potentially pointing to a skew towards NHEJ and away from HR, which is consistent with the known activity of DYNLL1 [236]. In addition, siRNA knockdown of DYNLL1 completely rescued the genetic instability phenotype in H662Q mutant cells, a phenotype that was predominantly S phase specific, suggesting a major cause of replication stress induced DNA damage was in fact caused by high DYNLL1 protein levels. To test if HR pathway choice and the subsequent response to replication stress are affected in H662Q mutant cells, I treated the cells with two drugs that either impairs HR pathway choice, or selectively targets HR deficient cells. I used mirin (Z-5-(4-hydroxybenzylidene)-2-imino-1,3-thiazolidin-4-one), an MRE11 inhibitor to prevent HR commitment [243], and olaparib, an inhibitor of PARP (poly(ADP-ribose) polymerase)1 and 2, proteins that are involved in the repair of single strand breaks by the base excision repair (BER) pathway [6]. Inhibition of the BER pathway by olaparib results in accumulation of unrepaired single strand breaks and the trapping of PARP proteins at these sites, leading to the formation of double strand breaks that require HR-mediated repair [6]. Therefore HR deficient cancers are sensitized to PARP inhibition resulting in the accumulation of deleterious double strand breaks that are cell lethal.

The results indicate that upon increasing doses of mirin treatment, viability of both wildtype and H662Q mutant cells decrease (**Figure 3.20A**). Upon olaparib treatment, the reduction in viability in H662Q mutant cells was significantly greater than the viability reduction of wildtype cells starting at olaparib concentration of 5 μM, indicating H662Q mutant cells are sensitized to olaparib treatment (**Figure 3.20B**). It should be noted that at the highest concentration of olaparib used (10 μM), H662Q mutant cells essentially stopped dividing and I noticed the cell count number per volume used to assess viability was lower than the starting number plated at the start of the experiment, suggesting cells could have undergone necrosis. The mirin treatment data showed inhibiting MRE11 did not appreciably affect viability in H662Q mutant cells, possibly because mirin inhibits the exonuclease activity of MRE11, and it has been suggested that MRE11 endonuclease activity initiates resection and HR commitment [244]. The olaparib treatment data showed inhibiting PARP led to a reduction in viability in H662Q mutant cells when compared to wildtype, perhaps due to a defect in repair pathway choice that promotes NHEJ due to high DYNLL1 protein levels in H662Q mutant cells.

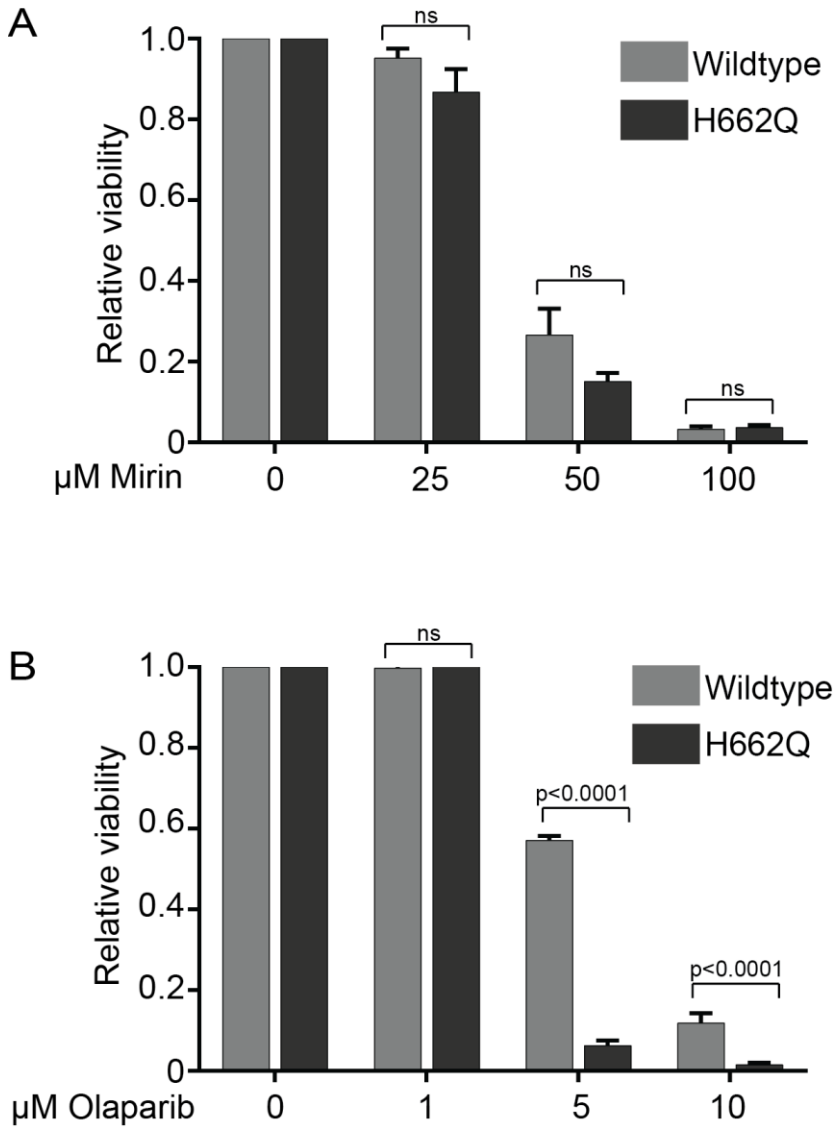

Figure 3. 20. SF3B1 mutant cells are sensitive to olaparib treatment.

Cell viability of NALM6 wildtype and NALM6 SF3B1-H662Q cells 120 hours (48+72 hours) after (A) mirin or (B) olaparib treatment at the indicates doses. 0 μM treatment indicates cells were treated with vehicle only (0.1% DMSO). For the viability counts, a mixture of acridine orange and DAPI was used to counterstain living and dead cells, and to stain dead cells, respectively. ANOVA was used to determine statistical significance. Mean values with S.E.M. error bars are shown, $n = 3$.

3.4 Conclusion

A growing body of evidence has shown splicing factor mutations play a role in genome maintenance across species [53, 63, 85, 165, 166]. The observation that splicing factors, in particular *SF3B1*, are frequently mutated in a broad range of cancers suggests spliceosome dysfunction is a driver of disease [29, 153-159]. While transcriptome analyses have characterized aberrant splicing events induced by SF3B1 mutations [78, 79, 120, 159-164], the biological consequences of these missplicing events is not well understood. To date the only clear examples of how aberrant splicing caused by *SF3B1* mutations could contribute to cancer progression includes: global splicing deregulation of multiple cellular processes including DNA damage response, telomere maintenance, and cell signaling in CLL patients [79], specific changes in expression of genome maintenance factors [78, 165], deregulation of genes involved in metabolism such as those involved in mitochondrial respiration and serine synthesis [245], and aberrant splicing of the iron transporter *ABCB7*, which has been linked to a specific type of MDS called MDS-RARS (myelodysplastic syndrome with refractory anemia with ring sideroblasts), that is characterized by erythroid precursors with abnormal mitochondrial iron accumulation (ring sideroblasts) [171].

Splicing occurs co-transcriptionally, therefore defects in splicing activity could perturb the transcriptional process and lead to accumulation of R-loop structures that contribute to genome instability by exposing single-stranded DNA (ssDNA) and by blocking replication forks, causing replication stress induced genome instability [119, 120]. While many studies across species have provided evidence that splicing defects induce R-loop mediated DNA damage [12, 51, 55, 63, 64, 119, 167], the role of *SF3B1* hotspot mutations in R-loop accumulation and genetic instability in cancer has not yet been reported.

Overall, my data reveal how a frequent *SF3B1* hotspot mutation contributes to genetic instability in two ways: 1) by aberrant accumulation of R-loops that leads to replication stress, and 2) aberrant splicing of the multifunctional protein DYNLL1, in which the resulting protein upregulation may enhance NHEJ and/or inhibit HR, and potentially affect the response to replication stress induced DNA damage. Importantly, this dysregulation of repair in the DYNLL1-high H662Q cells is associated with a significant sensitization to the PARP inhibitor

olaparib. This sensitization is consistent with a problem in HR caused by high DYNLL1 and the potential for clinical impact of these observations is an exciting area of future study for the lab. More broadly, while I have previously shown that a conditional mutant of *HSH155*, the yeast homolog of *SF3B1*, causes genetic instability by missplicing of a key transcript important for mitotic spindle assembly (Chapter 2), the cancer-associated *SF3B1* mutations in the equivalent Hsh155 protein residues do not appear to induce the same genetic instability phenotype in yeast. This shows that while genetic instability appears to be a general consequence of splicing defects, the biological consequences of splicing factor mutations is dependent on the penetrance of the allele and the activity of the splicing factor in shifting splicing events. Taken together, my results show such mutants with broad impacts on gene expression can selectively impair a specific aspect of genome maintenance that can be identified, such as α-tubulin in yeast or DYNLL1 in human cells.

Chapter 4: Overall discussion and conclusions

4.1 Summary of data

The goal of my thesis work was to understand how splicing factor mutations in yeast and human cells contribute to genetic instability. To this end, I have identified how select splicing factors in yeast and human cells contribute to genetic instability through R-loop accumulation and fluctuations in gene expression, adding to a growing body of evidence that splicing factors play a key role in genome maintenance across species.

In Chapter 2, I selected yeast strains with conditional mutations in each of the core snRNP complexes involved in establishing the splicing reaction – *YHC1* (U1), *HSH155* (U2), and *SNU114* (U4/U6.U5), and observed evidence of R-loop induced DNA damage in a mutant allele of *SNU114*, while all splicing mutants tested caused genetic instability by missplicing of *TUB1*, the protein product of which, α-tubulin, is critical in forming the mitotic spindle.

In Chapter 3, to gain functional insights to how *HSH155* could influence genetic instability in the context of cancer progression, I extended my analysis to include five cancer-associated *SF3B1* point mutations in the yeast homolog Hsh155. While splicing activity in these two homologs was conserved, it was clear that *SF3B1* and *HSH155* cancer-associated mutations did not induce the same types of transcriptional changes, likely due to the differences in intron features between yeast and humans. In humans cells, cancer-associated *SF3B1* mutations most frequently induce cryptic 3' splice site usage due to defects in recognizing the branch site sequence during splicing [78, 160, 162, 183, 184]. In yeast cells, the equivalent *HSH155* mutations result in defective branch site selection, but do not induce cryptic 3' splice site usage due to the inherently strong consensus sequences within yeast introns [161, 180]. I used isogenic NALM6 human leukemia cell lines to investigate how a specific *SF3B1* hotspot mutation contributes to genetic instability. My data indicates that the SF3B1 H662Q hotspot mutation may cause genetic instability in at least two ways: 1) by inducing R-loop-mediated replication stress either directly or indirectly through suboptimal expression of an unidentified R-loop modulating factor, and 2) aberrant splicing of the multifunctional protein DYNLL1, resulting in protein overexpression that potentially perturbs the balance of double strand break repair pathway choice.

In both yeast and mammalian systems, I have identified the cause of genetic instability when select splicing factors are mutated (summarized in **Figure 4.1**). My results indicate that while these mutants could cause global changes to gene expression, I was able to identify a precise aspect of genome maintenance that was perturbed by missplicing of a specific gene. In addition, although I was unable to determine the exact cause of aberrant R-loop accumulation in *SNU114* mutant yeast and SF3B1 H662Q mutant human cells, my data adds to the existing understanding that R-loop accumulation caused by splicing defects appears to be influenced by the penetrance of the allele, as well as the specific splicing factor mutated, organism studied, and cell type. For speculation on why only some splicing factor mutants induce R-loops, please refer to Section 4.2.1.

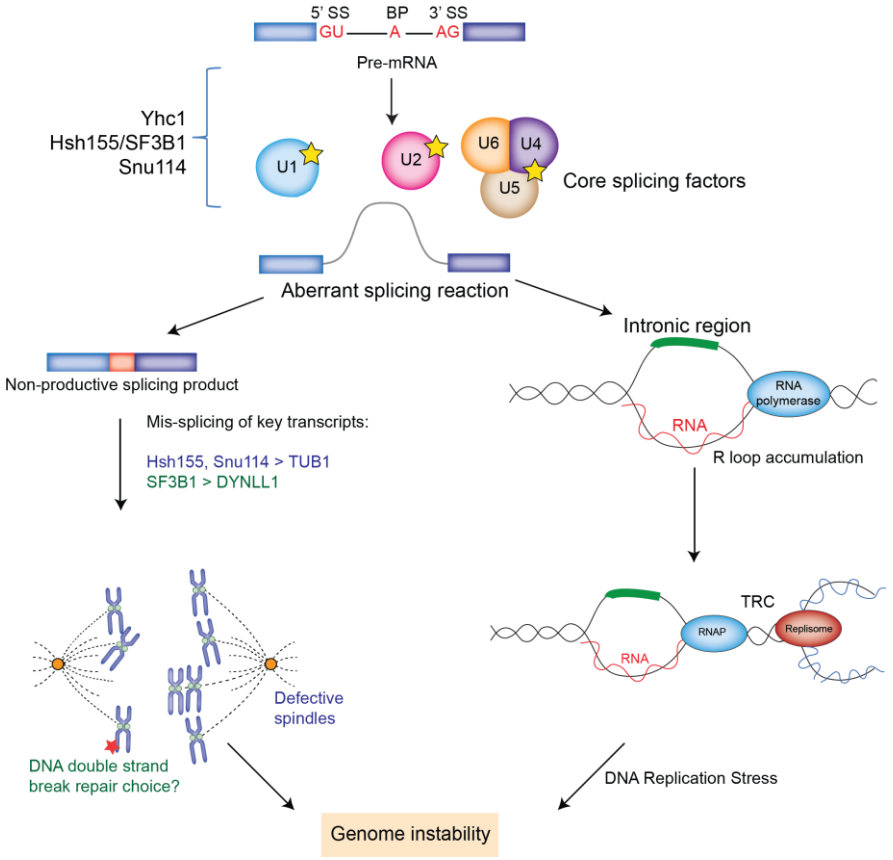

Figure 4. 1 Model of defective splicing-induced genome instability in yeast and human cells. Schematic summarizing causes of genome instability when specific splicing factors are mutated. Core splicing factor mutations can cause genome instability by changes in expression of transcripts that have roles in genome maintenance or by aberrant accumulation of R-loops that can drive transcription-replication conflicts (TRC) and cause DNA replication stress. Adapted and modified from Tam A.S. and Stirling P.C., 2019 *Curr Genet* [246].

4.2 Genetic instability is a general consequence of splicing factor mutations

Many transcriptome analysis studies have characterized aberrant splicing events induced by splicing factor mutations in cancer to understand if some commonly affected splicing events contribute to the disease [78, 79, 120, 159-164]. Mutated components of the spliceosome were mainly detected in the U2 snRNP and U2-related proteins that play key roles in branchsite and 3' splice site recognition like SF3B1, SRSF2, U2AF1, and ZRSR2, suggesting that proper recognition of these intron features is important for optimal splicing and cell function. Each of these mutations has been shown to induce a large number of abnormal splicing events in cell lines, mouse models, and human cancers [79, 171, 247]. However, it is not entirely clear how modulating splicing patterns affect tumourigenesis, as it has also been shown that these splicing factor mutations induce unique sets of splicing changes in distinct groups of transcripts depending on the cell type studied, and also in the same cell type carrying different splicing factor mutations [64, 81, 166]. Evidence suggests that numerous transcriptional changes induced by aberrant *SF3B1* mutations are subtle and broad, and it has been hypothesized that this pattern of activity affects oncogenesis by allowing cells to tolerate many changes in expression that may affect cell viability, and thus allow changes to be propagated, while changes in gene expression diversity could allow cells to select for pathways that promote cell survival [79]. Indeed, it has been proposed that splicing factor mutations cause genetic instability by changes in gene expression that synergize with additional R-loop accumulation that mounts the DNA damage response [64]. Mutations in *U2AF1* and *SRSF2* was observed to induce aberrant R-loop accumulation that increased DNA damage so that ATR but not ATM pathway was activated, suggesting a limited genetic instability phenotype in these cells [63, 64]. Additionally, these mutations appeared to affect expression of the DNA damage response factors *CLSPN* and *PMS2*, pointing to a synergistic relationship between R-loop accumulation and the subsequent response to R-loops as contributors to genetic instability [64]. Therefore, it is important to understand the context in which splicing factor mutations cause R-loop mediated damage.

4.2.1 R-loop accumulation and changes in gene expression cause genetic instability

Unbiased screens revealed a link between splicing and R-loop mediated genetic instability [11, 54, 248], but there are also many cases where splicing factor mutations induce genetic

instability that is independent of R-loop accumulation [54, 72-74, 249]. This suggests that a single splicing factor disruption can lead to genome instability through multiple mechanisms. However as previously discussed, the interaction between R-loop mediated damage and changes in gene expression that cause damage is complex and context specific.

I used yeast as a simplified model to understand the impact of splicing on genome maintenance. In yeast, about ~5% of genes contain introns, and these introns have been well annotated and characterized [176]. Intron deletions are also well tolerated in yeast, therefore a clean deletion of intron sequences can be used to assess splicing function. In Chapter 2, I characterized splicing factor mutants that affect early steps of splicing and observed that defects in distinct splicing factors can induce DNA damage through changes in gene expression, or changes in gene expression in conjunction with R-loop accumulation, suggesting multiple parallel mechanisms can synergize to induce genetic instability. This is consistent with the literature where mutations in the splicing regulator SLU7 was found to indirectly cause R-loop mediated damage as well as missplicing of transcripts important for sister chromatid cohesion [250]. Taken together with these studies, my data suggests that splicing factor mutations can cause genetic instability through missplicing of key transcripts important for genome maintenance, but a subset of splicing factors cause R-loop mediated genetic instability likely through activity that does not directly involve splicing. In Chapter 2, I measured hyperrecombination frequencies in *snu114-60* mutants using an integrated intronless genomic reporter, and upon RNase H1 overexpression, this suppressed the recombination phenotype, suggesting R-loops cause the recombination phenotype, but this was not dependent on the splicing function of Snu114 (Figure 2.5B). In addition, removal of the intron in *TUB1*, the target of missplicing leading to DNA damage, did not suppress R-loop accumulation or hyperrecombination, demonstrating two distinct mechanisms, operating concurrently, create genome instability in this strain (Figure 2.16). It has been suggested that perhaps RNA binding or spliceosome recruitment during transcription, not RNA splicing, causes R-loop accumulation in splicing factor mutations. Indeed, many RNA binding proteins have been identified as factors that prevent aberrant R-loop levels [11, 23, 194, 195, 251]. Perhaps most compellingly, R-ChIP (chromatin immunoprecipitation using a catalytically inactive RNase H1) used to map R-loops genome-wide in *U2AF1* and *SRSF2* mutant cells indicate that only a small fraction of altered R-loops (up or downregulated levels relative to wildtype control) was detected in gene bodies, and

an even smaller fraction of aberrant R-loops were associated with splice sites, suggesting R-loop formation is not necessarily caused by splicing defects in *cis* [64]. It should be noted that in yeast lacking both RNase H1 and 2, DRIP (immunoprecipitation using S9.6 antibody to pulldown hybrids) analysis revealed that intron-containing genes appeared to have significantly lower hybrid levels compared to intronless genes [59], pointing to a protective role of introns in R-loop prevention which is consistent with observations previously made by this group [252]. Additional experiments using yeast with mutated intron consensus sites affecting various levels of spliceosome recruitment showed that spliceosome recruitment but subsequent incomplete splicing reaction was sufficient to inhibit R-loop accumulation at these intron-containing genes, indicating the act of spliceosome recruitment, and not splicing, affects R-loop accumulation [59]. It is not entirely clear how splicing factors like Snu114 affect spliceosome recruitment and cause R-loop accumulation, since the three splicing factors in yeast studied in Chapter 2 have functions in different early steps of splicing and appear to have different mechanisms of genetic instability.

It has been shown that splicing factor mutations induce unique sets of splicing changes in distinct groups of transcripts that is cell type and cancer specific [64, 79, 81, 166, 171, 247]. To control for differences in cell type, I used an isogenic pair of *SF3B1* wildtype or mutant NALM6 cell lines to understand how *SF3B1* mutations cause genetic instability. My results suggest that SF3B1 H662Q mutant cells cause genetic instability by changes in gene expression of a specific transcript that synergizes with R-loop mediated replication stress, which has been proposed to be a unifying mechanism for tumourigenesis induced by cancer-associated splicing factor mutations [64]. It is interesting to note that while prior work did not include *SF3B1* in their analysis, my results are highly consistent with findings using *U2AF1* and *SRSF2* mutant cell lines that found these splicing factor mutations induced R-loops, replication stress, and activation of the ATR-CHK1 pathway. It was shown that *U2AF1* and *SRSF2* mutations do not activate the ATM pathway [64], and while ATM appears to be phosphorylated when SF3B1 H662Q is mutated in NALM6 cells, CHK2 signaling was not activated (Figure 3.9). The authors suggest that fluctuations in expression of DNA damage response proteins CLSPN and PMS2 in conjunction with R-loop accumulation cause a minor genetic instability phenotype that did not reach a level to cause DNA double strand breaks, although I have not tested whether these factors are misspliced in the SF3B1 H662Q mutant NALM6 cells [64]. My data fits with the hypothesis that transcriptional changes induced by mutated components of the U2 snRNP and U2-related

proteins affects oncogenesis by allowing cells to tolerate factors that may affect cell viability and genome stability, such as aberrant R-loop accumulation, thus imparting selective advantages that promote cell adaptation.

4.2.2 Interplay between DNA damage induction and subsequent response drives genetic instability

My results indicate that one of the ways SF3B1 H662Q mutant cells cause genetic instability is by aberrant splicing of DYNLL1, in which the resulting protein upregulation may be promoting NHEJ. As discussed in Chapter 3, DYNLL1 has been shown to specifically bind to 53BP1 and modulate the efficiency of 53BP1-mediated NHEJ [236, 237]. It is tempting to speculate that DYNLL1 protein upregulation and the resulting skew towards NHEJ is the main cause of genetic instability when SF3B1 H662Q is mutated, however as previously reported, downregulation of DNA damage response proteins in splicing mutant cells can also affect genetic instability phenotypes [64]. Additionally, DYNLL1 has many regulatory roles in processes like protein dimerization of targets that include transcription factors, DNA damage response proteins, regulators of apoptosis, amongst others [224-227, 231]. Nevertheless, the data indicates DYNLL1 protein upregulation caused by missplicing in SF3B1 H662Q mutant cells is correlated with increase in 53BP1 foci and reduction of RAD51 foci, suggesting a skew towards NHEJ over homologous recombination (HR) (Figure 3.19). Additionally, depletion of DYNLL1 and inhibition of replication completely rescued the genetic instability phenotype, suggesting a major cause of S phase dependent DNA damage was high DYNLL1 levels. Lastly, upon treatment with olaparib, H662Q mutant cells were selectively sensitive to PARP inhibition, mimicking, albeit subtly, the effect seen in HR-deficient cancers (Figure 3.20). Thus my model of how SF3B1 H662Q mutation leads to genetic instability is by R-loop mediated replication stress, and the promotion of NHEJ by high DYNLL1 protein levels, ultimately leading to a more mutagenic DNA damage response. It is interesting to note that to date, only a handful of papers have identified a role of DYNLL1 in DNA repair. In addition, to my knowledge this is the first time that the function of DYNLL1 has been implicated in replication stress tolerance. Therefore my thesis work has expanded our understanding of how this multifunctional protein maintains genome stability.

4.3 Conclusion

The goal of my thesis was to understand how splicing factor mutations in yeast and human cells contribute to genetic instability. The results in my thesis show splicing factor mutations in yeast and mammalian cells cause genetic instability by fluctuations in gene expression and aberrant R-loop accumulation. Additionally in both systems, I have identified specific transcripts that when misspliced lead to DNA damage phenotypes. My data enhances our understanding of how splicing factor mutations contribute to genetic instability through multiple mechanisms that operate concurrently. My results also contribute to the idea that transcriptional changes induced by mutated components of the spliceosome can manifest in highly specific yet one-step-removed phenotypes such as genetic instability. This is part of a broader effort to map seemingly peripheral cellular pathways that have been implicated in genome stability [248], back to fundamental DNA transactions of replication, repair or mitosis.

4.4 Future Directions

SUGP1 and the mechanistic SF3B1-R-loop connection: While my results show R-loop accumulation is a cause of genetic instability in splicing factor mutations, I was not able to identify the source of aberrant R-loops. It appears that spliceosome recruitment is important to prevent R-loop accumulation. The cancer-associated mutations in *SF3B1* cluster in the HEAT domain and are heterozygous missense mutations, suggesting these are gain of function mutations that alter the function of this domain [172]. It has been suggested that these point mutations may affect interactions with other splicing factors and proteins, because in yeast it was shown that the Hsh155 HEAT domain mutations affect interactions with the splicing factor Prp5 [180]. Due to the dynamic nature of protein-protein/RNA and RNA-RNA interactions, it has been difficult to identify loss of interactions caused by *SF3B1* mutations. Recently it was found that *SF3B1* HEAT domain mutations resulted in weakened interaction with splicing factor SUGP1, and the weakening of this interaction was responsible for the defects in branch site selection and cryptic 3' splice site when *SF3B1* is mutated [253]. While this direct protein-protein interaction is important for SF3B1 function in branch site selection, it is not entirely clear if loss of SUGP1 recruitment contributes to R-loop accumulation. Moving forward, to determine the cause of R-loop accumulation in *SF3B1* mutant cells, it will be important to identify factors

that directly interact with or are recruited by SF3B1, and how loss of these interactions, including loss of SUGP1, could affect R-loop accumulation.

DYNLL1 functions in S phase: As noted DYNLL1 has numerous cellular functions including those directly linked to DNA repair through interactions with 53BP1 and MRE11 [236, 238]. Nevertheless, to my knowledge, the results in Chapter 3 are among the first to show a potential function for DYNLL1 in S phase. Additional understanding of the mechanism of action for DYNLL1 in replication could have major impacts and I would like to explore this further. For example, DYNLL1-high tumours may be more sensitive to PARP inhibitors which is of clinical importance. In addition, tools to subtly shift DNA repair pathway choice, even while cells are cycling could have implications for CRISPR genome editing. Thus, exploring a novel role for DYNLL1 in modulating replication is an outcome and future direction of my work.

Bibliography

1. Herr, A.J., et al., *DNA replication error-induced extinction of diploid yeast.* Genetics, 2014. **196**(3): p. 677-91.
2. Loeb, L.A., *Human cancers express mutator phenotypes: origin, consequences and targeting.* Nat Rev Cancer, 2011. **11**(6): p. 450-7.
3. Stratton, M.R., P.J. Campbell, and P.A. Futreal, *The cancer genome.* Nature, 2009. **458**(7239): p. 719-24.
4. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
5. Lord, C.J. and A. Ashworth, *Mechanisms of resistance to therapies targeting BRCA-mutant cancers.* Nat Med, 2013. **19**(11): p. 1381-8.
6. Morales, J., et al., *Review of poly (ADP-ribose) polymerase (PARP) mechanisms of action and rationale for targeting in cancer and other diseases.* Crit Rev Eukaryot Gene Expr, 2014. **24**(1): p. 15-28.
7. Germano, G., et al., *Inactivation of DNA repair triggers neoantigen generation and impairs tumour growth.* Nature, 2017. **552**(7683): p. 116-120.
8. Le, D.T., et al., *PD-1 Blockade in Tumors with Mismatch-Repair Deficiency.* N Engl J Med, 2015. **372**(26): p. 2509-20.
9. Rizvi, N.A., et al., *Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer.* Science, 2015. **348**(6230): p. 124-8.
10. 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.
11. 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.
12. Stirling, P.C., et al., *R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants.* Genes Dev, 2012. **26**(2): p. 163-75.
13. Felipe-Abrio, I., et al., *RNA polymerase II contributes to preventing transcription-mediated replication fork stalls.* EMBO J, 2015. **34**(2): p. 236-50.

14. Dahan, O. and M. Kupiec, *Mutations in genes of Saccharomyces cerevisiae encoding pre-mRNA splicing factors cause cell cycle arrest through activation of the spindle checkpoint.* Nucleic Acids Res, 2002. **30**(20): p. 4361-70.
15. Li, X. and J.L. Manley, *Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability.* Cell, 2005. **122**(3): p. 365-78.
16. Mischo, H.E., et al., *Yeast Sen1 helicase protects the genome from transcription-associated instability.* Mol Cell, 2011. **41**(1): p. 21-32.
17. Jimeno, S., et al., *The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability.* EMBO J, 2002. **21**(13): p. 3526-35.
18. Gavalda, S., et al., *R-loop mediated transcription-associated recombination in trf4Delta mutants reveals new links between RNA surveillance and genome integrity.* PLoS One, 2013. **8**(6): p. e65541.
19. Luna, R., et al., *Interdependence between transcription and mRNP processing and export, and its impact on genetic stability.* Mol Cell, 2005. **18**(6): p. 711-22.
20. Mikolaskova, B., et al., *Maintenance of genome stability: the unifying role of interconnections between the DNA damage response and RNA-processing pathways.* Curr Genet, 2018. **64**(5): p. 971-983.
21. Bernier, F.P., et al., *Haploinsufficiency of SF3B4, a component of the pre-mRNA spliceosomal complex, causes Nager syndrome.* Am J Hum Genet, 2012. **90**(5): p. 925-33.
22. Lines, M.A., et al., *Haploinsufficiency of a spliceosomal GTPase encoded by EFTUD2 causes mandibulofacial dysostosis with microcephaly.* Am J Hum Genet, 2012. **90**(2): p. 369-77.
23. Buskin, A., et al., *Disrupted alternative splicing for genes implicated in splicing and ciliogenesis causes PRPF31 retinitis pigmentosa.* Nat Commun, 2018. **9**(1): p. 4234.
24. Agrawal, A.A., et al., *Targeting splicing abnormalities in cancer.* Curr Opin Genet Dev, 2018. **48**: p. 67-74.
25. Seiler, M., et al., *Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types.* Cell Rep, 2018. **23**(1): p. 282-296 e4.
26. Edery, P., et al., *Association of TALS developmental disorder with defect in minor splicing component U4atac snRNA.* Science, 2011. **332**(6026): p. 240-3.

27. He, H., et al., *Mutations in U4atac snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I.* Science, 2011. **332**(6026): p. 238-40.
28. Lynch, D.C., et al., *Disrupted auto-regulation of the spliceosomal gene SNRPB causes cerebro-costo-mandibular syndrome.* Nat Commun, 2014. **5**: p. 4483.
29. Coltri, P.P., M.G.P. Dos Santos, and G.H.G. da Silva, *Splicing and cancer: Challenges and opportunities.* Wiley Interdiscip Rev RNA, 2019. **10**(3): p. e1527.
30. Furukawa, T., et al., *Whole-exome sequencing uncovers frequent GNAS mutations in intraductal papillary mucinous neoplasms of the pancreas.* Sci Rep, 2011. **1**: p. 161.
31. Kurtovic-Kozaric, A., et al., *PRPF8 defects cause missplicing in myeloid malignancies.* Leukemia, 2015. **29**(1): p. 126-36.
32. Quigley, D.A., et al., *Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer.* Cell, 2018. **175**(3): p. 889.
33. Tien, J.F., et al., *CDK12 regulates alternative last exon mRNA splicing and promotes breast cancer cell invasion.* Nucleic Acids Res, 2017. **45**(11): p. 6698-6716.
34. Hamard, P.J., et al., *PRMT5 Regulates DNA Repair by Controlling the Alternative Splicing of Histone-Modifying Enzymes.* Cell Rep, 2018. **24**(10): p. 2643-2657.
35. Li, Y., et al., *PRMT5 is required for lymphomagenesis triggered by multiple oncogenic drivers.* Cancer Discov, 2015. **5**(3): p. 288-303.
36. Hastings, M.L. and A.R. Krainer, *Pre-mRNA splicing in the new millennium.* Curr Opin Cell Biol, 2001. **13**(3): p. 302-9.
37. Anna, A. and G. Monika, *Splicing mutations in human genetic disorders: examples, detection, and confirmation.* J Appl Genet, 2018. **59**(3): p. 253-268.
38. Matlin, A.J., F. Clark, and C.W. Smith, *Understanding alternative splicing: towards a cellular code.* Nat Rev Mol Cell Biol, 2005. **6**(5): p. 386-98.
39. Wang, Y., et al., *Mechanism of alternative splicing and its regulation.* Biomed Rep, 2015. **3**(2): p. 152-158.
40. Scotti, M.M. and M.S. Swanson, *RNA mis-splicing in disease.* Nat Rev Genet, 2016. **17**(1): p. 19-32.
41. Ares, M., Jr., L. Grate, and M.H. Pauling, *A handful of intron-containing genes produces the lion's share of yeast mRNA.* RNA, 1999. **5**(9): p. 1138-9.
42. Will, C.L. and R. Luhrmann, *Spliceosome structure and function.* Cold Spring Harb Perspect Biol, 2011. **3**(7).

43. Nguyen, H., et al., *The matrices and constraints of GT/AG splice sites of more than 1000 species/lineages.* Gene, 2018. **660**: p. 92-101.
44. Schwartz, S.H., et al., *Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes.* Genome Res, 2008. **18**(1): p. 88-103.
45. Wahl, M.C., C.L. Will, and R. Luhrmann, *The spliceosome: design principles of a dynamic RNP machine.* Cell, 2009. **136**(4): p. 701-18.
46. Fica, S.M. and K. Nagai, *Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine.* Nat Struct Mol Biol, 2017. **24**(10): p. 791-799.
47. Sollier, J., et al., *Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability.* Mol Cell, 2014. **56**(6): p. 777-85.
48. Ruiz, J.F., B. Gomez-Gonzalez, and A. Aguilera, *AID induces double-strand breaks at immunoglobulin switch regions and c-MYC causing chromosomal translocations in yeast THO mutants.* PLoS genetics, 2011. **7**(2): p. e1002009.
49. Pefanis, E., et al., *Noncoding RNA transcription targets AID to divergently transcribed loci in B cells.* Nature, 2014. **514**(7522): p. 389-93.
50. Helmrich, A., et al., *Transcription-replication encounters, consequences and genomic instability.* Nat Struct Mol Biol, 2013. **20**(4): p. 412-8.
51. Li, X. and J.L. Manley, *Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability.* Cell, 2005. **122**(3): p. 365-378.
52. Paulsen, R.D., et al., *A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability.* Molecular cell, 2009. **35**(2): p. 228-239.
53. Stirling, P.C., et al., *The Complete Spectrum of Yeast Chromosome Instability Genes Identifies Candidate CIN Cancer Genes and Functional Roles for ASTRA Complex Components.* PLoS genetics, 2011. **7**(4): p. e1002057.
54. Chan, Y.A., et al., *Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip.* PLoS Genet, 2014. **10**(4): p. e1004288.
55. Wan, Y., et al., *Splicing function of mitotic regulators links R-loop-mediated DNA damage to tumor cell killing.* J Cell Biol, 2015. **209**(2): p. 235-46.
56. Tuduri, S., et al., *Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription.* Nature cell biology, 2009. **11**(11): p. 1315-1324.

57. Jangi, M., et al., *SMN deficiency in severe models of spinal muscular atrophy causes widespread intron retention and DNA damage.* Proc Natl Acad Sci U S A, 2017. **114**(12): p. E2347-E2356.
58. Jimenez, M., et al., *Splicing events in the control of genome integrity: role of SLU7 and truncated SRSF3 proteins.* Nucleic Acids Res, 2019.
59. Bonnet, A., et al., *Introns Protect Eukaryotic Genomes from Transcription-Associated Genetic Instability.* Mol Cell, 2017. **67**(4): p. 608-621 e6.
60. Huertas, P. and A. Aguilera, *Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination.* Molecular cell, 2003. **12**(3): p. 711-721.
61. Gaillard, H. and A. Aguilera, *Transcription as a Threat to Genome Integrity.* Annu Rev Biochem, 2016. **85**: p. 291-317.
62. Komeno, Y., et al., *SRSF2 Is Essential for Hematopoiesis, and Its Myelodysplastic Syndrome-Related Mutations Dysregulate Alternative Pre-mRNA Splicing.* Mol Cell Biol, 2015. **35**(17): p. 3071-82.
63. Nguyen, H.D., et al., *Spliceosome Mutations Induce R Loop-Associated Sensitivity to ATR Inhibition in Myelodysplastic Syndromes.* Cancer Res, 2018. **78**(18): p. 5363-5374.
64. Chen, L., et al., *The Augmented R-Loop Is a Unifying Mechanism for Myelodysplastic Syndromes Induced by High-Risk Splicing Factor Mutations.* Mol Cell, 2018. **69**(3): p. 412-425 e6.
65. Pleiss, J.A., et al., *Transcript specificity in yeast pre-mRNA splicing revealed by mutations in core spliceosomal components.* PLoS Biol, 2007. **5**(4): p. e90.
66. Buratti, E., et al., *Aberrant 5' splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization.* Nucleic Acids Res, 2007. **35**(13): p. 4250-63.
67. Jumaa, H. and P.J. Nielsen, *The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation.* EMBO J, 1997. **16**(16): p. 5077-85.
68. Sureau, A., et al., *SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs.* EMBO J, 2001. **20**(7): p. 1785-96.
69. Burns, C.G., et al., *Removal of a single alpha-tubulin gene intron suppresses cell cycle arrest phenotypes of splicing factor mutations in Saccharomyces cerevisiae.* Molecular and cellular biology, 2002. **22**(3): p. 801-815.

70. Chawla, G., et al., *Dependence of pre-mRNA introns on PRP17, a non-essential splicing factor: implications for efficient progression through cell cycle transitions.* Nucleic Acids Res, 2003. **31**(9): p. 2333-43.
71. Dahan, O. and M. Kupiec, *The Saccharomyces cerevisiae gene CDC40/PRP17 controls cell cycle progression through splicing of the ANC1 gene.* Nucleic Acids Res, 2004. **32**(8): p. 2529-40.
72. Sundaramoorthy, S., et al., *Functional genomics identifies a requirement of pre-mRNA splicing factors for sister chromatid cohesion.* EMBO J, 2014. **33**(22): p. 2623-42.
73. Oka, Y., et al., *UBL5 is essential for pre-mRNA splicing and sister chromatid cohesion in human cells.* EMBO Rep, 2014. **15**(9): p. 956-64.
74. van der Lelij, P., et al., *SNW1 enables sister chromatid cohesion by mediating the splicing of sororin and APC2 pre-mRNAs.* EMBO J, 2014. **33**(22): p. 2643-58.
75. Watrin, E., et al., *Sororin pre-mRNA splicing is required for proper sister chromatid cohesion in human cells.* EMBO Rep, 2014. **15**(9): p. 948-55.
76. Tam, A.S., et al., *Selective defects in gene expression control genome instability in yeast splicing mutants.* Mol Biol Cell, 2019. **30**(2): p. 191-200.
77. Darman, R.B., et al., *Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3? Splice Site Selection through Use of a Different Branch Point.* Cell Rep, 2015. **13**(5): p. 1033-45.
78. Dolatshad, H., et al., *Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells.* Leukemia, 2015. **29**(8): p. 1798.
79. Wang, L., et al., *Transcriptomic Characterization of SF3B1 Mutation Reveals Its Pleiotropic Effects in Chronic Lymphocytic Leukemia.* Cancer Cell, 2016. **30**(5): p. 750-763.
80. Maciejowski, J. and T. de Lange, *Telomeres in cancer: tumour suppression and genome instability.* Nat Rev Mol Cell Biol, 2017. **18**(3): p. 175-186.
81. Ilagan, J.O., et al., *U2AF1 mutations alter splice site recognition in hematological malignancies.* Genome Res, 2015. **25**(1): p. 14-26.
82. Tien, J.F., et al., *CDK12 regulates alternative last exon mRNA splicing and promotes breast cancer cell invasion.* Nucleic Acids Res, 2017.

83. Blazek, D., et al., *The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes.* Genes Dev, 2011. **25**(20): p. 2158-72.
84. Quigley, D.A., et al., *Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer.* Cell, 2018. **174**(3): p. 758-769 e9.
85. Savage, K.I., et al., *Identification of a BRCA1-mRNA splicing complex required for efficient DNA repair and maintenance of genomic stability.* Mol Cell, 2014. **54**(3): p. 445-59.
86. Vohhodina, J., et al., *The RNA processing factors THRAP3 and BCLAF1 promote the DNA damage response through selective mRNA splicing and nuclear export.* Nucleic Acids Res, 2017. **45**(22): p. 12816-12833.
87. Marechal, A., et al., *PRP19 transforms into a sensor of RPA-ssDNA after DNA damage and drives ATR activation via a ubiquitin-mediated circuitry.* Mol Cell, 2014. **53**(2): p. 235-246.
88. Zhang, N., et al., *Cdc5L interacts with ATR and is required for the S-phase cell-cycle checkpoint.* EMBO Rep, 2009. **10**(9): p. 1029-35.
89. White, R.R. and J. Vijg, *Do DNA Double-Strand Breaks Drive Aging?* Mol Cell, 2016. **63**(5): p. 729-38.
90. Bouwman, B.A.M. and N. Crosetto, *Endogenous DNA Double-Strand Breaks during DNA Transactions: Emerging Insights and Methods for Genome-Wide Profiling.* Genes (Basel), 2018. **9**(12).
91. Jackson, S.P. and J. Bartek, *The DNA-damage response in human biology and disease.* Nature, 2009. **461**(7267): p. 1071-8.
92. Ceccaldi, R., B. Rondinelli, and A.D. D'Andrea, *Repair Pathway Choices and Consequences at the Double-Strand Break.* Trends Cell Biol, 2016. **26**(1): p. 52-64.
93. Li, X. and W.D. Heyer, *Homologous recombination in DNA repair and DNA damage tolerance.* Cell Res, 2008. **18**(1): p. 99-113.
94. Arnaudeau, C., C. Lundin, and T. Helleday, *DNA double-strand breaks associated with replication forks are predominantly repaired by homologous recombination involving an exchange mechanism in mammalian cells.* J Mol Biol, 2001. **307**(5): p. 1235-45.

95. Saleh-Gohari, N., et al., *Spontaneous homologous recombination is induced by collapsed replication forks that are caused by endogenous DNA single-strand breaks.* Mol Cell Biol, 2005. **25**(16): p. 7158-69.
96. Bryant, H.E., et al., *PARP is activated at stalled forks to mediate Mre11-dependent replication restart and recombination.* EMBO J, 2009. **28**(17): p. 2601-15.
97. Saintigny, Y., et al., *Characterization of homologous recombination induced by replication inhibition in mammalian cells.* EMBO J, 2001. **20**(14): p. 3861-70.
98. Lundin, C., et al., *Different roles for nonhomologous end joining and homologous recombination following replication arrest in mammalian cells.* Mol Cell Biol, 2002. **22**(16): p. 5869-78.
99. Hastings, P.J., et al., *Mechanisms of change in gene copy number.* Nat Rev Genet, 2009. **10**(8): p. 551-64.
100. Paull, T.T., *20 Years of Mre11 Biology: No End in Sight.* Mol Cell, 2018. **71**(3): p. 419-427.
101. Panier, S. and S.J. Boulton, *Double-strand break repair: 53BP1 comes into focus.* Nat Rev Mol Cell Biol, 2014. **15**(1): p. 7-18.
102. Bouwman, P., et al., *53BP1 loss rescues BRCA1 deficiency and is associated with triple-negative and BRCA-mutated breast cancers.* Nat Struct Mol Biol, 2010. **17**(6): p. 688-95.
103. Bunting, S.F., et al., *53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks.* Cell, 2010. **141**(2): p. 243-54.
104. Chapman, J.R., et al., *BRCA1-associated exclusion of 53BP1 from DNA damage sites underlies temporal control of DNA repair.* J Cell Sci, 2012. **125**(Pt 15): p. 3529-34.
105. Jasin, M. and R. Rothstein, *Repair of strand breaks by homologous recombination.* Cold Spring Harb Perspect Biol, 2013. **5**(11): p. a012740.
106. Gaillard, H., T. Garcia-Muse, and A. Aguilera, *Replication stress and cancer.* Nat Rev Cancer, 2015. **15**(5): p. 276-89.
107. Pacek, M. and J.C. Walter, *A requirement for MCM7 and Cdc45 in chromosome unwinding during eukaryotic DNA replication.* EMBO J, 2004. **23**(18): p. 3667-76.
108. Zou, L. and S.J. Elledge, *Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes.* Science, 2003. **300**(5625): p. 1542-8.
109. MacDougall, C.A., et al., *The structural determinants of checkpoint activation.* Genes Dev, 2007. **21**(8): p. 898-903.

110. Chaturvedi, P., et al., *Mammalian Chk2 is a downstream effector of the ATM-dependent DNA damage checkpoint pathway.* Oncogene, 1999. **18**(28): p. 4047-54.
111. Matsuoka, S., et al., *Ataxia telangiectasia-mutated phosphorylates Chk2 in vivo and in vitro.* Proc Natl Acad Sci U S A, 2000. **97**(19): p. 10389-94.
112. Kumagai, A. and W.G. Dunphy, *Claspin, a novel protein required for the activation of Chk1 during a DNA replication checkpoint response in Xenopus egg extracts.* Mol Cell, 2000. **6**(4): p. 839-49.
113. Kumagai, A. and W.G. Dunphy, *Repeated phosphopeptide motifs in Claspin mediate the regulated binding of Chk1.* Nat Cell Biol, 2003. **5**(2): p. 161-5.
114. Liu, Q., et al., *Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint.* Genes Dev, 2000. **14**(12): p. 1448-59.
115. Jazayeri, A., et al., *ATM- and cell cycle-dependent regulation of ATR in response to DNA double-strand breaks.* Nat Cell Biol, 2006. **8**(1): p. 37-45.
116. Nam, E.A. and D. Cortez, *ATR signalling: more than meeting at the fork.* Biochem J, 2011. **436**(3): p. 527-36.
117. Richard, P. and J.L. Manley, *R Loops and Links to Human Disease.* J Mol Biol, 2017. **429**(21): p. 3168-3180.
118. Joshi, P., S. Halene, and O. Abdel-Wahab, *How do messenger RNA splicing alterations drive myelodysplasia?* Blood, 2017. **129**(18): p. 2465-2470.
119. Chan, Y.A., P. Hieter, and P.C. Stirling, *Mechanisms of genome instability induced by RNA-processing defects.* Trends Genet, 2014. **30**(6): p. 245-53.
120. Aguilera, A. and T. Garcia-Muse, *R loops: from transcription byproducts to threats to genome stability.* Mol Cell, 2012. **46**(2): p. 115-24.
121. Skourti-Stathaki, K., N.J. Proudfoot, and N. Gromak, *Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination.* Mol Cell, 2011. **42**(6): p. 794-805.
122. Stirling, P.C., et al., *R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants.* Genes & development, 2012. **26**(2): p. 163-175.
123. Gomez-Gonzalez, B., et al., *Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles.* The EMBO journal, 2011.
124. Parenteau, J., et al., *Deletion of many yeast introns reveals a minority of genes that require splicing for function.* Molecular biology of the cell, 2008. **19**(5): p. 1932-1941.

125. Chang, E.Y., et al., *RECQ-like helicases Sgs1 and BLM regulate R-loop-associated genome instability.* J Cell Biol, 2017. **216**(12): p. 3991-4005.
126. Novoa, C.A., J.S. Ang, and P.C. Stirling, *The A-Like Faker Assay for Measuring Yeast Chromosome III Stability*. Methods Mol Biol, 2018. **1672**: p. 1-9.
127. Costanzo, M., et al., *A global genetic interaction network maps a wiring diagram of cellular function.* Science, 2016. **353**(6306).
128. Young, B.P. and C.J. Loewen, *Balony: a software package for analysis of data generated by synthetic genetic array experiments.* BMC Bioinformatics, 2013. **14**: p. 354.
129. Supek, F., et al., *REVIGO summarizes and visualizes long lists of gene ontology terms.* PLoS One, 2011. **6**(7): p. e21800.
130. Usaj, M., et al., *TheCellMap.org: A Web-Accessible Database for Visualizing and Mining the Global Yeast Genetic Interaction Network.* G3 (Bethesda), 2017. **7**(5): p. 1539-1549.
131. Woodruff, J.B., D.G. Drubin, and G. Barnes, *Dynein-driven mitotic spindle positioning restricted to anaphase by She1p inhibition of dynactin recruitment.* Mol Biol Cell, 2009. **20**(13): p. 3003-11.
132. Wahba, L., et al., *RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability.* Molecular cell, 2011. **44**(6): p. 978-988.
133. Galy, V., et al., *Nuclear retention of unspliced mRNAs in yeast is mediated by perinuclear Mlp1*. Cell, 2004. **116**(1): p. 63-73.
134. de la Loza, M.C., R.E. Wellinger, and A. Aguilera, *Stimulation of direct-repeat recombination by RNA polymerase III transcription.* DNA Repair (Amst), 2009. **8**(5): p. 620-6.
135. Aksenova, A.Y., et al., *Genome rearrangements caused by interstitial telomeric sequences in yeast.* Proc Natl Acad Sci U S A, 2013. **110**(49): p. 19866-71.
136. Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis.* Nat Methods, 2012. **9**(7): p. 671-5.
137. Measday, V. and P.C. Stirling, *Navigating yeast genome maintenance with functional genomics.* Brief Funct Genomics, 2015.
138. Gavalda, S., et al., *Excess of Yra1 RNA-Binding Factor Causes Transcription-Dependent Genome Instability, Replication Impairment and Telomere Shortening.* PLoS Genet, 2016. **12**(4): p. e1005966.

139. Garcia-Rubio, M., et al., *Yra1-bound RNA-DNA hybrids cause orientation-independent transcription-replication collisions and telomere instability.* Genes Dev, 2018. **32**(13-14): p. 965-977.
140. Rodriguez-Navarro, S., K. Strasser, and E. Hurt, *An intron in the YRA1 gene is required to control Yra1 protein expression and mRNA export in yeast.* EMBO Rep, 2002. **3**(5): p. 438-42.
141. Barta, I. and R. Iggo, *Autoregulation of expression of the yeast Dbp2p 'DEAD-box' protein is mediated by sequences in the conserved DBP2 intron.* EMBO J, 1995. **14**(15): p. 3800-8.
142. Davis, C.A., et al., *Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast.* Nucleic Acids Res, 2000. **28**(8): p. 1700-6.
143. van Leeuwen, J., et al., *Exploring genetic suppression interactions on a global scale.* Science, 2016. **354**(6312).
144. Amon, A., et al., *Mechanisms that help the yeast cell cycle clock tick: G2 cyclins transcriptionally activate G2 cyclins and repress G1 cyclins.* Cell, 1993. **74**(6): p. 993-1007.
145. Lara-Gonzalez, P., F.G. Westhorpe, and S.S. Taylor, *The spindle assembly checkpoint.* Curr Biol, 2012. **22**(22): p. R966-80.
146. Hartwell, L.H. and T.A. Weinert, *Checkpoints: controls that ensure the order of cell cycle events.* Science, 1989. **246**(4930): p. 629-34.
147. Schatz, P.J., F. Solomon, and D. Botstein, *Isolation and characterization of conditional-lethal mutations in the TUB1 alpha-tubulin gene of the yeast Saccharomyces cerevisiae.* Genetics, 1988. **120**(3): p. 681-95.
148. Gronroos, E. and C. Lopez-Garcia, *Tolerance of Chromosomal Instability in Cancer: Mechanisms and Therapeutic Opportunities.* Cancer Res, 2018. **78**(23): p. 6529-6535.
149. Spingola, M., et al., *Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae.* RNA, 1999. **5**(2): p. 221-34.
150. Hoyt, M.A., T. Stearns, and D. Botstein, *Chromosome instability mutants of Saccharomyces cerevisiae that are defective in microtubule-mediated processes.* Molecular and cellular biology, 1990. **10**(1): p. 223-234.

151. Katz, W., B. Weinstein, and F. Solomon, *Regulation of tubulin levels and microtubule assembly in Saccharomyces cerevisiae: consequences of altered tubulin gene copy number.* Mol Cell Biol, 1990. **10**(10): p. 5286-94.
152. Tresini, M., et al., *The core spliceosome as target and effector of non-canonical ATM signalling.* Nature, 2015. **523**(7558): p. 53-8.
153. Papaemmanuil, E., et al., *Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts.* N Engl J Med, 2011. **365**(15): p. 1384-95.
154. Yoshida, K., et al., *Frequent pathway mutations of splicing machinery in myelodysplasia.* Nature, 2011. **478**(7367): p. 64-9.
155. Rossi, D., et al., *Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness.* Blood, 2011. **118**(26): p. 6904-8.
156. Quesada, V., A.J. Ramsay, and C. Lopez-Otin, *Chronic lymphocytic leukemia with SF3B1 mutation.* N Engl J Med, 2012. **366**(26): p. 2530.
157. Hou, H.A., et al., *Splicing factor mutations predict poor prognosis in patients with de novo acute myeloid leukemia.* Oncotarget, 2016. **7**(8): p. 9084-101.
158. Zhao, Y., et al., *High expression of Ki-67 acts a poor prognosis indicator in locally advanced nasopharyngeal carcinoma.* Biochem Biophys Res Commun, 2017. **494**(1-2): p. 390-396.
159. Furney, S.J., et al., *SF3B1 mutations are associated with alternative splicing in uveal melanoma.* Cancer Discov, 2013. **3**(10): p. 1122-1129.
160. Darman, R.B., et al., *Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point.* Cell Rep, 2015. **13**(5): p. 1033-45.
161. Carrocci, T.J., et al., *SF3b1 mutations associated with myelodysplastic syndromes alter the fidelity of branchsite selection in yeast.* Nucleic Acids Res, 2017. **45**(8): p. 4837-4852.
162. Mupo, A., et al., *Hemopoietic-specific Sf3b1-K700E knock-in mice display the splicing defect seen in human MDS but develop anemia without ring sideroblasts.* Leukemia, 2017. **31**(3): p. 720-727.
163. Gentien, D., et al., *A common alternative splicing signature is associated with SF3B1 mutations in malignancies from different cell lineages.* Leukemia, 2014. **28**(6): p. 1355-7.

164. Maguire, S.L., et al., *SF3B1 mutations constitute a novel therapeutic target in breast cancer.* J Pathol, 2015. **235**(4): p. 571-80.
165. Te Raa, G.D., et al., *The impact of SF3B1 mutations in CLL on the DNA-damage response.* Leukemia, 2015. **29**(5): p. 1133-42.
166. Qiu, J., et al., *Distinct splicing signatures affect converged pathways in myelodysplastic syndrome patients carrying mutations in different splicing regulators.* RNA, 2016. **22**(10): p. 1535-49.
167. Tanikawa, M., et al., *The spliceosome U2 snRNP factors promote genome stability through distinct mechanisms; transcription of repair factors and R-loop processing.* Oncogenesis, 2016. **5**(12): p. e280.
168. Ben-Aroya, S., et al., *Toward a comprehensive temperature-sensitive mutant repository of the essential genes of Saccharomyces cerevisiae.* Mol Cell, 2008. **30**(2): p. 248-58.
169. Kofoed, M., et al., *An Updated Collection of Sequence Barcoded Temperature-Sensitive Alleles of Yeast Essential Genes.* G3 (Bethesda), 2015. **5**(9): p. 1879-87.
170. Mathew, V., et al., *The histone-fold protein CHRAC14 influences chromatin composition in response to DNA damage.* Cell Rep, 2014. **7**(2): p. 321-330.
171. Dolatshad, H., et al., *Cryptic splicing events in the iron transporter ABCB7 and other key target genes in SF3B1-mutant myelodysplastic syndromes.* Leukemia, 2016. **30**(12): p. 2322-2331.
172. Cretu, C., et al., *Molecular Architecture of SF3b and Structural Consequences of Its Cancer-Related Mutations.* Mol Cell, 2016. **64**(2): p. 307-319.
173. Carrocci, T.J., J.C. Paulson, and A.A. Hoskins, *Functional analysis of Hsh155/SF3b1 interactions with the U2 snRNA/branch site duplex.* RNA, 2018. **24**(8): p. 1028-1040.
174. Gnugge, R. and F. Rudolf, *Saccharomyces cerevisiae Shuttle vectors.* Yeast, 2017. **34**(5): p. 205-221.
175. Parenteau, J., et al., *Deletion of many yeast introns reveals a minority of genes that require splicing for function.* Mol Biol Cell, 2008. **19**(5): p. 1932-41.
176. Parenteau, J. and S. Abou Elela, *Introns: Good Day Junk Is Bad Day Treasure.* Trends Genet, 2019. **35**(12): p. 923-934.
177. Blencowe, B.J., *Splicing regulation: the cell cycle connection.* Curr Biol, 2003. **13**(4): p. R149-51.

178. Burns, C.G. and K.L. Gould, *Connections between pre-mRNA processing and regulation of the eukaryotic cell cycle.* Front Horm Res, 1999. **25**: p. 59-82.
179. Burns, C.G., et al., *Removal of a single alpha-tubulin gene intron suppresses cell cycle arrest phenotypes of splicing factor mutations in Saccharomyces cerevisiae.* Mol Cell Biol, 2002. **22**(3): p. 801-15.
180. Tang, Q., et al., *SF3B1/Hsh155 HEAT motif mutations affect interaction with the spliceosomal ATPase Prp5, resulting in altered branch site selectivity in pre-mRNA splicing.* Genes Dev, 2016. **30**(24): p. 2710-2723.
181. Lesser, C.F. and C. Guthrie, *Mutational analysis of pre-mRNA splicing in Saccharomyces cerevisiae using a sensitive new reporter gene, CUP1.* Genetics, 1993. **133**(4): p. 851-63.
182. Xu, Y.Z. and C.C. Query, *Competition between the ATPase Prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly.* Mol Cell, 2007. **28**(5): p. 838-49.
183. DeBoever, C., et al., *Transcriptome sequencing reveals potential mechanism of cryptic 3' splice site selection in SF3B1-mutated cancers.* PLoS Comput Biol, 2015. **11**(3): p. e1004105.
184. Alsafadi, S., et al., *Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage.* Nat Commun, 2016. **7**: p. 10615.
185. Gould, G.M., et al., *Identification of new branch points and unconventional introns in Saccharomyces cerevisiae.* RNA, 2016. **22**(10): p. 1522-34.
186. Grate, L. and M. Ares, Jr., *Searching yeast intron data at Ares lab Web site.* Methods Enzymol, 2002. **350**: p. 380-92.
187. Kinner, A., et al., *Gamma-H2AX in recognition and signaling of DNA double-strand breaks in the context of chromatin.* Nucleic Acids Res, 2008. **36**(17): p. 5678-94.
188. Fernandez-Capetillo, O., et al., *H2AX: the histone guardian of the genome.* DNA Repair (Amst), 2004. **3**(8-9): p. 959-67.
189. Singh, N.P. and R.E. Stephens, *Microgel electrophoresis: sensitivity, mechanisms, and DNA electrostretching.* Mutat Res, 1997. **383**(2): p. 167-75.
190. Collins, A.R., *The comet assay for DNA damage and repair: principles, applications, and limitations.* Mol Biotechnol, 2004. **26**(3): p. 249-61.
191. Boguslawski, S.J., et al., *Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids.* J Immunol Methods, 1986. **89**(1): p. 123-30.

192. Hartono, S.R., et al., *The Affinity of the S9.6 Antibody for Double-Stranded RNAs Impacts the Accurate Mapping of R-Loops in Fission Yeast.* J Mol Biol, 2018. **430**(3): p. 272-284.
193. Cerritelli, S.M. and R.J. Crouch, *Ribonuclease H: the enzymes in eukaryotes.* FEBS J, 2009. **276**(6): p. 1494-505.
194. Huertas, P. and A. Aguilera, *Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination.* Mol Cell, 2003. **12**(3): p. 711-21.
195. Wellinger, R.E., F. Prado, and A. Aguilera, *Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex.* Mol Cell Biol, 2006. **26**(8): p. 3327-34.
196. Tuduri, S., et al., *Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription.* Nat Cell Biol, 2009. **11**(11): p. 1315-24.
197. Aguilera, A. and B. Gomez-Gonzalez, *Genome instability: a mechanistic view of its causes and consequences.* Nat Rev Genet, 2008. **9**(3): p. 204-17.
198. Gan, W., et al., *R-loop-mediated genomic instability is caused by impairment of replication fork progression.* Genes Dev, 2011. **25**(19): p. 2041-56.
199. Salic, A. and T.J. Mitchison, *A chemical method for fast and sensitive detection of DNA synthesis in vivo.* Proc Natl Acad Sci U S A, 2008. **105**(7): p. 2415-20.
200. Buck, S.B., et al., *Detection of S-phase cell cycle progression using 5-ethynyl-2'-deoxyuridine incorporation with click chemistry, an alternative to using 5-bromo-2'-deoxyuridine antibodies.* Biotechniques, 2008. **44**(7): p. 927-9.
201. Fei, L. and H. Xu, *Role of MCM2-7 protein phosphorylation in human cancer cells.* Cell Biosci, 2018. **8**: p. 43.
202. Lei, M., et al., *Mcm2 is a target of regulation by Cdc7-Dbf4 during the initiation of DNA synthesis.* Genes Dev, 1997. **11**(24): p. 3365-74.
203. Ramer, M.D., et al., *Dbf4 and Cdc7 proteins promote DNA replication through interactions with distinct Mcm2-7 protein subunits.* J Biol Chem, 2013. **288**(21): p. 14926-35.
204. Hamperl, S., et al., *Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses.* Cell, 2017. **170**(4): p. 774-786 e19.

205. Olson, E., et al., *RPA2 is a direct downstream target for ATR to regulate the S-phase checkpoint.* J Biol Chem, 2006. **281**(51): p. 39517-33.
206. Vassin, V.M., et al., *Human RPA phosphorylation by ATR stimulates DNA synthesis and prevents ssDNA accumulation during DNA-replication stress.* J Cell Sci, 2009. **122**(Pt 22): p. 4070-80.
207. Xu, N., et al., *Cdk-mediated phosphorylation of Chk1 is required for efficient activation and full checkpoint proficiency in response to DNA damage.* Oncogene, 2012. **31**(9): p. 1086-94.
208. Gamper, A.M., et al., *ATR kinase activation in G1 phase facilitates the repair of ionizing radiation-induced DNA damage.* Nucleic Acids Res, 2013. **41**(22): p. 10334-44.
209. Boe, C.A., et al., *A novel role for ATR/Rad3 in G1 phase.* Sci Rep, 2018. **8**(1): p. 6880.
210. Delcuve, G.P., S. He, and J.R. Davie, *Mitotic partitioning of transcription factors.* J Cell Biochem, 2008. **105**(1): p. 1-8.
211. Hsiung, C.C., et al., *A hyperactive transcriptional state marks genome reactivation at the mitosis-G1 transition.* Genes Dev, 2016. **30**(12): p. 1423-39.
212. Marechal, A. and L. Zou, *RPA-coated single-stranded DNA as a platform for post-translational modifications in the DNA damage response.* Cell Res, 2015. **25**(1): p. 9-23.
213. Nguyen, H.D., et al., *Functions of Replication Protein A as a Sensor of R Loops and a Regulator of RNaseH1.* Mol Cell, 2017. **65**(5): p. 832-847 e4.
214. Montagnoli, A., et al., *A Cdc7 kinase inhibitor restricts initiation of DNA replication and has antitumor activity.* Nat Chem Biol, 2008. **4**(6): p. 357-65.
215. Jiang, W., et al., *Mammalian Cdc7-Dbf4 protein kinase complex is essential for initiation of DNA replication.* EMBO J, 1999. **18**(20): p. 5703-13.
216. Aronica, L., et al., *The spliceosome-associated protein Nrl1 suppresses homologous recombination-dependent R-loop formation in fission yeast.* Nucleic Acids Res, 2016. **44**(4): p. 1703-17.
217. Amon, J.D. and D. Koshland, *RNase H enables efficient repair of R-loop induced DNA damage.* Elife, 2016. **5**.
218. Deem, A., et al., *Break-induced replication is highly inaccurate.* PLoS Biol, 2011. **9**(2): p. e1000594.
219. Malkova, A. and G. Ira, *Break-induced replication: functions and molecular mechanism.* Curr Opin Genet Dev, 2013. **23**(3): p. 271-9.

220. Pellagatti, A., et al., *Impact of spliceosome mutations on RNA splicing in myelodysplasia: dysregulated genes/pathways and clinical associations.* Blood, 2018. **132**(12): p. 1225-1240.
221. Dolatshad, H., et al., *Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells.* Leukemia, 2015. **29**(5): p. 1092-103.
222. Dean, M., Y. Hamon, and G. Chimini, *The human ATP-binding cassette (ABC) transporter superfamily.* J Lipid Res, 2001. **42**(7): p. 1007-17.
223. King, S.M. and R.S. Patel-King, *The M(r) = 8,000 and 11,000 outer arm dynein light chains from Chlamydomonas flagella have cytoplasmic homologues.* J Biol Chem, 1995. **270**(19): p. 11445-52.
224. Barbar, E., *Dynein light chain LC8 is a dimerization hub essential in diverse protein networks.* Biochemistry, 2008. **47**(2): p. 503-8.
225. King, S.M., *Dynein-independent functions of DYNLL1/LC8: redox state sensing and transcriptional control.* Sci Signal, 2008. **1**(47): p. pe51.
226. Rapali, P., et al., *Directed evolution reveals the binding motif preference of the LC8/DYNLL hub protein and predicts large numbers of novel binders in the human proteome.* PLoS One, 2011. **6**(4): p. e18818.
227. Puthalakath, H., et al., *The proapoptotic activity of the Bcl-2 family member Bim is regulated by interaction with the dynein motor complex.* Mol Cell, 1999. **3**(3): p. 287-96.
228. Yien, Y.Y., et al., *TMEM14C is required for erythroid mitochondrial heme metabolism.* J Clin Invest, 2014. **124**(10): p. 4294-304.
229. Kremer, B.E., L.A. Adang, and I.G. Macara, *Septins regulate actin organization and cell-cycle arrest through nuclear accumulation of NCK mediated by SOCS7.* Cell, 2007. **130**(5): p. 837-50.
230. Rivas, M.A., et al., *Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome.* Science, 2015. **348**(6235): p. 666-9.
231. Jaffrey, S.R. and S.H. Snyder, *PIN: an associated protein inhibitor of neuronal nitric oxide synthase.* Science, 1996. **274**(5288): p. 774-7.
232. Jurado, S., et al., *ATM substrate Chk2-interacting Zn2+ finger (ASCIZ) Is a bi-functional transcriptional activator and feedback sensor in the regulation of dynein light chain (DYNLL1) expression.* J Biol Chem, 2012. **287**(5): p. 3156-64.

233. McNees, C.J., et al., *ASCIZ regulates lesion-specific Rad51 focus formation and apoptosis after methylating DNA damage.* EMBO J, 2005. **24**(13): p. 2447-57.
234. Kanu, N., et al., *The ATM cofactor ATMIN protects against oxidative stress and accumulation of DNA damage in the aging brain.* J Biol Chem, 2010. **285**(49): p. 38534-42.
235. Wong, D.M., et al., *The Transcription Factor ASCIZ and Its Target DYNLL1 Are Essential for the Development and Expansion of MYC-Driven B Cell Lymphoma.* Cell Rep, 2016. **14**(6): p. 1488-1499.
236. Becker, J.R., et al., *The ASCIZ-DYNLL1 axis promotes 53BP1-dependent non-homologous end joining and PARP inhibitor sensitivity.* Nat Commun, 2018. **9**(1): p. 5406.
237. Lo, K.W., et al., *The 8-kDa dynein light chain binds to p53-binding protein 1 and mediates DNA damage-induced p53 nuclear accumulation.* J Biol Chem, 2005. **280**(9): p. 8172-9.
238. He, Y.J., et al., *DYNLL1 binds to MRE11 to limit DNA end resection in BRCA1-deficient cells.* Nature, 2018. **563**(7732): p. 522-526.
239. Kilic, S., et al., *Phase separation of 53BP1 determines liquid-like behavior of DNA repair compartments.* EMBO J, 2019. **38**(16): p. e101379.
240. Allen, C., et al., *More forks on the road to replication stress recovery.* J Mol Cell Biol, 2011. **3**(1): p. 4-12.
241. Michl, J., et al., *FANCD2 limits replication stress and genome instability in cells lacking BRCA2.* Nat Struct Mol Biol, 2016. **23**(8): p. 755-757.
242. Feng, W. and M. Jasin, *BRCA2 suppresses replication stress-induced mitotic and G1 abnormalities through homologous recombination.* Nat Commun, 2017. **8**(1): p. 525.
243. Dupre, A., et al., *A forward chemical genetic screen reveals an inhibitor of the Mre11-Rad50-Nbs1 complex.* Nat Chem Biol, 2008. **4**(2): p. 119-25.
244. Shibata, A., et al., *DNA double-strand break repair pathway choice is directed by distinct MRE11 nuclease activities.* Mol Cell, 2014. **53**(1): p. 7-18.
245. Dalton, W.B., et al., *Hotspot SF3B1 mutations induce metabolic reprogramming and vulnerability to serine deprivation.* J Clin Invest, 2019. **130**: p. 4708-4723.
246. Tam, A.S. and P.C. Stirling, *Splicing, genome stability and disease: splice like your genome depends on it!* Curr Genet, 2019. **65**(4): p. 905-912.

247. Obeng, E.A., et al., *Physiologic Expression of Sf3b1(K700E) Causes Impaired Erythropoiesis, Aberrant Splicing, and Sensitivity to Therapeutic Spliceosome Modulation.* Cancer Cell, 2016. **30**(3): p. 404-417.
248. Stirling, P.C., et al., *The complete spectrum of yeast chromosome instability genes identifies candidate CIN cancer genes and functional roles for ASTRA complex components.* PLoS Genet, 2011. **7**(4): p. e1002057.
249. Salas-Armenteros, I., et al., *Depletion of the MFAP1/SPP381 Splicing Factor Causes R-Loop-Independent Genome Instability.* Cell Rep, 2019. **28**(6): p. 1551-1563 e7.
250. Jimenez, M., et al., *Splicing events in the control of genome integrity: role of SLU7 and truncated SRSF3 proteins.* Nucleic Acids Res, 2019. **47**(7): p. 3450-3466.
251. Gomez-Gonzalez, B., et al., *Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles.* EMBO J, 2011. **30**(15): p. 3106-19.
252. Bonnet, A., H. Bretes, and B. Palancade, *Nuclear pore components affect distinct stages of intron-containing gene expression.* Nucleic Acids Res, 2015. **43**(8): p. 4249-61.
253. Zhang, J., et al., *Disease-Causing Mutations in SF3B1 Alter Splicing by Disrupting Interaction with SUGP1.* Mol Cell, 2019. **76**(1): p. 82-95 e7.

Appendices

Appendix 1. SGA scores and hit list from *hsh155-1* screen.

Linked genes and and uracil biosynthetic genes are marked in red and were not included as true negative hits. The list is ordered in 4 groups, which reflect alleles with significant p-values broken into negative hits (E-C <0.5), negative scores above this threshold, positive scores below the threshold, and positive hits (E-C >0.5).

Part I: Raw Score List for SGA

ALLELE	E-C value	Standard Error	P-value	LIST
sec12-4	-1.6670748	0.096081318	6.48E-05	Negative Hit
yhc1-1	-1.4430814	0.076931593	4.76E-05	Negative Hit
irr1-1	-1.2533872	0.070037616	5.73E-05	Negative Hit
bet1-1	-1.2294623	0.081703011	0.00011365	Negative Hit
ame1-4	-1.1949676	0.083470104	0.00013831	Negative Hit
cof1-8	-1.1727876	0.05248632	2.38E-05	Negative Hit
mob2-34	-1.1657811	0.08529095	0.00016594	Negative Hit
YJL130C	-1.0918865	0.051831554	3.00E-05	Uracil pathway
luc7-1	-1.0404847	0.040952013	1.43E-05	Negative Hit
mob2-22	-1.0381851	0.118913035	0.00094822	Negative Hit
pse1-41	-1.0212777	0.116503835	0.00093362	Linked
nnf1-77	-1.019193	0.102781604	0.00058063	Negative Hit
YLR078C	-1.0144774	0.022385933	1.42E-06	Negative Hit
prp16-ts	-1.009512	0.091735418	0.00038754	Negative Hit
cdc1-2	-1.0036835	0.062549085	8.82E-05	Negative Hit
YLR420W	-0.9941896	0.055172346	5.58E-05	Uracil pathway
stu1-12	-0.9924442	0.072572031	0.00016561	Negative Hit
uso1-1	-0.985901	0.029897517	5.04E-06	Negative Hit
YMR289W	-0.9840513	0.040568436	1.71E-05	linked
las17-1	-0.9746828	0.186264328	0.00637152	Negative Hit
sec23-1	-0.9592566	0.197092918	0.00823735	Negative Hit
YMR288W	-0.9511593	0.029750829	5.71E-06	Linked
YLR368W	-0.9426114	0.105083665	0.00085469	Negative Hit
YLR085C	-0.939447	0.055525544	7.15E-05	Negative Hit
mob2-11	-0.93565	0.100627022	0.00074437	Negative Hit
mob2-36	-0.9340399	0.06829159	0.00016552	Negative Hit
YDR090C	-0.9300439	0.067418499	0.00016003	Negative Hit
YMR285C	-0.9214785	0.060966987	0.00011169	Linked
YHR086W	-0.9204186	0.051107863	5.59E-05	Negative Hit
YNL258C	-0.9034749	0.067634443	0.0001816	Negative Hit

YOR308C	-0.9032615	0.05143666	6.18E-05	Negative Hit
YML041C	-0.900494	0.045200248	3.75E-05	Negative Hit
YDR334W	-0.8929859	0.05061852	6.06E-05	Negative Hit
cdc28-td	-0.8926709	0.056229423	9.20E-05	Negative Hit
YHR030C	-0.8910358	0.052241231	6.93E-05	Negative Hit
sec26-F856AW860A	-0.8860211	0.043229101	3.35E-05	Negative Hit
snu114-60	-0.8843177	0.085359129	0.00049002	Negative Hit
YPR082C	-0.8714206	0.039780171	2.57E-05	Negative Hit
mif2-3	-0.8600568	0.058097078	0.00012122	Negative Hit
abf1-103	-0.8540169	0.176241851	0.00836538	Negative Hit
rpb3-2	-0.8525871	0.102310575	0.00113317	Negative Hit
YMR291W	-0.8428722	0.154762495	0.00552023	Linked
las17-14	-0.8306812	0.081039297	0.00051066	Negative Hit
YKL216W	-0.8280795	0.04608792	5.64E-05	Uracil pathway
hrp1-7	-0.8250982	0.061084155	0.00017384	Negative Hit
YOR327C	-0.8205854	0.095067788	0.00099059	Negative Hit
mtw1-ts	-0.8177015	0.077029571	0.0004458	Negative Hit
YAL011W	-0.8106727	0.043786405	5.01E-05	Negative Hit
mob2-40	-0.8049846	0.11277175	0.00203711	Negative Hit
YKL074C	-0.790008	0.053314912	0.00012077	Negative Hit
YEL061C	-0.7879271	0.104125223	0.00163488	Negative Hit
YLR268W	-0.7872938	0.063850014	0.00024857	Negative Hit
YMR286W	-0.7697947	0.043668592	6.08E-05	Linked
cdc34-2	-0.7683087	0.122165598	0.00326536	Negative Hit
pan1-4	-0.7660759	0.04190564	5.27E-05	Negative Hit
mcd1-73	-0.7615815	0.084468455	0.00083803	Negative Hit
sqt1-201	-0.7508929	0.088844845	0.00107371	Negative Hit
YDR017C	-0.7482997	0.043303375	6.58E-05	Negative Hit
YJL095W	-0.7441054	0.053532022	0.00015532	Negative Hit
sgv1-23	-0.7415739	0.099843516	0.00175417	Negative Hit
YOR058C	-0.7273548	0.073258011	0.00057779	Negative Hit
sft1-15	-0.7230674	0.073416341	0.00059612	Negative Hit
YKL048C	-0.7159452	0.034656662	3.24E-05	Negative Hit
abf1-102	-0.7013729	0.052616806	0.00018312	Negative Hit
dsn1-7	-0.6987609	0.101882713	0.00236627	Negative Hit
YNL249C	-0.6963337	0.067912115	0.00051006	Negative Hit
YDR273W	-0.6897353	0.059210478	0.00031044	Negative Hit
spc24 4-2	-0.6809539	0.129059313	0.0061857	Negative Hit
YIL062C	-0.6783126	0.052679611	0.00020977	Negative Hit
YOR216C	-0.6758798	0.037351815	5.48E-05	Negative Hit
taf1-1	-0.6752385	0.094291601	0.00201263	Negative Hit
YIR005W	-0.6722183	0.026807248	1.50E-05	Negative Hit

YPL055C	-0.6585474	0.047672553	0.00015917	Negative Hit
pkc1-2	-0.6517669	0.116194117	0.00496226	Negative Hit
cdc10-1	-0.6470596	0.096939977	0.00261844	Negative Hit
YOR027W	-0.6463279	0.042100485	0.00010503	Negative Hit
YEL062W	-0.6453486	0.098483623	0.00280441	Negative Hit
YLR337C	-0.6448177	0.031946728	3.56E-05	Negative Hit
YLL038C	-0.6429763	0.024482106	1.25E-05	Negative Hit
taf10-ts34	-0.6399109	0.084938945	0.0016624	Negative Hit
ask1-2	-0.6377868	0.078117857	0.00122523	Negative Hit
taf8-7	-0.6334957	0.106208175	0.00396747	Negative Hit
YNL286W	-0.6330967	0.039619541	8.97E-05	Negative Hit
ssl2-ts	-0.632111	0.075451259	0.0011104	Negative Hit
med7-141	-0.6315624	0.085488501	0.00178997	Negative Hit
YPL079W	-0.628038	0.062551537	0.00055331	Negative Hit
act1-108	-0.6242737	0.083318437	0.00169717	Negative Hit
YPL106C	-0.6238352	0.067914326	0.00078012	Negative Hit
cog2-1	-0.6216964	0.113152089	0.00534795	Negative Hit
sln1-ts4	-0.6092648	0.058406482	0.00047711	Negative Hit
tub4-?DSY	-0.5964548	0.097126162	0.00356503	Negative Hit
nnf1-48	-0.593916	0.088520329	0.00256859	Negative Hit
YPR095C	-0.5915616	0.036111454	8.13E-05	Negative Hit
act1-3	-0.5888086	0.081868076	0.00198023	Negative Hit
YJR084W	-0.5852585	0.057879477	0.00053834	Negative Hit
YNL069C	-0.5849932	0.069990349	0.00112035	Negative Hit
YGR080W	-0.5840037	0.053617427	0.00040336	Negative Hit
ret3-1	-0.5828721	0.062480074	0.00073497	Negative Hit
ssu72-2	-0.5728064	0.085599941	0.00259401	Negative Hit
YNL097C	-0.570089	0.040430299	0.00014682	Negative Hit
YNL266W	-0.5662038	0.065626434	0.00099231	Negative Hit
ask1-3	-0.5645261	0.080030672	0.00213002	Negative Hit
YBR036C	-0.5613982	0.063563395	0.00090711	Negative Hit
YBR189W	-0.5568713	0.035322144	9.46E-05	Negative Hit
mob2-14	-0.553198	0.120041902	0.00996784	Negative Hit
duo1-2	-0.5530201	0.08817593	0.00329873	Negative Hit
YMR280C	-0.5514462	0.095037274	0.00438798	Linked
sec39-1	-0.5500379	0.077718773	0.0021037	Negative Hit
rna15-58	-0.5489952	0.08522245	0.00298783	Negative Hit
spc105-4	-0.543954	0.074870572	0.00190631	Negative Hit
erg11-td	-0.5398286	0.08833748	0.00362993	Negative Hit
YFL018C	-0.5375316	0.048090852	0.00036472	Negative Hit
YKL114C	-0.537384	0.046273949	0.00031419	Negative Hit
act1-119	-0.5359062	0.08401948	0.00309957	Negative Hit

YDL043C	-0.5356472	0.024175227	2.46E-05	Negative Hit
YML106W	-0.5343398	0.051757009	0.00049667	Uracil pathway
YJL183W	-0.5334958	0.04185748	0.00021833	Negative Hit
dbf2-3	-0.5318166	0.0821964	0.00293991	Negative Hit
YJL168C	-0.5314176	0.030455169	6.33E-05	Negative Hit
YMR179W	-0.5310119	0.040649193	0.00019823	Linked
YLR413W	-0.5207211	0.059106122	0.0009159	Negative Hit
sgv1-35	-0.519888	0.087838958	0.00408151	Negative Hit
act1-136	-0.5191726	0.062153339	0.00112296	Negative Hit
dsn1-8	-0.5144537	0.076703024	0.00257186	Negative Hit
YKR024C	-0.5112959	0.047218535	0.00041268	Negative Hit
YGL192W	-0.5060326	0.05742322	0.00091494	Negative Hit
YLR211C	-0.5052152	0.037643647	0.00017828	Negative Hit
abf1-101	-0.5026232	0.068325493	0.00181894	Negative Hit
YER113C	-0.5021809	0.065052399	0.00151592	Negative Hit
YDL006W	0.5012971	0.071456605	0.00217414	Positive Hit
YJL064W	0.50174206	0.064443734	0.0014677	Positive Hit
YIL032C	0.50234977	0.101288567	0.00770836	Positive Hit
mex67-ts5	0.50508734	0.071536171	0.00212243	Positive Hit
YJL175W	0.50550237	0.066292549	0.00158816	Positive Hit
YKL176C	0.5056668	0.025671909	3.92E-05	Positive Hit
YPL062W	0.50602045	0.090590948	0.00503833	Positive Hit
YBL071W-A	0.50988785	0.051343915	0.00057731	Positive Hit
lcb2-2	0.51139705	0.054249682	0.00070601	Positive Hit
YEL033W	0.51384622	0.054200225	0.00069068	Positive Hit
YPR133W-A	0.51508207	0.031658699	8.35E-05	Positive Hit
rho3-Ser228	0.51567836	0.069028219	0.00171619	Positive Hit
YOR157C	0.51791797	0.022758809	2.21E-05	Positive Hit
YMR253C	0.5179267	0.062777033	0.00117733	Positive Hit
YBR266C	0.52798646	0.053640661	0.00059749	Positive Hit
YML094W	0.52985583	0.066539092	0.00134739	Positive Hit
YCR034W	0.52992824	0.06415303	0.00117184	Positive Hit
YDR463W	0.53014934	0.064441525	0.00119017	Positive Hit
YGL105W	0.53312	0.058356588	0.00079669	Positive Hit
YGR200C	0.53329693	0.057116604	0.00073253	Positive Hit
YJR117W	0.5353901	0.045528535	0.00029919	Positive Hit
ura6-5	0.53645035	0.065961506	0.00124349	Positive Hit
YLR074C	0.53754808	0.081110013	0.00268892	Positive Hit
YOR251C	0.53759287	0.053615452	0.0005562	Positive Hit
sec10-2	0.5377821	0.102345708	0.00627767	Positive Hit

YFL002C	0.54034845	0.078069649	0.00228686	Positive Hit
YGR285C	0.54064191	0.054639575	0.00058551	Positive Hit
cdc123-4	0.54356141	0.063535765	0.00102488	Positive Hit
emg1-1	0.54458754	0.040558042	0.00017795	Positive Hit
YOR138C	0.5456581	0.080127144	0.00242996	Positive Hit
gcd1-502	0.55448437	0.082415895	0.00254235	Positive Hit
nop2-3	0.55779354	0.060180472	0.00075354	Positive Hit
YLR403W	0.55828931	0.063351217	0.00091484	Positive Hit
YML092C	0.55907793	0.041856521	0.00018166	Positive Hit
YKL213C	0.55973253	0.053612785	0.00047555	Positive Hit
sec59-ts	0.56922561	0.111080843	0.006865	Positive Hit
cdc48-4601	0.57288231	0.068443778	0.00111427	Positive Hit
YPR103W	0.57385432	0.040739641	0.00014742	Positive Hit
YCR044C	0.57809988	0.054306453	0.00044098	Positive Hit
sec11-2	0.57832924	0.104926963	0.00528731	Positive Hit
YOR117W	0.58666594	0.064152476	0.00079358	Positive Hit
cdc39-1	0.58973832	0.087072531	0.00247981	Positive Hit
YML010C-B	0.59147608	0.096946489	0.00365168	Positive Hit
YJR118C	0.59316374	0.055223454	0.00042585	Positive Hit
YDR363W-A	0.59437005	0.037687834	9.44E-05	Positive Hit
YGR178C	0.59465372	0.093722417	0.00316057	Positive Hit
YDR382W	0.59907179	0.08130634	0.001808	Positive Hit
pre2-2	0.59985183	0.114634707	0.00637182	Positive Hit
YOL038C-A	0.60786937	0.114784777	0.00610477	Positive Hit
nop4-3	0.61476612	0.096612557	0.00312695	Positive Hit
YHR064C	0.62835306	0.05170427	0.00026308	Positive Hit
YGR279C	0.63104185	0.12174686	0.00659166	Positive Hit
YLR374C	0.63341657	0.090482881	0.00219166	Positive Hit
alg2-1	0.63348434	0.079081725	0.00131731	Positive Hit
afg2-18	0.64291249	0.083086167	0.0015024	Positive Hit
lcb1-5	0.65335324	0.0684461	0.00067271	Positive Hit
YBR267W	0.65613787	0.129232436	0.00709464	Positive Hit
YCL032W	0.67049339	0.05764436	0.00031225	Positive Hit
YLR372W	0.67298636	0.042965361	9.70E-05	Positive Hit
YLR402W	0.68026392	0.053701874	0.00022365	Positive Hit
YLR021W	0.68092637	0.039026908	6.34E-05	Positive Hit
YMR314W	0.68417265	0.036728081	4.89E-05	Positive Hit
YMR259C	0.68649195	0.128228593	0.00587122	Positive Hit
YKL010C	0.68701996	0.046094991	0.00011802	Positive Hit
YBR058C	0.70725425	0.053646605	0.00019122	Positive Hit
taf9-ts2	0.70920751	0.100347943	0.00211466	Positive Hit
qri1-ts6	0.71395083	0.096984697	0.00181412	Positive Hit

YLR373C	0.7195316	0.044284648	8.40E-05	Positive Hit
YMR060C	0.75187669	0.071605653	0.00046509	Positive Hit
rpn1-821	0.75421432	0.089226889	0.0010732	Positive Hit
nop7-1	0.76173521	0.114016702	0.0026096	Positive Hit
YBR021W	0.76956541	0.058866382	0.00019764	Positive Hit
YJL117W	0.77317938	0.163745549	0.00915916	Positive Hit
cdc48-3	0.77765383	0.13962437	0.00509142	Positive Hit
YML084W	0.77904413	0.073999583	0.0004604	Positive Hit
YPL161C	0.78304142	0.085200181	0.00077849	Positive Hit
YER012W	0.78724549	0.056132654	0.00014997	Positive Hit
YOR293W	0.81576077	0.053297023	0.00010628	Positive Hit
YBR173C	0.82163272	0.027271499	7.23E-06	Positive Hit
dcp2-7	0.82542506	0.133696583	0.00349574	Positive Hit
YFR012W	0.83550264	0.075295078	0.00037521	Positive Hit
pre2-V214A	0.85394019	0.144244319	0.00407786	Positive Hit
YIL009C-A	0.8576141	0.104597697	0.00120555	Positive Hit
rpt4-145	0.87584199	0.121112625	0.00193988	Positive Hit
YIL024C	0.88052346	0.078587242	0.00036131	Positive Hit
YER079W	0.88173951	0.129506547	0.00243189	Positive Hit
sup45-ts	0.89361403	0.18751236	0.00886848	Positive Hit
nop2-4	0.90973192	0.1049331	0.00097404	Positive Hit
YNL091W	0.97918672	0.155739031	0.00326866	Positive Hit
YER086W	0.98237847	0.15542467	0.00320559	Positive Hit
YCR063W	0.99401597	0.088472649	0.00035745	Positive Hit
ura6-4	1.06816758	0.086330654	0.00024523	Positive Hit
YGR135W	1.12624518	0.059437407	4.57E-05	Positive Hit
rpt6-20	1.17285551	0.195444812	0.00388027	Positive Hit
YML081W	1.19220714	0.145196663	0.00119895	Positive Hit
YML083C	1.204461	0.136806168	0.00091821	Positive Hit
YML080W	1.34333134	0.155794961	0.00099462	Positive Hit
YML061C	1.3521638	0.125824522	0.00042504	Positive Hit
YML088W	1.35263305	0.157180102	0.00100208	Positive Hit
YML082W	1.35371157	0.161412714	0.0011059	Positive Hit
YML078W	1.46041589	0.151652946	0.00065021	Positive Hit
YML089C	1.50853238	0.132148503	0.00033596	Positive Hit
YML063W	1.51307	0.187701429	0.00128615	Positive Hit
YML087C	1.54745837	0.18449682	0.00110549	Positive Hit
YML079W	1.58736523	0.201130572	0.00139395	Positive Hit
YML051W	1.63946429	0.111777491	0.00012572	Positive Hit
exo84-102	1.70165445	0.158422657	0.00042584	Positive Hit
ndc1-4	2.08819287	0.228617711	0.00079722	Positive Hit

Part II: negative-genetic interaction hit list used in subsequent analyses

ALLELE	E-C value	Standard error	P-value	Gene name	Intron
sec12-4	-1.6670748	0.096081318	6.48E-05	SEC12	NO
yhc1-1	-1.4430814	0.076931593	4.76E-05	YHC1	NO
irr1-1	-1.2533872	0.070037616	5.73E-05	IRR1	NO
bet1-1	-1.2294623	0.081703011	0.000114	BET1	YES
ame1-4	-1.1949676	0.083470104	0.000138	AME1	NO
cof1-8	-1.1727876	0.05248632	2.38E-05	COF1	YES
mob2-34	-1.1657811	0.08529095	0.000166	MOB2	YES
luc7-1	-1.0404847	0.040952013	1.43E-05	LUC7	NO
mob2-22	-1.0381851	0.118913035	0.000948	MOB2	YES
nnf1-77	-1.019193	0.102781604	0.000581	NNF1	NO
YLR078C	-1.0144774	0.022385933	1.42E-06	BOS1	YES
prp16-ts	-1.009512	0.091735418	0.000388	PRP16	NO
cdc1-2	-1.0036835	0.062549085	8.82E-05	CDC1	NO
stu1-12	-0.9924442	0.072572031	0.000166	STU1	NO
uso1-1	-0.985901	0.029897517	5.04E-06	USO1	NO
las17-1	-0.9746828	0.186264328	0.006372	LAS17	NO
sec23-1	-0.9592566	0.197092918	0.008237	SEC23	NO
YLR368W	-0.9426114	0.105083665	0.000855	MDM30	NO
YLR085C	-0.939447	0.055525544	7.15E-05	ARP6	NO
mob2-11	-0.93565	0.100627022	0.000744	MOB2	YES
mob2-36	-0.9340399	0.06829159	0.000166	MOB2	YES
YDR090C	-0.9300439	0.067418499	0.00016	YDR090C	NO

Total genes in list containing:

No intron	93
Intron	10

Total genes in SGD containing:

No intron	5905
Intron	273

Fisher's exact test:
The two-tailed P value equals 0.0255

YHR086W	-0.9204186	0.051107863	5.59E-05	NAM8	NO
YNL258C	-0.9034749	0.067634443	0.000182	DSL1	NO
YOR308C	-0.9032615	0.05143666	6.18E-05	SNU66	NO
YML041C	-0.900494	0.045200248	3.75E-05	VPS71	NO
YDR334W	-0.8929859	0.05061852	6.06E-05	SWR1	NO
cdc28-td	-0.8926709	0.056229423	9.20E-05	CDC28	NO
YHR030C	-0.8910358	0.052241231	6.93E-05	SLT2	NO
sec26-F856AW860A	-0.8860211	0.043229101	3.35E-05	SEC26	NO
snu114-60	-0.8843177	0.085359129	0.00049	SNU114	NO
YPR082C	-0.8714206	0.039780171	2.57E-05	DIB1	NO
mif2-3	-0.8600568	0.058097078	0.000121	MIF	NO
abf1-103	-0.8540169	0.176241851	0.008365	ABF1	NO
rpb3-2	-0.8525871	0.102310575	0.001133	RPB3	NO
las17-14	-0.8306812	0.081039297	0.000511	LAS17	NO
hrp1-7	-0.8250982	0.061084155	0.000174	HRP1	NO
YOR327C	-0.8205854	0.095067788	0.000991	SNC2	NO
mtw1-ts	-0.8177015	0.077029571	0.000446	MTW1	NO
YAL011W	-0.8106727	0.043786405	5.01E-05	SWC3	NO
mob2-40	-0.8049846	0.11277175	0.002037	MOB2	YES
YKL074C	-0.790008	0.053314912	0.000121	MUD2	NO
YEL061C	-0.7879271	0.104125223	0.001635	CIN8	NO
YLR268W	-0.7872938	0.063850014	0.000249	SEC22	NO
cdc34-2	-0.7683087	0.122165598	0.003265	CDC34	NO
pan1-4	-0.7660759	0.04190564	5.27E-05	PAN1	NO
mcd1-73	-0.7615815	0.084468455	0.000838	MCD1	NO
sqt1-201	-0.7508929	0.088844845	0.001074	SQT1	NO
YDR017C	-0.7482997	0.043303375	6.58E-05	KCS1	NO
YJL095W	-0.7441054	0.053532022	0.000155	BCK1	NO

sgv1-23	-0.7415739	0.099843516	0.001754	SGV1	NO
YOR058C	-0.7273548	0.073258011	0.000578	ASE1	NO
sft1-15	-0.7230674	0.073416341	0.000596	SFT1	YES
YKL048C	-0.7159452	0.034656662	3.24E-05	ELM1	NO
abf1-102	-0.7013729	0.052616806	0.000183	ABF1	NO
dsn1-7	-0.6987609	0.101882713	0.002366	DSN1	NO
YNL249C	-0.6963337	0.067912115	0.00051	MPA43	NO
YDR273W	-0.6897353	0.059210478	0.00031	DON1	NO
spc24 4-2	-0.6809539	0.129059313	0.006186	SPC24	NO
YIL062C	-0.6783126	0.052679611	0.00021	ARC15	NO
YOR216C	-0.6758798	0.037351815	5.48E-05	RUD3	NO
taf1-1	-0.6752385	0.094291601	0.002013	TAF1	NO
YIR005W	-0.6722183	0.026807248	1.50E-05	IST3	NO
YPL055C	-0.6585474	0.047672553	0.000159	LGE1	NO
pkc1-2	-0.6517669	0.116194117	0.004962	PKC1	NO
cdc10-1	-0.6470596	0.096939977	0.002618	CDC10	NO
YOR027W	-0.6463279	0.042100485	0.000105	STI1	NO
YEL062W	-0.6453486	0.098483623	0.002804	NPR2	NO
YLR337C	-0.6448177	0.031946728	3.56E-05	VRP1	NO
YLL038C	-0.6429763	0.024482106	1.25E-05	ENT4	NO
taf10-ts34	-0.6399109	0.084938945	0.001662	TAF10	NO
ask1-2	-0.6377868	0.078117857	0.001225	ASK1	NO
taf8-7	-0.6334957	0.106208175	0.003967	TAF8	NO
YNL286W	-0.6330967	0.039619541	8.97E-05	CUS2	NO
ssl2-ts	-0.632111	0.075451259	0.00111	SSL2	NO
med7-141	-0.6315624	0.085488501	0.00179	MED7	NO
YPL079W	-0.628038	0.062551537	0.000553	RPL21B	YES
act1-108	-0.6242737	0.083318437	0.001697	ACT	NO
YPL106C	-0.6238352	0.067914326	0.00078	SSE1	NO

cog2-1	-0.6216964	0.113152089	0.005348	COG2	NO
sln1-ts4	-0.6092648	0.058406482	0.000477	SLN1	NO
tub4-?DSY	-0.5964548	0.097126162	0.003565	TUB4	NO
nnf1-48	-0.593916	0.088520329	0.002569	NNF1	NO
YPR095C	-0.5915616	0.036111454	8.13E-05	SYT1	NO
act1-3	-0.5888086	0.081868076	0.00198	ACT	YES
YJR084W	-0.5852585	0.057879477	0.000538	CSN12	NO
YNL069C	-0.5849932	0.069990349	0.00112	RPL16B	YES
YGR080W	-0.5840037	0.053617427	0.000403	TWF1	NO
ret3-1	-0.5828721	0.062480074	0.000735	RET3	NO
ssu72-2	-0.5728064	0.085599941	0.002594	SSU72	NO
YNL097C	-0.570089	0.040430299	0.000147	PHO23	NO
YNL266W	-0.5662038	0.065626434	0.000992	YNL266W	NO
ask1-3	-0.5645261	0.080030672	0.00213	ASK1	NO
YBR036C	-0.5613982	0.063563395	0.000907	CSG2	NO
YBR189W	-0.5568713	0.035322144	9.46E-05	RPS9B	YES
mob2-14	-0.553198	0.120041902	0.009968	MOB2	YES
duo1-2	-0.5530201	0.08817593	0.003299	DUO1	NO
sec39-1	-0.5500379	0.077718773	0.002104	SEC39	NO
rna15-58	-0.5489952	0.08522245	0.002988	RNA15	NO
spc105-4	-0.543954	0.074870572	0.001906	SPC105	NO
erg11-td	-0.5398286	0.08833748	0.00363	ERG11	NO
YFL018C	-0.5375316	0.048090852	0.000365	LPD1	NO
YKL114C	-0.537384	0.046273949	0.000314	APN1	NO
act1-119	-0.5359062	0.08401948	0.0031	ACT1	YES
YDL043C	-0.5356472	0.024175227	2.46E-05	PRP11	NO
YJL183W	-0.5334958	0.04185748	0.000218	MNN11	NO
dbf2-3	-0.5318166	0.0821964	0.00294	DBF2	NO
YJL168C	-0.5314176	0.030455169	6.33E-05	SET2	NO

YLR413W	-0.5207211	0.059106122	0.000916	YLR413W	NO
sgv1-35	-0.519888	0.087838958	0.004082	SGV1	NO
act1-136	-0.5191726	0.062153339	0.001123	ACT1	YES
dsn1-8	-0.5144537	0.076703024	0.002572	DSN1	NO
YKR024C	-0.5112959	0.047218535	0.000413	DBP7	NO
YGL192W	-0.5060326	0.05742322	0.000915	IME4	NO
YLR211C	-0.5052152	0.037643647	0.000178	YLR211C	YES
abf1-101	-0.5026232	0.068325493	0.001819	ABF1	NO
YER113C	-0.5021809	0.065052399	0.001516	TMN3	NO

Appendix 2. Gene Ontology terms enriched among *hsh155-1* negative interacting partners.

GO Biological Process						
Gene Ontology term	Percent cluster	Percent Genome	FOLD ENRICHMENT	Corrected P-value	FDR	FALSE Positives
mitotic sister chromatid biorientation	3.90%	0.20%	19.50	0.00865	0.14%	0.08
histone exchange	4.90%	0.30%	16.33	0.00289	0.13%	0.06
attachment of spindle microtubules to kinetochore	5.80%	0.40%	14.50	0.00358	0.16%	0.08
microtubule polymerization	5.80%	0.40%	14.50	0.00358	0.16%	0.08
spliceosomal complex assembly	6.80%	0.50%	13.60	0.00028	0.08%	0.02
microtubule polymerization or depolymerization	6.80%	0.50%	13.60	0.00035	0.07%	0.02
retrograde vesicle-mediated transport, Golgi to ER	5.80%	0.50%	11.60	0.00433	0.16%	0.08
actin polymerization or depolymerization	5.80%	0.50%	11.60	0.00433	0.15%	0.08
positive regulation of protein polymerization	5.80%	0.50%	11.60	0.00433	0.15%	0.08
positive regulation of supramolecular fiber organization	6.80%	0.60%	11.33	0.00109	0.05%	0.02
regulation of protein polymerization	6.80%	0.60%	11.33	0.00153	0.14%	0.06
regulation of cellular localization	6.80%	0.60%	11.33	0.0018	0.14%	0.06
protein polymerization	10.70%	1.00%	10.70	2.02E-06	0.00%	0
positive regulation of cytoskeleton organization	7.80%	0.80%	9.75	0.00067	0.05%	0.02
supramolecular fiber organization	14.60%	1.60%	9.13	3.71E-08	0.00%	0
regulation of supramolecular fiber organization	6.80%	0.80%	8.50	0.00946	0.14%	0.08
positive regulation of cellular component biogenesis	7.80%	1.00%	7.80	0.00742	0.14%	0.08
ER to Golgi vesicle-mediated transport	9.70%	1.30%	7.46	0.00062	0.06%	0.02
regulation of protein complex assembly	9.70%	1.30%	7.46	0.00062	0.06%	0.02
positive regulation of organelle organization	10.70%	1.50%	7.13	0.00021	0.00%	0
actin filament organization	7.80%	1.10%	7.09	0.00996	0.13%	0.08
mRNA splicing, via spliceosome	11.70%	1.70%	6.88	7.58E-05	0.00%	0
RNA splicing, via transesterification reactions with bulged	11.70%	1.70%	6.88	8.32E-05	0.00%	0

adenosine as nucleophile						
RNA splicing, via transesterification reactions	11.70%	1.80%	6.50	0.00021	0.00%	0
microtubule cytoskeleton organization	9.70%	1.50%	6.47	0.00204	0.13%	0.06
cytokinesis	8.70%	1.40%	6.21	0.0098	0.13%	0.08
Golgi vesicle transport	16.50%	2.80%	5.89	2.03E-06	0.00%	0
cell division	25.20%	4.30%	5.86	4.86E-11	0.00%	0
regulation of cellular component biogenesis	12.60%	2.20%	5.73	0.00024	0.08%	0.02
RNA splicing	12.60%	2.20%	5.73	0.00029	0.07%	0.02
microtubule-based process	9.70%	1.70%	5.71	0.00617	0.15%	0.08
positive regulation of cellular component organization	11.70%	2.10%	5.57	0.00088	0.05%	0.02
mRNA processing	16.50%	3.00%	5.50	6.09E-06	0.00%	0
cytoskeleton organization	19.40%	3.60%	5.39	3.44E-07	0.00%	0
covalent chromatin modification	11.70%	2.60%	4.50	0.00849	0.14%	0.08
macromolecular complex subunit organization	41.70%	10.00%	4.17	1.30E-14	0.00%	0
cellular protein complex assembly	15.50%	3.80%	4.08	0.00111	0.05%	0.02
mRNA metabolic process	18.40%	4.60%	4.00	0.00012	0.00%	0
protein complex subunit organization	20.40%	5.20%	3.92	3.71E-05	0.00%	0
protein complex assembly	17.50%	4.50%	3.89	0.00039	0.07%	0.02
macromolecular complex assembly	32.00%	8.40%	3.81	3.78E-09	0.00%	0
protein complex biogenesis	17.50%	4.60%	3.80	0.00064	0.06%	0.02
cellular macromolecular complex assembly	30.10%	8.00%	3.76	2.32E-08	0.00%	0
regulation of organelle organization	16.50%	4.50%	3.67	0.00194	0.13%	0.06
vesicle-mediated transport	21.40%	5.90%	3.63	6.94E-05	0.00%	0
mitotic cell cycle process	18.40%	5.10%	3.61	0.00057	0.06%	0.02
mitotic cell cycle	19.40%	5.60%	3.46	0.00055	0.07%	0.02
regulation of cellular component organization	20.40%	6.20%	3.29	0.00059	0.06%	0.02
cellular component assembly	38.80%	11.90%	3.26	1.04E-09	0.00%	0
cell cycle	32.00%	11.20%	2.86	7.48E-06	0.00%	0
chromosome organization	23.30%	8.60%	2.71	0.00344	0.17%	0.08

protein localization	31.10%	11.70%	2.66	6.85E-05	0.00%	0
cell cycle process	23.30%	8.90%	2.62	0.00613	0.15%	0.08
cellular localization	33.00%	12.70%	2.60	3.86E-05	0.00%	0
organelle organization	50.50%	20.60%	2.45	7.98E-09	0.00%	0
macromolecule localization	31.10%	13.00%	2.39	0.00079	0.05%	0.02
cellular component organization	66.00%	28.30%	2.33	9.82E-13	0.00%	0
cellular component biogenesis	40.80%	18.10%	2.25	3.89E-05	0.00%	0
cellular component organization or biogenesis	68.00%	33.60%	2.02	5.84E-10	0.00%	0
localization	42.70%	21.90%	1.95	0.00107	0.05%	0.02
cellular process	91.30%	71.70%	1.27	0.00061	0.06%	0.02

GO Molecular Function

Gene Ontology term	Percent Cluster	Percent Genome	FOLD ENRICHMENT	Corrected P-value	FDR	FALSE Positives
cytoskeletal protein binding	10.70%	1.40%	7.64	2.57E-05	0.00%	0
protein binding	31.10%	12.30%	2.53	4.06E-05	0.00%	0
SNAP receptor activity	4.90%	0.30%	16.33	0.00234	0.00%	0
structural constituent of cytoskeleton	4.90%	0.40%	12.25	0.00617	0.50%	0.02
actin binding	5.80%	0.70%	8.29	0.00794	1.60%	0.08
binding	66.00%	47.10%	1.40	0.00944	1.33%	0.08

GO Cellular Component

Gene Ontology term	Percent cluster	Percent Genome	FOLD ENRICHMENT	Corrected P-value	FDR	FALSE Positives
MIS12/MIND type complex	2.91%	0.07%	41.74	0.00624	0.00%	0
nuclear MIS12/MIND complex	2.91%	0.07%	41.74	0.00624	0.00%	0
Swr1 complex	4.85%	0.20%	24.85	0.00022	0.00%	0

INO80-type complex	4.85%	0.35%	13.91	0.00519	0.00%	0
transport vesicle membrane	4.85%	0.38%	12.88	0.00772	0.00%	0
SNARE complex	4.85%	0.39%	12.42	0.00929	0.00%	0
Golgi-associated vesicle membrane	5.83%	0.47%	12.28	0.00163	0.00%	0
condensed nuclear chromosome kinetochore	8.74%	0.78%	11.18	1.75E-05	0.00%	0
condensed chromosome kinetochore	9.71%	0.93%	10.38	6.71E-06	0.00%	0
condensed nuclear chromosome, centromeric region	8.74%	0.87%	10.10	4.38E-05	0.00%	0
spliceosomal snRNP complex	8.74%	0.88%	9.94	5.05E-05	0.00%	0
coated vesicle membrane	5.83%	0.59%	9.94	0.0058	0.00%	0
condensed chromosome, centromeric region	9.71%	1.02%	9.53	1.56E-05	0.00%	0
small nuclear ribonucleoprotein complex	8.74%	0.92%	9.49	7.62E-05	0.00%	0
actin cortical patch	7.77%	0.82%	9.43	0.00037	0.00%	0
kinetochore	9.71%	1.05%	9.28	2.03E-05	0.00%	0
endocytic patch	7.77%	0.84%	9.28	0.00042	0.00%	0
Sm-like protein family complex	8.74%	0.95%	9.21	9.91E-05	0.00%	0
spindle	8.74%	0.98%	8.95	0.00012	0.00%	0
supramolecular complex	9.71%	1.09%	8.92	2.98E-05	0.00%	0
supramolecular polymer	9.71%	1.09%	8.92	2.98E-05	0.00%	0
supramolecular fiber	9.71%	1.09%	8.92	2.98E-05	0.00%	0
polymeric cytoskeletal fiber	9.71%	1.09%	8.92	2.98E-05	0.00%	0
vesicle membrane	6.80%	0.80%	8.54	0.00334	0.00%	0
cytoplasmic vesicle membrane	6.80%	0.80%	8.54	0.00334	0.00%	0
microtubule	7.77%	0.93%	8.31	0.00099	0.00%	0
U2-type spliceosomal complex	6.80%	0.82%	8.25	0.00421	0.00%	0
chromosome, centromeric region	11.65%	1.44%	8.11	4.63E-06	0.00%	0
cortical actin cytoskeleton	7.77%	1.00%	7.73	0.00171	0.00%	0
spliceosomal complex	9.71%	1.30%	7.48	0.00016	0.00%	0
Golgi-associated vesicle	7.77%	1.05%	7.42	0.00233	0.00%	0
cortical cytoskeleton	7.77%	1.05%	7.42	0.00233	0.00%	0

condensed nuclear chromosome	10.68%	1.55%	6.89	9.99E-05	0.00%	0
cytoskeletal part	22.33%	3.27%	6.84	2.64E-11	0.00%	0
cytoskeleton	23.30%	3.45%	6.76	9.15E-12	0.00%	0
actin cytoskeleton	7.77%	1.16%	6.71	0.00495	0.00%	0
microtubule cytoskeleton	12.62%	1.88%	6.70	1.17E-05	0.00%	0
condensed chromosome	11.65%	1.80%	6.47	5.82E-05	0.00%	0
Golgi membrane	12.62%	2.08%	6.07	3.80E-05	0.00%	0
Golgi subcompartment	13.59%	2.55%	5.32	6.36E-05	0.00%	0
cell cortex part	9.71%	1.95%	4.97	0.00651	0.00%	0
Golgi apparatus part	13.59%	2.89%	4.71	0.00028	0.00%	0
chromosomal region	11.65%	2.90%	4.01	0.00862	0.00%	0
nuclear chromosome part	17.48%	4.61%	3.79	0.00018	0.00%	0
Golgi apparatus	14.56%	3.92%	3.71	0.0022	0.00%	0
nuclear chromosome	17.48%	4.94%	3.54	0.00051	0.00%	0
chromosomal part	19.42%	6.45%	3.01	0.00154	0.00%	0
chromosome	20.39%	6.92%	2.95	0.00124	0.00%	0
protein complex	41.75%	16.26%	2.57	1.17E-07	0.00%	0
nuclear part	44.66%	18.62%	2.40	2.00E-07	0.00%	0
nuclear lumen	30.10%	13.40%	2.25	0.00151	0.00%	0
macromolecular complex	66.02%	32.24%	2.05	3.40E-10	0.00%	0
non-membrane-bounded organelle	42.72%	21.45%	1.99	0.00018	0.00%	0
intracellular non-membrane-bounded organelle	42.72%	21.45%	1.99	0.00018	0.00%	0
intracellular organelle part	78.64%	43.61%	1.80	6.25E-11	0.00%	0
organelle part	78.64%	43.68%	1.80	6.94E-11	0.00%	0
nucleus	58.25%	33.98%	1.71	7.30E-05	0.00%	0
intracellular organelle	90.29%	67.33%	1.34	8.11E-06	0.00%	0
organelle	90.29%	67.35%	1.34	8.23E-06	0.00%	0
intracellular membrane-bounded organelle	81.55%	61.89%	1.32	0.00253	0.00%	0
membrane-bounded organelle	81.55%	62.71%	1.30	0.00522	0.00%	0

intracellular	95.15%	79.39%	1.20	0.00118	0.00%	0
cell part	97.09%	81.55%	1.19	0.00034	0.00%	0
cell	97.09%	81.58%	1.19	0.00035	0.00%	0
intracellular part	94.17%	79.19%	1.19	0.00442	0.00%	0