Investigating the role of Drosha and RNA molecules in the repair of DNA double-strand breaks

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

Benjamin R. Hawley
MRC Toxicology Unit
University of Leicester

2018

Abstract - Investigating the role of Drosha and RNA molecules in the repair of DNA double-strand breaks

Benjamin R Hawley, MRC Toxicology Unit, University of Leicester, Leicester, LE1 9HN, UK

The error-free and efficient repair of DNA double-strand breaks is extremely important for cell survival. Over the past several years, RNA has been increasingly implicated in the DNA damage response, but the mechanism or mechanisms through which it can act remain poorly understood. Various hypotheses have been made in the literature: RNA as a signalling molecule or as a scaffold; RNA long-range interactions to reorganise or maintain chromatin structure; RNA as a template for repair.

In this thesis, the miRNA biogenesis apparatus was demonstrated to be involved in DNA repair outside of their normal roles in miRNA maturation. Drosha was required for efficient repair by the two major repair pathways, suggesting a central role in the DNA damage response. Previous reports looking at the involvement of Drosha in DNA damage had suggested that Drosha processes newly transcribed RNA, the product of which would carry out a role in repair. Here, a comprehensive series of next-generation sequencing approaches was unable to validate this.

Later experiments instead discovered that pre-existing RNA molecules could participate in repair. RNA invasion around DNA break sites was observed, and shown to be Drosha-dependent. These damage-induced DNA:RNA hybrid structures were important for efficient repair, showing that RNA can be a direct and critical mediator of DNA damage repair.

Acknowledgements

I would like to express my sincere gratitude to Professor Martin Bushell for his excellent supervision and support throughout the course of my PhD. His enthusiasm for science and constant encouragement have kept me inspired throughout this project, and to pursue an academic career now that it is over.

I would also like to thank the entirety of the Bushell lab, who have made this an enjoyable and incredibly rewarding four years. The lab meetings may go on for a bit too long last thing on a Friday, but the great discussions and bizarre tangents make it all worthwhile and have made me a better scientist. In particular, I would like to thank Dr Wei-Ting Lu for his mentoring for the duration of this project, Dr Ania Wilczynska for her invaluable advice regarding just about everything, and Dr Ewan Smith for his excellent counsel and proofreading of this thesis.

As always, a special thanks goes to my family and friends.

Contents

Table of figures

Table of tables

Abbreviations

53BP1	Tumour Suppressor p53-Binding Protein 1
Ago	Argonaute
AP site	Abasic Site
aRNA	Aberrant RNA
ATM	Ataxia-Telangiectasia Mutated
ATR	Ataxia- And Rad3-Related
ATRIP	ATR-Interacting Protein
BER	Break Excision Repair
BLM	Bloom Syndrome Protein
BRCA1/2	Breast Cancer 1/2
BRE	TFIIB Recognition Element
CDK9	Cyclin-Dependent Kinase 9
CETN2	Centrin-2
CFI/II	Cleavage Factors I/II
ChIP	Chromatin IP
CHK2	Checkpoint Kinase 2
CO/NCO	Crossover/Non-Crossover Event
CPA factors	Cleavage and Polyadenylation Factors
CPSF	Cleavage and Polyadenylation Specificity Factors
CSA/B	Cockayne Syndrome A/B
CstF	Cleavage Stimulatory Factors
CTD	C-Terminal Domain of RNA Pol II
CtIP	C-Terminal-Binding Protein 1-Interacting Protein
DCL	Dicer-Like Protein
DDB2	DNA Damage-Binding Protein 2
DDR	DNA Damage Response
DGCR8	DiGeorge Syndrome Critical Region 8
dHJ	Double Holliday Junction
DHX9	DEAH-Box Helicase 9
D-loop	Displacement Loop

DNA2	DNA Replication ATP-Dependent Helicase/Nuclease
DNA-PKcs	DNA-Dependent Protein Kinase Catalytic Subunit
DSB	Double Strand Break
DSIF	5,6-Dichloro-1-B-D-Ribofuranosylbenzimidazole-Sensitivity Inducing Factor
dsRBD	dsRNA Binding Domain
DSS1	Deleted in Spilt Hand/Spilt Foot
ERCC1/XPF	DNA Excision Repair Protein 1 & 4 Complex
EXO1	Exonuclease 1
GG-/TC-NER	Global Genomic/Transcription-Coupled NER
GTase	RNA Guanylyltransferase
Hi-C	High-Throughput Chromosomal Conformation Capture
HR	Homologous Recombination
IF	Immunofluorescence
indel	Insertion or Deletion (of Genetic Material)
IP	Immunoprecipitation
IR	Ionising Radiation
LIG4	DNA Ligase IV
lncRNA	Long Non-Coding RNA
MDC1	Mediator of DNA Damage Checkpoint 1
miRNA	MicroRNA
MMEJ	Microhomology-Mediated End Joining
MMR	Mismatch Repair
MRN	MRE11-RAD50-Nbs1 Complex
MutSα/β	Mutator S Protein
NELF	Negative Elongation Factor
NER	Nucleotide Excision Repair
NGS	Next-Generation Sequencing
NHEJ	Non-Homologous End Joining
NuRD	Nucleosome Remodelling Deacetylase
PABP	Poly(A) Binding Protein

PALB2	Partner and Localiser of BRCA2
PAP	Poly(A) Polymerase
PARP	Poly(ADP) Ribose Polymerase
PAS	Polyadenylation Signal
PASR	Promoter-Associated RNA
PIC	Pre-Initiation Complex
PIKK	Phosphatidylinositol 3-Kinase-Related Kinases
piRNA	Piwi-Interacting RNA
Pol	Polymerase (DNA Are Greek Letters, RNA Are Roman Numerals)
pre-miRNA	Precursor miRNA
pri-miRNA	Primary miRNA
P-TEFb	Positive Transcription Elongation Factor B
PTRF	Pol I Transcript Release Factor
RdDP	RNA-Dependent DNA Polymerase
RdRP	RNA-Dependent RNA Polymerase
RIIID	RNase III Domain
RISC	RNA-Induced Silencing Complex
R-loop	DNA:RNA Hybrid Displacement Loop
RNMT	Guanine-N7 Methyltransferase
RPA	Replication Protein A
RT	Reverse Transcriptase
RTEL1	Regulator of Telomere Elongation 1
SCIDA	Severe Combined Immunodeficiency, Athabaskan-Type
SDSA	Synthesis-Dependant Strand Annealing
SETX	Senataxin
sn/snoRNA	Small Nuclear/Nucleolar RNA
sncRNA	Small Non-Coding RNA
ss/dsDNA	Single/Double Stranded DNA
SSA	Single-Strand Annealing
SSB	Single Strand Break
SSBR	SSB Repair

SWI/SNF	Switch/Sucrose Non-Fermentable
TAF	TBP-Associated Factors
TBP	TATA Box-Binding Protein
TES/TTS	Transcription End/Termination Site
TFIIA-H	TF for Pol II
TFIIA-H	Transcription Factor
TNRC6	Trinucleotide Repeat-Containing Gene 6
TPase	RNA Triphosphatase
TSS	Transcription Start Site
TSSa-RNA	TSS-Associated small RNA
TTFI	Transcription Termination Factor I
UV	Ultraviolet Radiation
UVSSA	UV-Stimulated Scaffold Protein A
X4-L4	XRCC4-LIG4 Complex
XLF	XRCC4-Like Factor
XPA-G	Xeroderma Pigmentosum Complentation Group A-G
XPO5	Exportin-5
XRCC	X-Ray Repair Cross-Complementing Protein
XRN2	5’-3’ Exoribonuclease 2
γH2AX	H2A Histone Family, Member X (Ser139 P)

1 Chapter 1: Introduction

1.1 Prelude

Traditionally, DNA damage repair has been considered a purely protein-based process. The actions of kinases, histones, ubiquitin transferases, scaffold proteins, nucleases and polymerases work in concert to identify a break, initiate signalling cascades, modify and open the chromatin, recruit various competing factors and ultimately resolve the damage with as little error as possible. Known as the DNA damage response (DDR), this is reviewed in section 1.2.

Most studies into DNA repair have consequently ignored the most active nuclear process – transcription (section 1.3). RNA has been implicated in most other nuclear pathways: transcriptional regulation, DNA replication, and chromatin structure; however, the involvement of what is ultimately additional copies of the transcribed genome in DNA repair has been unexplored.

Many recent pieces of evidence have pointed towards an important role for RNA in DNA repair (discussed in section 1.2). The explosion of next-generation sequencing approaches has shown that transcription is remarkably pervasive, with 80-90% of the genome being actively transcribed (Wong, Passey, and Yu 2001; Clark et al. 2011; Djebali et al. 2012; Hangauer, Vaughn, and McManus 2013). Recent high-throughput proteomic screens have consistently identified RNA binding proteins as being one of the most significantly enriched groups of proteins involved in DNA repair (section 1.5). The DNA repair pathways undertaken depends on how transcriptionally active the damaged locus is, with genes undertaking the most energy-intensive but error-free pathway (section 1.5.1). Finally, compelling but preliminary data suggests that RNA may be a template for DNA repair (section 1.5.4), a concept that would spawn an entire new area of DNA repair research.

The aim of this PhD project was to investigate how RNA participates in DNA repair (see section 1.6 for detailed aims). A major focus was the involvement of the microRNA biogenesis components (reviewed in section 1.4.2), and the utilisation of next generation sequencing (NGS) approaches.

1.2 The DNA damage response

1.2.1 Induction of DNA damage

Our genomes are constantly barraged by an array of damaging agents, which come from both endogenous and exogenous sources. For example, intrinsic factors such as radical oxygen species (ROS) from metabolic processes can damage and oxidise DNA bases (Jena 2012). DNA replication can be erroneous, leading to misincorporation of nucleotides; additionally, the DNA replication machinery and the transcription machinery can collide, causing further instability (Price, Alm, and Arkin 2005; N. Kim and Jinks-Robertson 2012). Whilst these intrinsic forms of damage are very uncommon, the sheer size of human genomes means each individual cell can accumulate thousands of damage events every day: when considering external sources this number is estimated to be as high as 10^5 per day (Hoeijmakers 2009). Exogenous sources are the most recognisable DNA damaging agents. Ultraviolet (UV) radiation from sunlight can cause DNA base crosslinking or ROS: keratinocytes, being highly exposed to UV, are subject to an additional 10^5 of these DNA photoproducts daily (see section 1.2.8.3) (Hoeijmakers 2009). Ionizing radiation (IR) from cosmic rays and X-rays can cause single strand breaks (SSBs) and double strand breaks (DSBs). Finally, chemicals cause a wide variety of damage, which is the consequence of certain chemotherapeutics and smoking, and a side effect from the metabolism of other drugs.

The vast majority of these genotoxic insults are recognised rapidly by the cell and repaired with no loss of genetic information. However, those that are repaired erroneously result in mutations, insertions and deletions (indels), and chromosomal translocations. Inevitably, these errors lead to loss or gain of gene function, resulting in cell senescence and aging, cell death, or cellular transformation and oncogenesis. Research into all aspects of the DDR is therefore fundamental to understanding of disease progression and the development of new treatments.

1.2.2 Initial detection of a DSB

The recruitment of either the MRE11-RAD50-Nbs1 (MRN) complex or Ku to DNA termini is typically referred to as the first step in DSB detection and DDR signalling, initiating repair by homologous recombination (HR; section 1.2.6) or non-homologous end joining (NHEJ; section 1.2.5) respectively. Ku has high affinity for DSB ends and

protects them from MRN; when Ku is outcompeted by MRN this initiates resection and HR repair. However, the role of MRN as the primary damage sensor for resection-dependent pathways has been challenged by observations that poly(ADP-ribose) polymerase 1 (PARP1) binds DNA ends first, and is required for MRN recruitment (Haince et al. 2008; Ali et al. 2012). Like Ku, PARP1 has very high affinity for DNA ends, suggesting that these proteins (in high abundances in the nucleus) constantly probe the genome for damage. Interestingly, an additional model where recruitment of these factors is preceded by the sudden relaxation of DNA supercoiling and thus the alteration in chromatin structure due to the DSB has been proposed in reviews, but is dependent on an as-of-yet unidentified sensor (Thompson 2012). Regardless of this, all events lead to the initiation of the DNA damage response (DDR), the first step of which is propagation of the damage signal and chromatin remodelling.

1.2.3 Signal propagation

Following sensing of a break, the key event in the DDR is the initiation of signalling, orchestrated by one of the three DNA damage PIKKs; ATM, ATR, and DNA-PKcs. ATM is a master regulator of DSB DDR signalling, where it initiates cell cycle arrest through checkpoint kinase 2 (CHK2), DNA repair through histone H2AX, and potentially apoptosis through tumour suppressor p53 (Figure 1.1AB) (Paull 2015). ATR and DNA-PKcs possess a degree of redundancy for certain ATM targets, however they are typically activated by different forms of DNA damage or to promote certain repair pathways (Matsuoka et al. 2007). For example, DNA-PKcs, in concert with Ku, promote NHEJ whereas ATR would promote HR; ATR is also activated by single-stranded DNA (ssDNA) and so also promotes nucleotide excision repair (NER), the pathway that repairs UV-induced DNA adducts (section 1.2.8.3) (Caron et al. 2015).

In a normal situation, ATM exists as an inactive dimer. Upon DNA damage, autophosphorylation of ATM occurs *in trans* on the serine 1981 residue and the dimer dissociates (Bakkenist and Kastan 2003). Active pATM monomer is then recruited to DSBs by mediator of DNA damage checkpoint 1 (MDC1) and the MRN complex (Lou et al. 2006). At the DSB, pATM can phosphorylate H2AX on serine 139, known as γH2AX. This is found at all DSBs and can spread over megabases of chromatin (Iacovoni et al. 2010) as part of a positive feedback loop: MDC1 binds γH2AX, which in turn recruits more MRN and pATM to propagate this signal (Figure 1.1A).

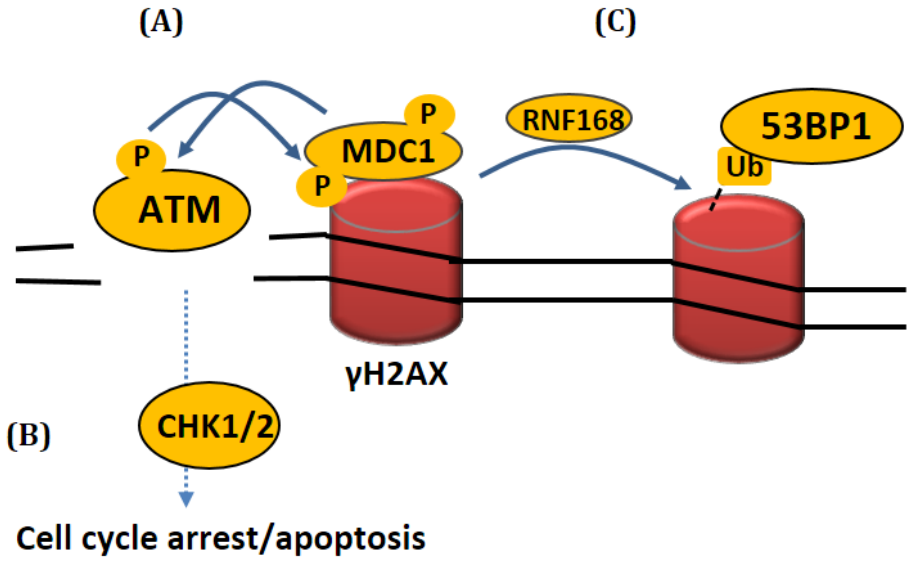


Figure 1.1. Early DDR and chromatin remodelling
(A) On damage induction, inactive ATM dimers undergo autophosphorylation and split into active pATM monomers. A key target for pATM is the phosphorylation of histone H2AX (γH2AX) which is present at every DSB. γH2AX recruits MDC1, which in turn recruits more pATM to spread the γH2AX signal along the chromatin. **(B)** Other pATM targets include the CHK1 and CHK2 kinases, which then initiate cell cycle arrest. Prolonged damage can also induce programmed cell death by apoptosis. **(C)** γH2AX is a platform for other downstream DDR factors. The E3 ubiquitin ligase RNF168 recognises this modification, and in turn ubiquitinates H2A and H2AX. These marks, as well as methylation and acetylation, promote chromatin unwinding and allow the recruitment of key DDR factors such as 53BP1.

1.2.4 Chromatin remodelling

Following detection of the break and the initiation of the DDR, the chromatin around DSBs is relaxed and heavily modified to allow access of the many repair factors to the sites of damage and to provide a platform for their recruitment. As mentioned above, the γH2AX signal can extend far along the chromatin, and in turn recruits more DDR factors (Figure 1.1A). Acetylation on histones H2 and H4 similarly spreads long distances along the chromatin (Downs et al. 2004); this is catalysed by the acetyltransferase Tip60 and is associated with opening of the chromatin structure (Murr et al. 2006). Other modifications such as methylation can stimulate the recruitment of factors that promote certain pathways other others. For example, di-methylated histone H4 on lysine 20 (H4K20me2) was shown to be required for tumour suppressor p53 binding protein 1 (53BP1) to DSBs to promote NHEJ repair (Tuzon et al. 2014).

Another critical modification is the ubiquitination (ub) of various nucleosomes. H2AK15ub was demonstrated to work in concert with H4K20me2 to recruit 53BP1 (Fradet-Turcotte et al. 2013). γH2AX recruits two E3 ubiquitin ligases, RING finger 8 (RNF8) and RNF168, which appear to act sequentially to add polyubiquitin chains onto H2A (and H2AX itself) to promote DDR signalling (Figure 1.1C) (Mattiroli et al. 2012; Fradet-Turcotte et al. 2013; Bohgaki et al. 2013). Thus, without chromatin modification and relaxation, DDR does not proceed efficiently beyond the earliest stages; indeed, loss of functional RNF168 and thus downstream DDR is associated with an immunodeficiency and radiosensitivity syndrome (Stewart et al. 2009). Once adequate modification has been achieved, the DDR can fully progress by committing to one of several repair pathways, the major two being NHEJ and HR.

1.2.5 Non-homologous end joining

Non-homologous end joining (NHEJ) is the dominant repair pathway in the cell, occurring at all stages of the cell cycle. NHEJ is the simple and direct ligation of DNA ends, which provides a quick resolution to a DSB but one that potentially introduces errors (Figure 1.2). For a cell to commit to NHEJ, the DSB must have had minimal processing. To prevent the resection machinery (discussed in section 1.2.6) from digesting DNA, the Ku heterodimer binds to and protects the DNA ends. Ku consists of Ku70 and Ku80 subunits (classically XRCC6 and 5, respectively) which create a toroidal

structure that encircles dsDNA ends (Figure 1.2A) (Walker, Corpina, and Goldberg 2001). The recruitment of Ku occurs within seconds and is the primary event in NHEJ (Mari et al. 2006; Mahaney, Meek, and Lees-Miller 2009). Once assembled upon DNA ends, Ku recruits DNA-PKcs, the kinase responsible for downstream NHEJ signalling (Suwa et al. 1994; Lehman, Hoelz, and Turchi 2008). On DNA-PKcs binding, Ku translocates inwards along the DNA allowing DNA-PKcs to contact the DSB termini (Figure 1.2B) (Yoo and Dynan 1999). As the recruitment of these factors occurs on both ends at a DSB, the two DNA-PKcs interact forming a synaptic complex where the two ends are kept in contact and protected from any resection or additional DNA damage (DeFazio et al. 2002). In the simplest scenario, where the DNA ends are blunt ended or have compatible overhangs, with 5' phosphate and 3' hydroxyl groups, the ends are directly ligated. This is canonically catalysed by the ATP-dependent DNA ligase IV (LIG4), which is found in the X4-L4 complex alongside XRCC4 and XLF (Figure 1.2C) (Grawunder et al. 1997; Ma et al. 2004). XRCC4 is a homodimer that has no enzymatic activity and is instead an essential scaffold protein that aids in recruitment of LIG4 as well as other factors (Mari et al. 2006). It stimulates the enzymatic activity of DNA ligase IV, as does XLF. XLF is not a core component of X4-L4, but is required for some NHEJ-repaired breaks and is either recruited to DSBs by XRCC4 or retained by it (Roy et al. 2015); XLF similarly stimulates LIG4 activity, however towards the ligation of non-compatible ends.

Typically, a DSB is not a clean break and will require additional processing before X4-L4 can re-ligate the ends. These steps involve exo- and endonucleases, DNA polymerases, helicases, and DNA modifying enzymes and as such majorly contribute to the error-prone nature of NHEJ (Figure 1.2D). Many processing factors have been described as having a role in NHEJ, yet very few cause impairment to DSB repair when depleted suggesting there may be a large degree of redundancy. For example, Artemis is a 5'-3' exonuclease with additional endonuclease activity that can process end termini and is critical for V(D)J recombination; loss of functional Artemis leads to immunodeficiency but no appreciable DSB defects (Moshous et al. 2001).

DNA that has been processed by nucleases such as Artemis may require synthesis of new DNA to fill in the resulting gaps. The DNA polymerase X family, which include Pol μ and Pol λ, has been extensively documented to participate in NHEJ. Pol λ is

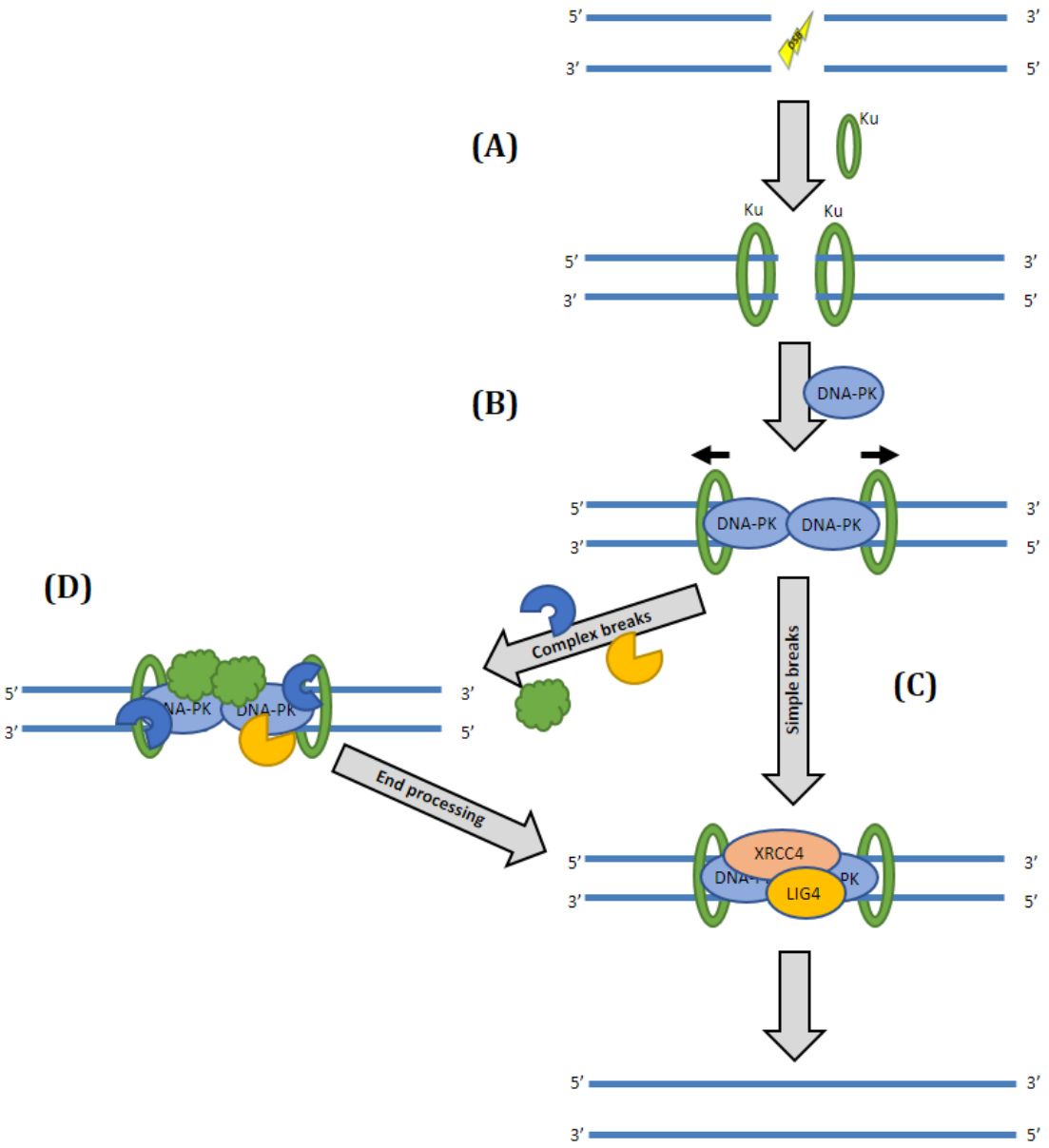


Figure 1.2. NHEJ repair pathway

(A) In G1-phase, DSB induction leads to the recruitment of the toroidal Ku heterodimer (Ku70-Ku80) to protect the DNA ends. **(B)** Stable Ku-DNA complexes recruit DNA-PK. This causes inwards translocation of Ku allowing DNA-PK to synapse the break. **(C)** Simple breaks (i.e. ligatable ends) are simply ligated together by the X4-L4 (XRCC4-DNA ligase IV) complex. **(D)** DSBs typically require processing. The recruitment and concerted actions of helicases, nucleases, polymerases, and DNA modification enzymes (indicated by an assortment of shapes) make the ends amenable to ligation and thus repair of the DSB.

template-dependent and as such requires processed DNA termini to be compatible overhangs; in contrast, Pol μ is template-independent and will synthesise new DNA on ends with little to no end complementarity (McElhinny et al. 2005). The DSB can then be resolved by the X4-L4 complex (Figure 1.2CD). The order of all these events is unclear, and may depend heavily on the nature of the break (Mahaney, Meek, and Lees-Miller 2009). The unexpected activity of LIG4 in annealing a single strand at a time for example could produce scenarios where an amenable strand is ligated easily, generating a complex SSB that is now more easily repaired in the absence of many competing DSB repair factors. This flexibility in recruitment is an important feature of NHEJ.

1.2.6 Homologous recombination

Homologous recombination (HR) is traditionally thought of as the error-free DSB repair pathway and is the next major pathway after NHEJ. HR requires several key steps, namely the commitment to DNA resection around the DSB and the invasion of the resulting ssDNA into the sister chromatid for templated repair. As such, HR only occurs in S/G2 phase when the sister chromatid is available.

1.2.6.1 Resection

In NHEJ, the DNA ends are protected from resection by Ku binding, which competes away the MRN complex (see section 1.2.5). In HR, the MRN complex binds to and synapses the DNA termini to initiate resection, as well as triggering the ATM-dependent signalling cascade as discussed previously in section 1.2.3 (Figure 1.3A). The MRE11 subunit of MRN is a 3'-5' ATP-independent exonuclease with ssDNA endonuclease activity (Trujillo et al. 1998). This activity of MRE11 is not required for HR however, as nuclease dead mutants still have adequate resection (Llorente and Symington 2004). The endonuclease CtIP (CTBP-interacting protein) has been shown to interact with MRN and facilitate initial resection for up to around 100 nt (Figure 1.3B) (Mehta and Haber 2014). CtIP associates with Breast Cancer 1 (BRCA1), both of which are phosphorylated by ATM to interact with MRN at DSBs (Cortez 1999; H. Wang et al. 2013). After the initial resection by MRN-BRCA1-CtIP, more extensive nucleolytic degradation is required to provide adequate substrates for the downstream homology search. This resection is carried out by the 5'-3' exonuclease EXO1 and the endonuclease DNA2, which resect at a rate of ~4 kb/hr (Figure 1.3C)

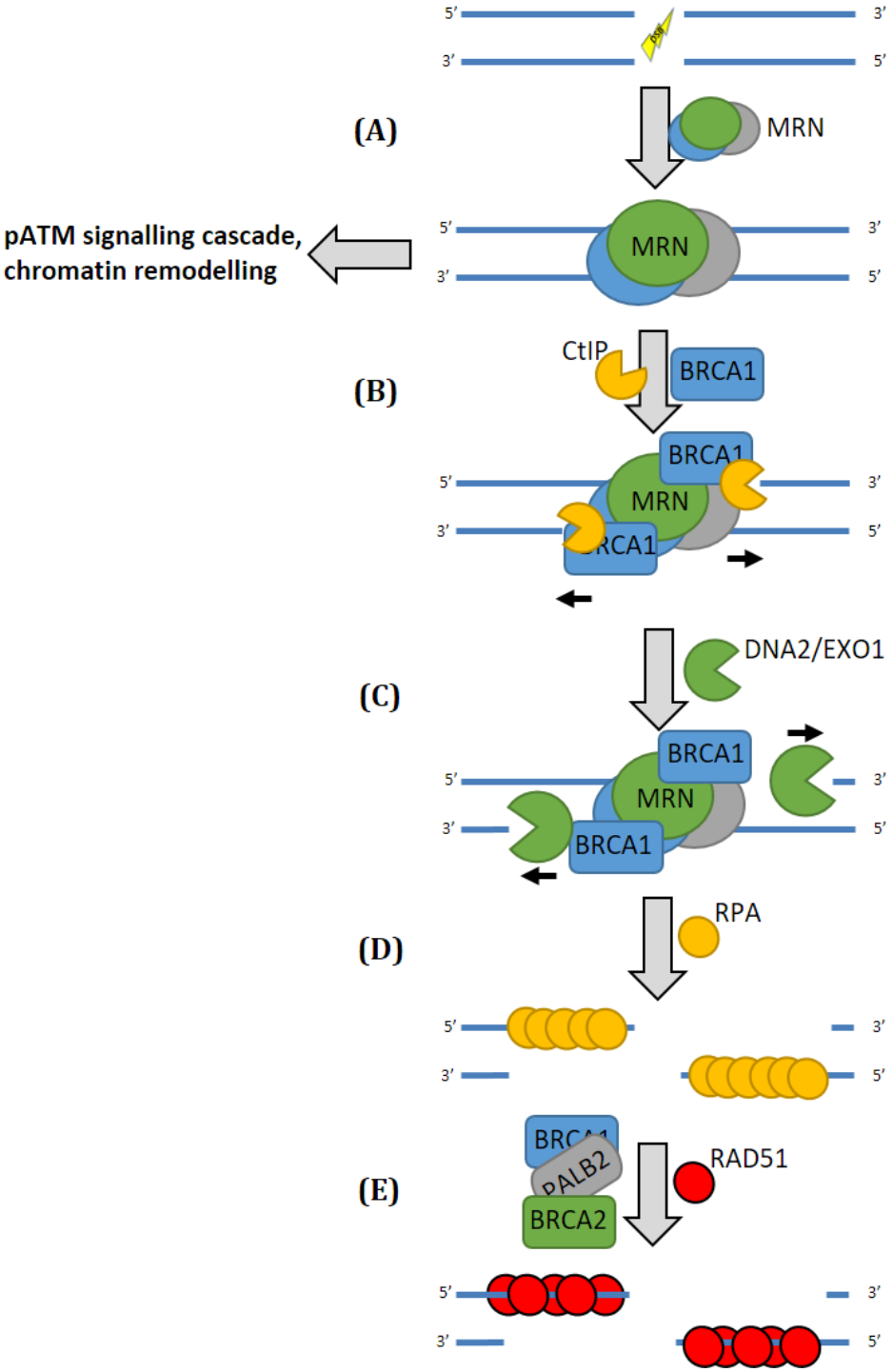


Figure 1.3. HR requires commitment to resection.

(A) MRN is recruited instead of Ku, synapsing DNA ends. This also recruits pATM to trigger DDR signalling. **(B)** MRN-BRCA1-CtIP initiates 5'-3' resection (~100 bp). **(C)** Extensive resection by DNA2 and EXO1 leads to large stretches of ssDNA. **(D)** ssDNA is bound by RPA to protect it and prevent secondary structure. **(E)** BRCA1-PALB2-BRCA2 loads RAD51 onto the ssDNA, replacing RPA to form the presynaptic nucleofilament.

(Mimitou and Symington 2008; Zhu et al. 2008; Mehta and Haber 2014). BRCA1 is a key mediator of HR, playing a central role in resection. Opposite BRCA1 is 53BP1, which prevents resection to promote NHEJ. As such, these two factors are considered antagonists, each competing with one another to promote their respective pathways (Bunting et al. 2010; Panier and Boulton 2013).

1.2.6.2 Assembly of the presynaptic filament

Following resection, the resulting 3' ssDNA is rapidly covered by Replication Protein A (RPA). RPA is a heterotrimer with very high sequence-nonspecific ssDNA affinity that prevents secondary structure from forming and protects the vulnerable ssDNA from further damage (Figure 1.3D) (Wold 1997; R. Chen and Wold 2014). For HR to proceed, RPA needs to be replaced with the recombinase RAD51. Oligomerisation of RAD51 along the DNA strand forms the presynaptic nucleoprotein filament, crucial for both invasion of the ssDNA into the sister chromatid and for annealing to regions of homology (Figure 1.3E). The displacement of RPA requires the activity of several mediators such as BRCA2. BRCA2 was demonstrated to bind to and facilitate RAD51 loading on RPA-bound ssDNA, and to then stabilise the nucleofilament (Carreira and Kowalczykowski 2011). How BRCA2 directs monomeric RAD51 to the DSB and physically exchanges with RPA was unknown until relatively recently. A DNA-binding domain in BRCA2 was identified which could also bind the protein DSS1 (H. Yang et al. 2002); however, loss of DNA-binding did not affect HR as long as PALB2 binding remained possible (Siaud et al. 2011). PALB2 can bind both BRCA1 and BRCA2 simultaneously, and thus directs RAD51-bound BRCA2 to BRCA1-bound ssDNA; loss of this causes severe HR defects (F. Zhang et al. 2009; Sy, Huen, and Chen 2009; Siaud et al. 2011). This suggests that the interaction with DSS1 is more important for BRCA2 function than DNA-binding. This is backed up by observations that DSS1 stimulates BRCA2-mediated RAD51 loading onto RPA-bound ssDNA (J. Liu et al. 2010). Excitingly, recent data showed that DSS1 can interact with RPA to promote its release from ssDNA (Zhao et al. 2015). This work demonstrated that DSS1 – as a small, highly acidic polypeptide – acts as a DNA mimic that can compete with ssDNA for RPA binding, allowing RAD51 to be exchanged upon the ssDNA. Taken together, all these data suggest a model where, after DNA resection mediated by BRCA1 and RPA loading,

PALB2 recruits BRCA2, DSS1 destabilises RPA, and RAD51 can then form the presynaptic nucleofilament, stabilised by BRCA2 (Figure 1.3).

1.2.6.3 Invasion into homologous donor

Once assembled, the presynaptic filament searches for and synapses with homologous duplex DNA. The homology search is the rate-limiting step of HR, yet the mechanisms at play are still not very well understood. Following replication, the sister chromatid is held close by cohesins (Uhlmann and Nasmyth 1998) which would suggest homology is found by simply being in close proximity; however, this does not guarantee perfect juxtaposition, and in fact sister loci can be more than 1 μm apart (B. Chen et al. 2013). Furthermore, interchromosomal recombination has been demonstrated to be possible, indicating a mechanism for whole-genome scanning (R. R. White et al. 2013). More evidence for genome-wide scanning of the nucleofilament came from RAD51 chromatin immunoprecipitation (ChIP) sequencing (Renkawitz et al. 2013), which demonstrated that RAD51 can contact many regions of the genome over many hours before repair was concluded. Much more work is required to elucidate how the homology search can be conducted efficiently. In contrast, the mechanisms for how homology is probed is better understood. Structural studies performed on the prototypic RAD51 – the bacterial RecA – showed that each RecA in the nucleoprotein filament binds the phosphate backbone of three nucleotides of the ssDNA (Z. Chen, Yang, and Pavletich 2008). This creates a local B-DNA structure at the triplets with surface-exposed bases, which is suggested to be the method of homology probing through Watson-Crick base pairing. When these triplets find a homologous triplet, additional contacts can be made. Once an undetermined degree of homology is captured, the now synaptic filament is a semi-stable paranaemic structure but is still dependent on RAD51 (Riddles and Lehman 1985). Eventually, complete strand exchange occurs and a plectonaemic joint (i.e. a stable helical duplex) is formed which is independent of the nucleofilament. This structure is known as a displacement loop (D-loop) as the plectonaemic joint displaces the other strand of the donor DNA complex (Figure 1.4A).

1.2.6.4 Templated DNA synthesis

Following formation of a D-loop, the 3’ end of the invading strand is utilised as a primer for extension by DNA polymerases (Figure 1.4B). The replication-associated

DNA polymerases Pol δ and Pol ε have both been shown to be the main enzymes in HR-associated DNA synthesis (McVey et al. 2016). They likely act sequentially, where Pol δ extends the invading strand within the D-loop to generate the first end, followed by Pol ε-mediated second-end synthesis to refabricate both resected strands around the DSB.

Two mechanisms have been described for this final stage of HR repair. The most likely mechanism in somatic cells is the synthesis-dependant strand annealing (SDSA) model in which only one resected strand forms a D-loop with the donor duplex (Figure 1.4C). Following synthesis of new DNA, the extended invading strand dissociates from the donor – likely catalysed the RTEL1 helicase (Barber et al. 2008) – and anneals with the resected DNA on the other side of the break. This intermediate then resembles a dsDNA duplex with gaps, which are filled by more DNA polymerases to fix the DSB. Evidence for SDSA being the major mitotic HR pathway comes mostly from the observations that all newly synthesised DNA is located at the recipient duplex and the donor is left intact (Ira, Satory, and Haber 2006).

The alternative pathway is the double Holliday junction (dHJ) mechanism which involves the invasion of both resected strands into the donor duplex (Figure 1.4D). This can either occur as two individual RAD51-mediated synapsing events, or from a SDSA-like model where after extension within the D-loop, the other strand anneals with the displaced strand instead. Within the dHJ structure, both DSB ends are extended and ligated to the other end of the original strand, fixing the DSB but remaining in the tangled dHJ. To complete dHJ they must be either cleaved or dissolved. Cleavage can result in non-crossover (NCO) or crossover (CO) events, where CO is the exchange of DNA from the recipient to the donor and vice versa (Figure 1.4D). In mitotic cells, as mentioned previously, NCO are the dominant form of repair carried out by SDSA. However, if dHJs form in mitotic cells, they can be resolved by dissolving of the structure by BLM helicase and topoisomerase IIIα, which is a NCO event. Dysregulation of the post-D-loop mechanisms can result in loss of heterozygosity, mutagenesis, and copy number variations, all of which are drivers for oncogenesis.

1.2.7 Alternative DSB repair pathways

The two major repair pathways are considered HR and NHEJ, however at least two more pathways have been described. These are alternative-NHEJ (alt-NHEJ), the most

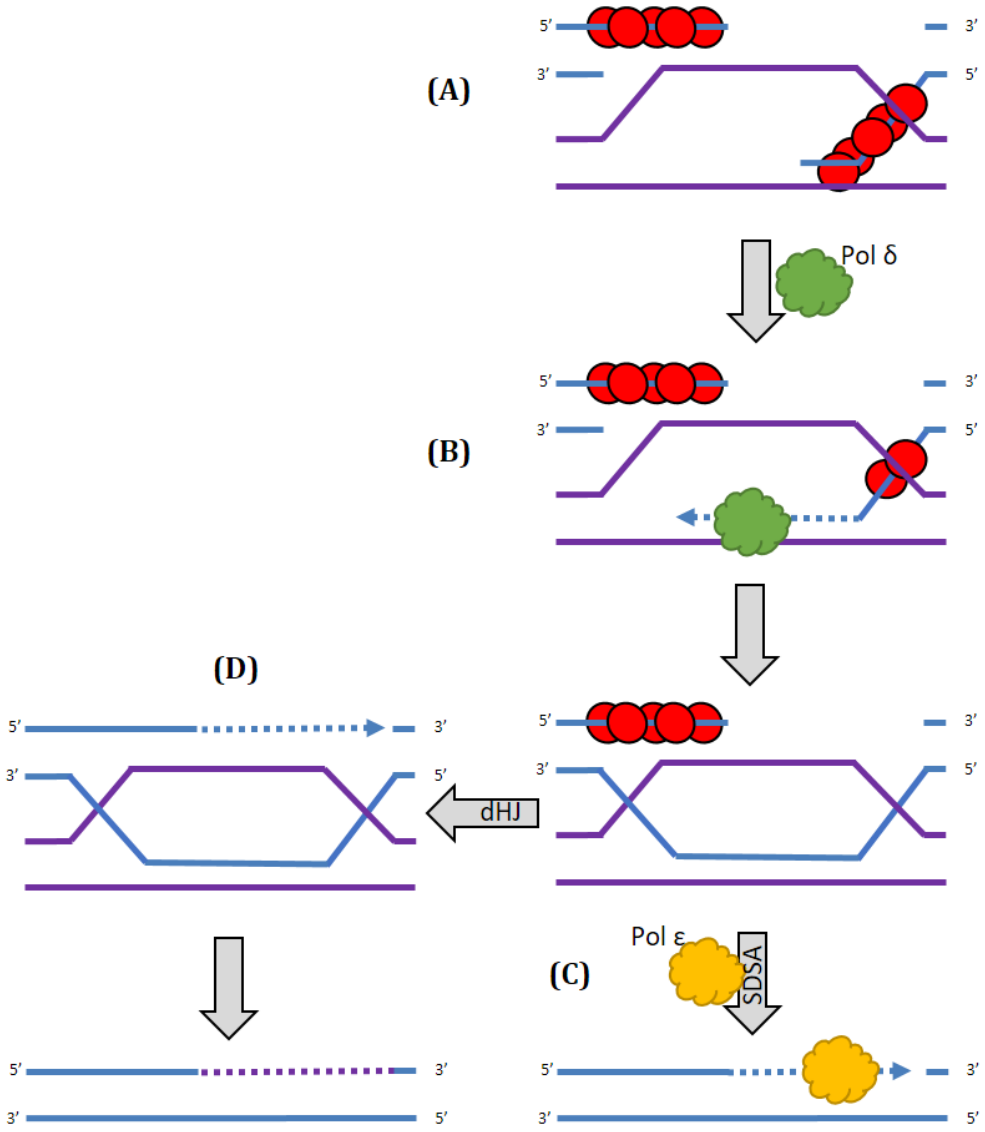


Figure 1.4. Templated HR repair and resolution

(A) Following homology search by probing for complementary sequence, the nucleofilament invades the sister chromatid forming a D-loop. **(B)** A DNA polymerase (Pol δ) synthesises a DNA strand using the homologous sequence as a template. This extends to and is annealed by ligases to the other end of the break. **(C)** Somatic cells then use the SDSA pathway to synthesise the other strand using the newly synthesised DNA as template. **(D)** If the D-loop captures the other nucleofilament, a double Holliday junction (dHJ) is formed. This can be resolved by crossing over in meiotic cells, transferring sequence between the chromatids.

established of which is microhomology-mediated end joining (MMEJ); and single-strand annealing (SSA). Both pathways are dependent on resection. As such, they may take over from NHEJ when uncontrolled resection occurs during the G1 phase, or during S/G2 before replication has occurred at that locus. Both SSA and MMEJ utilise homology regions within resected regions either side of a DSB that hybridise allowing polymerases to extend and ligate the ends together (Bhargava, Onyango, and Stark 2016). This comes with the consequence of losing genetic information, and may be a last-ditch effort by the cell to prevent chromosomal aberrations.

1.2.7.1 Single-strand annealing

SSA repairs DSBs between repetitive sequences. It is dependent on the extensive resection around the break, revealing homology regions within those repeats that can hybridise mediated by RAD52 (Rothenberg et al. 2008; Bhargava, Onyango, and Stark 2016). The 3' non-homologous regions between the hybridised repeats is then trimmed away by the heterodimer ERCC1/XPF (Motycka et al. 2004). Depending on the distance between the repeats, huge amount of genetic information can be lost this way. Following this, DNA polymerases and ligases can extend the gaps and fix the DSB, however which enzymes in particular is currently not very well understood (Bhargava, Onyango, and Stark 2016).

1.2.7.2 Microhomology-mediated end joining

MMEJ is, in principle, similar to SSA in that they both involve resection and annealing of homologous regions *in cis*. However, where SSA requires extensive resection and large regions of homology, MMEJ utilises very short (≤10 nt) microhomology and reduced resection (Bhargava, Onyango, and Stark 2016). For microhomologous annealing, RAD52 is dispensable and is likely to occur spontaneously, showing a key difference to SSA (Bennardo et al. 2008); however, following annealing, the 3' flaps are excised by ERCC1/XPF as in SSA. Both SSA and MMEJ also utilise the same standard resection machinery, namely the MRN complex, CtIP, and EXO1 (Truong et al. 2013). How these are regulated to only resect a small distance in MMEJ versus longer distances for HR and SSA was unclear, however the recent observations that 53BP1 in fact promotes HR by limiting extensive SSA-prone resection may provide a mechanism for this (Ochs et al. 2016). Indeed, an earlier paper highlighted a requirement for both

53BP1 and BRCA1 in MMEJ, suggesting regulation of resection is important for MMEJ repair (Xiong et al. 2015).

Other proteins involved in MMEJ are more established than SSA. DNA polymerase theta (Pol θ) was shown to be one of the main polymerases involved in MMEJ, as it can initiate DNA synthesis from microhomology priming (Kent et al. 2015). Following extension of the microhomology, the synapse is stabilised and resembles two ssDNA breaks in proximity. As such, the SSB repair ligase – DNA ligase IIIα – was shown to participate in MMEJ; as was the Okazaki fragment joiner, DNA ligase I (Liang et al. 2008) (see section 1.2.8.1 for a brief SSB repair overview). Overall regulation of MMEJ also appears to be more understood compared to SSA. MMEJ fundamentally requires poly-ADP-ribose polymerase (PARP) for both synapsis of the DNA ends and to compete with Ku for end binding, and is required for the recruitment of both Pol θ and ligase IIIα (Audebert, Salles, and Calsou 2004; M. Wang et al. 2006; McVey and Lee 2008; Bhargava, Onyango, and Stark 2016).

1.2.8 Repair of non-DSB DNA damage

As discussed in section 1.2.1, DSBs are only one manifestation of DNA damage within a cell. Other forms of damage include SSBs, pyrimidine dimers from UV-B radiation, base mismatches, DNA adduct formation, and modification of bases through oxidation, alkylation, and hydrolysis. Several pathways exist to repair each form of damage; singular base modifications can be repaired via base excision repair (BER), bulky modifications from adduction and UV by nucleotide excision repair (NER), DNA mismatches by mismatch repair (MMR), and SSBs by SSB repair (SSBR).

1.2.8.1 Single stranded break repair

SSBs occur from IR and ROS damage, and are intermediates from several other repair pathways such as BER (discussed below) and MMEJ (section 1.2.7.2). As the genetic information is found in the undamaged complementary strand of the DNA duplex, SSBs can be repaired very efficiently; however, unrepaired SSBs can cause replication fork collapse leading to the formation of DSBs (Kuzminov 2001). SSBs are detected by PARP1, which rapidly PARylates histones H1 and H2B which allows recruitment of downstream factors, including the scaffold protein XRCC1 (London 2015). Most forms of damage will modify the SSB termini in some way, necessitating the recruitment and activity of a variety of end processing factors to restore the 5’-monophosphate and

3’-OH. These ends can then be ligated following filling of any DNA gaps. SSB-associated DNA synthesis is mainly carried out by DNA pol β, however there is some evidence for the involvement of others such as Pol δ (Parsons et al. 2007). Ligation is carried out by DNA ligase IIIα, resolving the SSB.

1.2.8.2 Base excision repair

Base damage includes oxidation, alkylation, and deamination. When these forms of damage are small tracts (a single nucleotide) and do not distort the helical structure, they are repaired via BER. BER is effectively the removal of the damaged nucleotide, followed by SSB repair of the lesion. DNA glycosylases initiate the first step of BER by removing the damaged base. At least eleven BER DNA glycosylases have been described, each removing bases with different modifications. The resulting abasic site (AP site) nucleotides are cleaved by AP endonucleases, which produce a SSB that is then repaired via SSBR.

1.2.8.3 Nucleotide excision repair

UV radiation are high energy photons that can create free radicals (UV-A) leading to SSBs or bulky DNA lesions (UV-B). UV-B induces covalent crosslinking between adjacent thymine or cytosine bases, distorting the DNA helix and stalling replication and transcription. These pyrimidine dimers are repaired via NER, a similar pathway to BER in that the steps involved are damage recognition, excision of the lesion, and gap filling.

Two NER sub-pathways exist which differ mainly at the damage recognition and pre-processing stages. Global genomic (GG-) NER is active at all times and involves the constant probing (random, diffusion-limited collisions) of the genome for helix distortions by XPC (in a complex with RAD23B and CETN2) and DDB2 (in the UV-DDB complex containing DDB1, CRL) (Min and Pavletich 2007; Hoogstraten et al. 2008; Moser et al. 2005; Wittschieben, Iwai, and Wood 2005). When the bulky adducts occur on the template strand of a transcribing gene, Pol II gets stalled triggering transcription-coupled (TC-) NER. This leads to the recruitment of the Cockayne syndrome proteins CSA and CSB, which further recruit UV-stimulated scaffold protein A (UVSSA) and USP7 (Schwertman et al. 2012). Once the damage has been detected, the downstream steps are the same for both GG-NER and TC-NER. The transcription factor TFIIH is a multifactor complex containing the helicases XPB (ERCC3) and XPD

(ERCC2) within its core module (section 1.3.2.1). It is recruited through XPB to the damage site, where XPD opens the duplex through its helicase activity (Fuss and Tainer 2011). The TFIIH-stabilised open bubble allows the recruitment of XPA and RPA, which bind to the modified nucleotides and the ssDNA on the undamaged strand respectively. RPA binding triggers the ATR signalling cascade, leading to the phosphorylation of H2AX and recruitment of DDR factors (see section 1.2.3). This does not appear to be a general feature of NER but instead a backup mechanism when NER fails; the non-NER factor EXO1 (section 1.2.6.1) was shown to create extended ssDNA gaps, increasing RPA recruitment and thus increased ATR activation (Giannattasio et al. 2010). XPA binding acts to verify the bulky lesion, and initiates strand excision and gap filling by recruiting endonucleases and PCNA. The endonucleases ERCC1-XPF and XPG – positioned by XPA and RPA – cleave the damaged strand 5' and 3' of the adduct, leaving a 22-30 nt gap (De Laat et al. 1998; Mocquet et al. 2008; Staresincic et al. 2009). PCNA binding and ERCC1-XPF cleavage allows recruitment of DNA polymerases (Pols δ, ε, or κ) and ligases (DNA ligase I and IIIα) which synthesise new DNA, repairing the damage (Ogi et al. 2010).

1.2.8.4 Mismatch repair

MMR is another excision repair pathway that targets nucleotide misincorporations. The main replicative DNA polymerases provide a very high degree of fidelity thanks to proofreading activities, with an error rate of approximately 1 in 10^7 bp per cell division (Kunkel 2004). When base substitutions or indels slip by, MMR can identify and correct these, contributing an additional 10^{-2} to 10^{-3}-fold to the total replication fidelity. Different mismatch repair proteins identify different errors; for example, the MutSα human homologue heterodimer Msh2/6 recognises base substitutions whereas the MutSβ heterodimer Msh2/3 recognises small and large indels (Peña-Diaz and Jiricny 2012). These complexes recruit downstream factors including the endonuclease MutLα, a heterodimer consisting of MLH1 and PMS2. These nick the mismatched strand, which is then excised by EXO1 and resynthesized by DNA polymerases (Peña-Diaz and Jiricny 2012).

1.3 Transcription

Transcription is the process of generating RNA from DNA and thus the first step in gene expression. The 5'-most nucleotide of a gene is known as the transcriptional start site (TSS) and is the first nucleotide incorporated by one of the three RNA polymerase enzymes in humans. Transcription occurs in the 5'→3' direction, generating a complementary (or antiparallel) RNA molecule to the template strand of the DNA; this primary (or *de novo*) transcript is therefore identical to the non-template strand with the exception that thymine is replaced by uracil in RNA. The final 3' nucleotide of a gene is known as the transcription end site (TES; a.k.a. transcriptional termination site, TTS). In addition to the TSS and TES, all genes have a promoter element. This is the sequence upstream of the TSS that the RNA polymerase assembles upon to initiate transcription, but is not transcribed itself as part of the gene.

Each of the three human RNA polymerases transcribes different classes of RNA. Pol I exclusively transcribes the 45S pre-ribosomal RNA (rRNA), the most abundant RNA (by mass) within a cell. Pol III transcribes the smaller 5S rRNA, as well as transfer RNA (tRNA) and several other types of small non-coding RNA (sncRNA) (see section 1.4.1.1). The final and most well-studied polymerase, Pol II, synthesises messenger RNA (mRNA) from protein-coding genes, as well as microRNAs (miRNA) and the remainder of the short and long non-coding RNAs (ncRNA). All three polymerases have three stages to their overall activity: initiation, elongation, and termination. Initiation is the assembly of the polymerase upon the gene promoter, elongation is the escape of the polymerase from the promoter element and synthesis of the RNA, and termination is the eventual release of the polymerase and the nascent transcript. Each of these steps is incredibly complex and subject to an enormous degree of regulation. This section will focus mainly on Pol II as it is the most well-studied and appropriate to the work carried out in this thesis.

1.3.1 The nuclear RNA polymerases

The human RNA polymerases are multi-subunit enzymes; Pol II contains 12 subunits whereas Pol I has 14 and Pol III has 17. Most of these subunits are shared between or have homologous counterparts in each polymerase (Jasiak et al. 2006; Kuhn et al. 2007). Of particular interest in terms of transcriptional progression and regulation is the largest Pol II subunit, RPB1, which contains the C-terminal domain (CTD). This is

the conserved and unstructured heptapeptide $Y_1S_2P_3T_4S_5P_6S_7$ which occurs 52 times in tandem; it acts as both a scaffold for various co-transcriptional factors and is also subject to heavy phosphorylation at different stages of transcription (Corden 1990).

1.3.2 Initiation

1.3.2.1 Assembly of the pre-initiation complex

Transcription of a gene first requires the assembly of the pre-initiation complex (PIC; not to be confused with the ribosomal 43S PIC) at the promoter of the gene (Figure 1.5). This consists of Pol II and the general transcription factors (TFs) TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH (TF for Pol II). With the exception of TFIIB, the general TFs are multi-subunit complexes; the largest, TFIID, consists of at least 14 subunits totalling over 1 MDa with the final PIC working out to be around 2.5 MDa (Bieniossek et al. 2013).

The classic model of initiation is a stepwise assembly of the PIC on exposed promoter elements following chromatin remodelling and modification at the promoter. The binding of upstream activators, cell type-specific and gene-specific TFs, and inducible TFs to the limited exposed regions between nucleosomes in the promoter region recruits those histone modifiers that make the assembly of the PIC possible.

The first stage in PIC assembly once the promoter is amenable to its formation is the recruitment of TFIID (Figure 1.5B). TFIID contains the crucial factor TATA box-binding protein (TBP), a saddle shaped protein that can bind the TATA box, and 13 TBP-associated factors (TAFs) (J. L. Kim, Nikolov, and Burley 1993; Y. Kim et al. 1993; Louder et al. 2016). The TATA box is the consensus sequence $TATA^{A}_{T}A^{A\ A}_{T\ G}$ and is located at position -200 to -50 in around 20% of yeast genes (Basehoar, Zanton, and Pugh 2004). In humans, this is located ~30 nt upstream of the TSS and is only present at around 10% of genes (Carninci et al. 2006; C. Yang et al. 2007). Observations that TBP is located at promoters genome-wide suggests that the TATA box is not the only feature of promoters that is important for formation of the PIC (Sung Rhee and Franklin Pugh 2012). The 13 TAFs within TFIID can recognise additional sequences such as the initiator sequence (YYANWYY) which is found at nearly half of all promoters, suggesting that TFIID through TBP primarily binds non-specifically to the DNA minor groove and is directed to promoters with varying binding elements by its repertoire of

TAFs (Chalkley and Verrijzer 1999; C. Yang et al. 2007). Once bound, TBP causes a ~90° kink and partial unwinding in the dsDNA (Figure 1.5B) (J. L. Kim, Nikolov, and Burley 1993; Y. Kim et al. 1993). TFIID can be stabilised on DNA by TFIIA, however this is not essential and likely dependent on binding elements within the promoter (Louder et al. 2016).

TFIIB further stabilises TFIID binding to the promoter, and is also important for Pol II recruitment (Figure 1.5C) (Bushnell et al. 2004; Sainsbury, Niesser, and Cramer 2012). It can interact with GC-rich regions upstream or downstream of TATA/TBP binding which may help orientate further PIC assembly on the correct strand (Lagrange et al. 1998; Bushnell et al. 2004). The domains at the N-terminus of TFIIB enter the Pol II active site through the RNA exit channel and makes direct contact with the template strand of the DNA (Sainsbury, Niesser, and Cramer 2012). This is thought to help Pol II find the correct TSS, keep the initial transcription bubble open, and participate in abortive initiation (discussed below in section 1.3.2.2) by sterically blocking the elongating RNA strand (Bushnell et al. 2004; Sainsbury, Niesser, and Cramer 2012). The binding of TFIIF stabilises TFIIB and the PIC, however, this is not essential for all promoters (Cabart et al. 2011; Fishburn and Hahn 2012).

The final general TFs required to form the complete PIC are TFIIE and TFIIH (Figure 1.5D). TFIIE binds to Pol II and can associate with ssDNA, potentially contributing to the formation of the transcription bubble; however, its primary role appears to be to recruit TFIIH to the PIC (Okamoto et al. 1998; Maxon, Goodrich, and Tjian 1994). TFIIH consists of a core module (which is involved in the NER pathway, discussed in section 1.2.7) and a kinase module, both of which are critical for transcription initiation (Ohkuma and Roeder 1994; Holstege, van der Vliet, and Timmers 1996). Where XPD activity is required in NER, XPB is required for transcription initiation (Guzmán and Lis 1999). XPB binds to DNA downstream of the PIC-bound promoter but, instead of utilising its ATP-dependent helicase activity, may unconventionally translocate the downstream DNA whilst the upstream PIC:DNA remains stationary; this torque then causes the melting of DNA (T.-K. Kim, Ebright, and Reinberg 2000). This requires ATP and changes the PIC from a “closed” state to an “open” state.

In the presence of NTPs, Pol II within the open PIC can begin DNA-dependent RNA synthesis starting at the TSS nucleotide of the unwound template DNA strand.

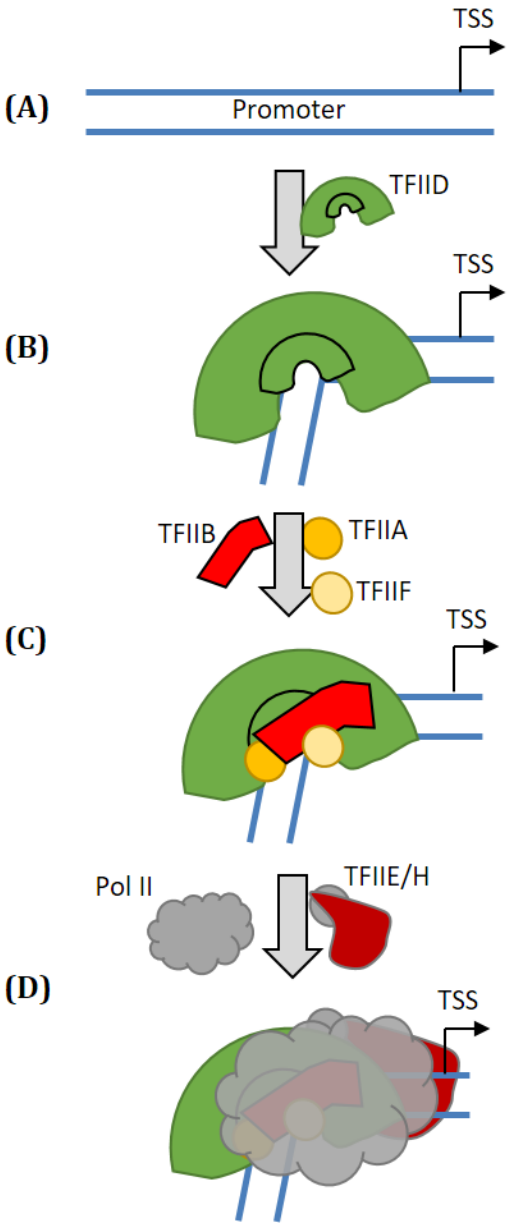


Figure 1.5. Assembly of the pre-initiation complex at promoters.
(A) Upstream of the transcription start site (TSS) of a gene is the promoter, on which the pre-initiation complex (PIC) forms. **(B)** First, TFIID binds the promoter through its TBP subunit. This kinks the DNA ~90°. **(C)** Recruitment of TFIIA, B, and F stabilise TFIID and begin the melting of the duplex DNA. **(D)** Recruitment of Pol II and TFIIH forms the complete PIC. When NTPs are present, TFIIH phosphorylates Pol II on its CTD tail (not shown) and elongation begins.

1.3.2.2 Abortive transcription

Before RNA Pol II can dissociate from the PIC, transcription is very frequently aborted resulting in the release of truncated transcripts <10 nt. This abortive initiation can occur for many cycles until Pol II eventually escapes the promoter, at which point the CDK7 component within the kinase module of TFIIH phosphorylates serine 5 (Ser5; Ser5P) on the Pol II CTD and transcription enters the elongation phase. How and why abortive transcription exists is still under investigation, however several key pieces of evidence suggest that abortive transcription may be to generate enough free energy to overcome the stability of the PIC (Henderson et al. 2017). A DNA scrunching mechanism for transcription initiation was first described in two companion publications (Kapanidis et al. 2006; Revyakin et al. 2006). Here, it was demonstrated that Pol II within the PIC open complex remains in place on the promoter and TSS and pulls the downstream template DNA strand into its active site. This causes flexible DNA scrunching within and external to Pol II as the nascent RNA strand is being produced. In rounds of abortive transcription, the <10 nt transcript is released, as is the scrunched DNA. Very recently, studies on *E. coli* RNA polymerase demonstrated that each cycle of scrunching generates free energy which allows progressively longer abortive transcription; eventually, this stored energy is great enough to overcome the open complex allowing promoter escape into elongation (Henderson et al. 2017).

1.3.3 Elongation

Elongation is split into two main phases: early and productive (Jonkers and Lis 2015). In early elongation, Pol II pauses within 30-60 nt downstream of the TSS and is a source of additional regulation. Productive elongation continues after unpausing until termination of the transcript and synthesises the entire RNA molecule. During elongation, the nascent transcript is processed through capping and splicing.

1.3.3.1 Early elongation

Promoter-proximal pausing is mediated by negative elongation factor (NELF) and DSIF, which interact with the TFs at the promoter elements and tethers Pol II. Additionally, the +1 nucleosome (that is, the first nucleosome downstream of the open promoter) acts as a barrier to pause transcription (Weber, Ramachandran, and Henikoff 2014). Unpausing of Pol II is dependent on P-TEFb (positive transcription elongation factor b) which is complex containing the kinase CDK9. Recruitment of P-TEFb leads to

phosphorylation of the repressive NELF and DSIF, switching DSIF to a positive elongation factor and leading to dissociation of NELF. P-TEFb is required at the vast majority, if not all, genes (Henriques et al. 2013; Jonkers, Kwak, and Lis 2014). The release of Pol II from promoter-proximal pause sites is also coupled to the additional phosphorylation of Pol II CTD at Ser2 (Ser2P), catalysed by P-TEFb.

1.3.3.2 Productive elongation

Productive elongation proceeds the entire gene length until it reaches the termination signals (see section 1.3.4). Where open PIC-associated Pol II and promoter-proximal Pol II is marked by Ser5P CTD, productively elongating Pol II is marked by increasing Ser2P CTD (Zaborowska, Egloff, and Murphy 2016). This appears to coincide with variable elongation rates of Pol II; acceleration from 0.5 kb/min to 2-5 kb/min takes place over many kilobases (Jonkers, Kwak, and Lis 2014; Jonkers and Lis 2015). Additionally, productively elongating Pol II is also subject to pausing and deaccelerations. For example, Pol II stalls at exons however this may be due to splicing. Yet more variability in elongation arises from the chromatin architecture ahead of Pol II; tightly packed nucleosomes require co-transcriptional histone remodelling to allow elongation to proceed (Jonkers and Lis 2015).

1.3.3.3 Co-transcriptional processing

Maturation of a Pol II-transcribed RNA occurs co-transcriptionally as many of these factors are recruited through the Pol II CTD. Two key events occur co-transcriptionally: 5’ capping and exon splicing.

Messenger RNAs (as well as primary miRNA and other long non-coding RNA) require capping of their 5’ ends to protect against surveillance exonucleases, and for association of eIF4E in the eIF4F complex in the cytoplasm to initiate cap-dependent translation of the mRNA. Capping occurs post-initiation in early elongation; one proposed function of promoter-proximal pausing is to ensure mRNAs have been fully capped. The capping machinery consists of RNA triphosphatase (TPase), RNA guanylyltransferase (GTase), and guanine-N7 methyltransferase (RNMT) and associates with Ser5P on the Pol II CTD. TPase removes the γ-phosphate from the first nucleotide of the nascent RNA, then GTase ligates a GMP nucleotide to form the unusual G(5’)ppp(5’)RNA triphosphate linkage. Finally, RNMT catalyses the methylation of position N7 of the guanosine cap, generating the final m7GpppRNA cap.

A major event during elongation is splicing. This is the process of excising intronic sequences from the nascent mRNA and joining of exons. Splicing is carried out by the spliceosome, a large complex containing multiple protein and RNA subunits; this interacts with the Pol II CTD and processes the mRNA co-transcriptionally (Matera and Wang 2014). Splicing allows the generation of multiple isoforms and thus different proteins from the same gene unit by differentially retaining some exons over others, known as alternative splicing (M. Chen and Manley 2009).

1.3.4 Transcriptional termination

The final phase of transcription is the termination of elongation. There are at least two competing models of termination in the literature, however these are not necessarily mutually exclusive. Nevertheless, termination is much simpler and less regulated than initiation and elongation.

1.3.4.1 Polyadenylation signal-dependent release of mRNA

As elongation proceeds, the initial Ser5P CTD is increasingly replaced by Ser2P which causes a switch from pro-elongation factors associated with Pol II to pro-termination factors known as cleavage and polyadenylation (CPA) factors (Figure 1.6A) (Davidson, Muniz, and West 2014). These include cleavage and polyadenylation specificity factors (CPSF), cleavage stimulatory factors (CstF), and cleavage factors I and II (CFI/II). When elongating Pol II reaches the polyadenylation signal (PAS) near the TES of the gene, the endonuclease CPSF73 binds and cleaves the nascent transcript 20-30 nt downstream of the PAS (Figure 1.6B) (Ryan, Calvo, and Manley 2004). CPSF-dependent cleavage releases the nascent transcript, which is polyadenylated by poly(A) polymerase (PAP). Polyadenylation is the process of adding up to several hundred adenosines without requiring a template (Figure 1.6C) and is important for mRNA stability and translational regulation. The poly(A) tail is coated by poly(A) binding protein (PABP) and the now fully mature mRNA is exported from the nucleus for translation in the cytoplasm.

1.3.4.2 Termination of Pol II

Whilst the release, polyadenylation, and export of the mRNA ends the story of transcription for the mRNA, Pol II is still associated with the DNA and actively transcribing. If Pol II does not get dissociated, it continues elongating leading to read-through into neighbouring genes; at best, this is a waste of cellular resources and at worst can lead to aberrant transcripts or collisions with other RNA polymerases or

the replication machinery causing genomic instability (Price, Alm, and Arkin 2005; N. Kim and Jinks-Robertson 2012).

As mentioned briefly, there are two models for how Pol II is released from the DNA. The first is known as the allosteric model, which hypothesises that the binding of CPA factors to Pol II CTD and the PAS of the nascent chain causes pausing of and conformational changes to Pol II (H. Zhang, Rigo, and Martinson 2015). This appears to be sufficient to dissociate Pol II from the template *in vitro*, but *in vivo* may be too inefficient.

The alternative and more accepted model is referred to as the torpedo model. Following PAS-dependent cleavage and release of the mRNA, the presence of the resulting 5'-monophosphate nascent RNA recruits 5'-3' exoribonuclease 2 (XRN2) (Figure 1.6D). XRN2 digests the still-elongating nascent transcript until it reaches Pol II; this collision causes termination and release of the polymerase (Figure 1.6E). This clearly requires some form of Pol II slow down or pausing as evidenced by the use of slowly elongating Pol II mutants which terminate sooner versus faster mutants that terminate more downstream (Fong et al. 2015). This pausing or reduction in rate of elongation could be due to the association of factors with Pol II as in the allosteric model, the presence of downstream heterochromatin, or, as shown recently, the formation of R-loops (Skourti-Stathaki, Proudfoot, and Gromak 2011). In this model, the nascent transcript hybridises with the single stranded non-template or antisense DNA strand within the transcription bubble, reducing the rate of Pol II elongation (Figure 1.6D). This allows XRN2 to catch up to the polymerase; the helicase senataxin (SETX) was also demonstrated to unwind the hybrid ahead of XRN2 allowing it to exonucleolytically digest the ssRNA (Figure 1.6DE). This model is further backed up by the observations that SETX is recruited to Pol II CTD through the SMN protein, the removal of which results in termination defects (Yanling Zhao et al. 2015).

1.3.4.3 PAS-independent release of nascent RNA

Messenger RNA only makes up a small proportion of the steady state RNA levels. As such, most transcription will not rely on PAS-dependent termination.

RNA polymerase III transcribes various smaller non-coding RNA (sncRNA, section 1.4.1.1) including the 5S rRNA and tRNA. When it reaches a stretch of 4 or more thymines in the non-template strand, elongation is heavily reduced and this causes

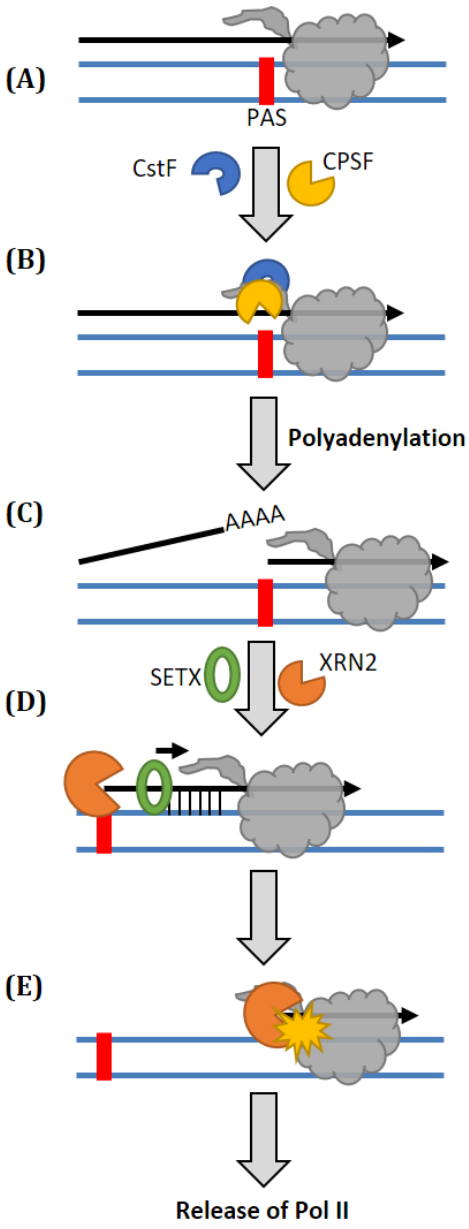


Figure 1.6. Torpedo model for transcriptional termination.
(A) When an elongating Pol II transcribes the polyadenylation signal (PAS), CPA factors are recruited to the CTD tail **(B)** CPA factors such as CPSF cleave the nascent RNA at the PAS. **(C)** The 3' end of the mRNA is polyadenylated by PAP (not shown) and is then exported to the cytoplasm. RNA Pol II however continues transcribing. **(D)** In the R-loop-driven torpedo model, R-loops form behind the elongating Pol II, causing it to slow down. XRN2 exonucleolytically digests the 5' end of the RNA, requiring SETX to unwind the R-loop ahead of it. **(E)** XRN2 catches up to Pol II and collides with it. This collision causes a conformation change in Pol II leading to its release from the DNA, concluding transcription.

allosteric termination of Pol III (Bogenhagen and Brown 1981; Matsuzaki, Kassavetis, and Geiduschek 1994). More recently it was demonstrated that the poly(T) signal may cause backtracking (Nielsen, Yuzenkova, and Zenkin 2013). When the reversing Pol III eventually reaches RNA secondary structure it clashes causing allosteric termination; however, this has been somewhat controversial (Arimbasseri, Kassavetis, and Maraia 2014; Nielsen and Zenkin 2014).

Pol I transcribes the 45S pre-rRNA which, when released, is processed into the 18S, 5.8S, and 28S rRNAs which form major components of the ribosome. Downstream of each rDNA repeat is a series of the Sal-box terminator element, an 18 bp sequence that is bound by TTFI (transcription termination factor I) (Smid, Finsterer, and Grummt 1992; Bartsch, Schoneberg, and Grummt 1988). This kinks the DNA causing Pol I to pause; the association of Pol I transcript release factor (PTRF) then dissociates Pol I from the template (Mason, Sander, and Grummt 1997). It is currently not clear as to what initiates the cleavage of the 3' end of the terminated pre-rRNA in higher eukaryotes; however, in yeast the RNase III enzyme Rnt1 was shown to cleave RNA hairpins at the 3' end to release the pre-rRNA (Kawauchi et al. 2008). Here, it was also demonstrated that Rnt1 could cleave Pol II-elongated transcripts in a PAS-independent manner, again reliant on XRN2 and a R-loop-mediated torpedo model.

In mammalian cells, the nuclear RNase III enzyme is Drosha (see section 1.4.2). Recently, it was shown that Drosha was involved in the termination of certain long non-coding RNAs (lncRNA; section 1.4.1.2) (Dhir et al. 2015). Loss of Drosha did not affect termination of mRNAs with PAS, but did lead to readthrough of microRNA-containing lncRNA. Whether Drosha can terminate rRNA and other RNA at secondary structures is yet to be determined.

1.4 Non-coding RNAs

Whilst most studies on transcription focus on the production and processing of mRNAs, the most abundant transcripts within a cell are in fact non-coding RNAs (ncRNAs). These consist of rRNA (80-90% of total by mass), tRNA (10-15%), long ncRNA (lncRNA) (<1%), and small ncRNA (sncRNA) (<1%); mRNA constitutes 3-7% of the total RNA (Palazzo and Lee 2015).

1.4.1 Classes of ncRNA

1.4.1.1 Small non-coding RNA

The small non-coding RNA are those less than 200 nt. There are many classes of sncRNA with very clearly defined functions, such as tRNA, miRNA (section 1.4.2), piwi-interacting RNA (piRNA), small nuclear RNA (snRNA), and small nucleolar RNA (snoRNA). Many more classes exist with less understood functions, such as enhancer RNA (eRNA), TSS-associated/promoter-associated sRNA (TSSa-RNA/PASR), and sense/antisense termini-associated sRNA ((a)TASR) (Kapranov et al. 2007; Seila et al. 2008; Fejes-Toth et al. 2009; Taft et al. 2009; Kapranov et al. 2010).

1.4.1.2 Long non-coding RNA

Long non-coding RNA are those longer than ~200 nt, and so include both the 18S and 28S rRNA and the telomerase RNA component. Those long ncRNA that resemble mRNA are simply referred to as lncRNA, whereas those that originate from intergenic regions are sometimes referred to as long intergenic ncRNA (lincRNA), although this is not used as often. LncRNA has been implicated in many cellular pathways such as chromatin organisation (Melé and Rinn 2016), the most famous example of which is the inactivation of one copy of the X chromosome in female cells by Xist (Penny et al. 1996; Froberg, Yang, and Lee 2013). Other roles for lncRNA include transcriptional regulation (Engreitz et al. 2016) and epigenetic regulation (Gupta et al. 2010), with many more likely to be described.

1.4.2 MicroRNAs

The miRNA-mediated translational repression pathway is a critical form of post-transcriptional gene regulation within the cell. This involves the imperfect base pairing of the ~22 nt mature miRNA to its target mRNA, leading to negative control of translation initiation or elongation and destabilisation of the mRNA itself (Figure 1.7).

miRNAs can have potentially many targets, and many miRNAs can have the same target. It is now thought that most, if not all, mRNAs (and by extension cellular pathways) are controlled in some way by miRNAs. In development, the miRNA pathway is essential and knockouts of components of the biogenesis pathway are embryonically lethal (Pan et al. 2011; Fan et al. 2013). A major function of miRNAs is fine control over the cell cycle, as well as to provide a means of rapid translational reprogramming in response to stress stimuli and signalling (Carleton, Cleary, and Linsley 2007; Gulyaeva and Kushlinskiy 2016).

The steps involved in the biogenesis of miRNAs is now very well-established, with several key events defining the process. They are encoded as genes which can contain a singular miRNA or a cluster of many. These genes typically contain only miRNAs; however, many miRNAs have also been found to be located within the introns and even the exons of mRNA. The primary miRNA (pri-miRNA) is transcribed by Pol II and – like mRNAs – are capped, polyadenylated and sometimes spliced (Figure 1.7A) (Mondol, Ahn, and Pasquinelli 2015). Additionally, many pri-miRNAs can undergo RNA editing. For example, the Adenosine Deaminase Acting on RNA (ADAR) proteins can catalyse deamination of adenosine residues to inosine, which can inhibit processing of the miRNA, affect stability, and change targeting specificity (Kawahara et al. 2008; Kawahara et al. 2007; W. Yang et al. 2006). Maturation of a miRNA involves sequential processing by Drosha in the nucleus and Dicer in the cytoplasm.

1.4.2.1 Nuclear processing by Drosha

Primary miRNAs contain secondary structures consisting of a ~33 bp stem – which harbours the mature miRNA sequence – and an apical loop (Auyeung et al. 2013). This structure is recognised by the Microprocessor complex within the nucleus and liberates the hairpin 11 bp from the base of the stem, resulting in a smaller ~70 nt hairpin with a 2 nt 3' overhang, known as the precursor miRNA (pre-miRNA).

The Microprocessor complex is heterotrimeric, consisting of Drosha and two DGCR8 proteins. DGCR8 (DiGeorge syndrome critical region 8; a.k.a. Pasha [Partner of Drosha] in *Drosophila* and *C. elegans*) is a nuclear RNA binding protein, with a ssRNA-binding Rhed domain and two dsRNA binding domains (dsRBDs) which it uses to bind the upper stem and apical loop of a pri-miRNA. An additional C-terminal region is critical for Drosha interaction (Yeom et al. 2006). It was thought to direct Drosha to

pri-miRNAs by binding as dimers to both the basal and apical junctions of the stem loop, totalling 4 DGCR8 molecules per pri-miRNA. The guiding role of DGCR8 was challenged when studies on recombinant DGCR8 demonstrated that it will bind non-specifically to dsRNA (Roth, Ishimaru, and Hennig 2013); and, yet more recently, the stoichiometry of the functional Microprocessor was solved showing that there is only a single DGCR8 dimer per pri-miRNA (Figure 1.7B) (Nguyen et al. 2015).

Drosha is a nuclear endoribonuclease containing two RNase III domains (RIIIDa/b) and a dsRBD at the C-terminus. The two RIIIDs form an intramolecular dimer, with each catalysing the cleavage of a single RNA strand (Han et al. 2004): RIIIDa cuts the 3' strand and RIIIDb cuts the 5' strand. Drosha "measures" 11 bp from the basal junction most likely due to steric hindrance from domains in its central regions. Curiously, a UG motif in the ssRNA basal segment was found to be important for effective processing by Drosha, however no obvious part of Drosha was found to contact this when the crystal structure was solved (Auyeung et al. 2013; Nguyen et al. 2015; Kwon et al. 2016). Instead, Drosha likely measures 11 bp from any ssRNA-dsRNA junction; this is backed up from the identification of non-productive processed products where Drosha cleaves from the apical junction of the pri-miRNA (Nguyen et al. 2015). DGCR8 therefore acts to increase the likelihood that Drosha binds to pri-miRNAs and in the correct orientation (Figure 1.7B).

The pre-miRNA is then exported from the nucleus by Exportin-5 (XPO5), a sequence-independent dsRNA binding protein which requires Ran-GTP activity (Figure 1.7C) (Yi et al. 2003; Bohnsack, Czaplinski, and Gorlich 2004). Whilst mutations and knockout of XPO5 strongly reduce miRNA maturation within the cell, recent data suggests that there may be redundancy from other factors (Xie et al. 2013; Y.-K. Kim, Kim, and Kim 2016; Melo et al. 2010).

1.4.2.2 Cytoplasmic processing by Dicer

Within the cytoplasm, the pre-miRNA is processed by Dicer, another RNase III endoribonuclease (Figure 1.7D). Human Dicer anchors both ends of the stem loop; its PAZ domain recognises the 2 nt 3' overhang whilst the 5'-pocket domain binds the recessed 5' monophosphate (Park et al. 2011). Dicer cleaves away the terminal loop of the pre-miRNA ~22 nt from the 5' end generating another 2 nt 3' overhang. It utilises two RIIIDs in an intramolecular dimer to cut each strand, similarly to Drosha. This ruler

activity is known as the 5’ counting rule and appears to be specific to miRNA-producing Dicers; the 5’-pocket motif is conserved in *Drosophila* Dicer-1 but not the siRNA processing enzyme Dicer- 2 (Park et al. 2011). Likewise, the miRNA-devoid lower eukaryote *Giardia lamblia* Dicer – which most Dicer structural studies have been performed upon – also lacks this motif and the 5’ counting rule and instead cuts 25 nt from a blunt dsRNA end (MacRae 2006).

X-ray crystallography studies identified huge structural similarities between Drosha and Dicer, even though they are very dissimilar based on sequence (Kwon et al. 2016). As Dicer is found in most eukaryotes, and Drosha only in metazoa, it is likely that Drosha evolved from an early ancestral animal Dicer.

1.4.2.3 Incorporation into Argonaute

The diced pre-miRNA is a ~22 nt dsRNA with 2 nt 3’ overhangs which is then incorporated into an Argonaute (Ago) protein. Ago discards one strand of the duplex – referred to as the passenger – and retains the other as the mature guide miRNA.

The choice of which strand is the guide and which is the passenger appears to be determined by the relative thermodynamic stability of each end; the strand with the least stable 5’ end is incorporated as the guide and the passenger is degraded (Krol et al. 2004; Schwarz et al. 2003). However, either strand can be selected as the guide from many miRNA genes, and so the miRNA has an additional identifier as -5p or -3p depending on which arm of the pre-miRNA hairpin it was found in. Curiously, some cell and tissue types have a preference for different strand miRNAs to others, indicating there may be other active processes that determine strand selection (Biasiolo et al. 2011; He et al. 2012). In *Drosophila* and human cells, the Translin-Trax octamer C3PO was demonstrated to actively process the passenger strand associated with Ago (Y. Liu et al. 2009; Ye et al. 2011). Furthermore, Dicer and its cofactors TRBP and PACT were shown to directly interact with Ago to transfer the mature miRNA and assist in strand selection (H.-W. Wang et al. 2009; Noland and Doudna 2013). At a similar time, the RNA helicase DHX9 was also shown to participate in this complex (Fu and Yuan 2013). Effectively, this process doubles the miRNA repertoire of the cell without requiring twice the transcription and initial processing, a highly energy intensive process.

The PAZ domain of Ago, like Dicer, binds the 3’-hydroxyl (-OH) moiety of the miRNA, whilst the interface between the PIWI-Mid domains binds the 5’-monophosphate

(Boland et al. 2011). A miRNA-loaded Ago protein forms the basis of the RNA-induced silencing complex (RISC). An 8 nt sequence within the miRNA – known as the seed sequence – directly and imperfectly base pairs with its target mRNAs, typically within the 3' untranslated region (UTR). All four Ago proteins, as part of RISC, lead to the repression of the translation of the mRNA target (Figure 1.7E) (Su et al. 2009). Additionally, Ago2 possesses an endoribonuclease activity from its PIWI domain that can catalyse the degradation of mRNA targets when the miRNA complementarity is substantial (Meister et al. 2004).

1.4.2.4 Mechanism of miRNA-mediated repression

The mechanism of miRNA-mediated repression is currently under debate. Many studies have highlighted a block on translation initiation, meaning the ribosome does not fully assemble on an mRNA (Mathonnet et al. 2007; Meijer et al. 2013). Other work has suggested that the miRNA mode of action is instead at the level of translation elongation, where ribosomes are stalled whilst actively translating (Maroney et al. 2006). In either case, the repressed mRNA is either stored within a cellular body known as processing (P-) bodies or is deadenylated and degraded. It is entirely possible that both mechanisms may exist to some degree, meaning certain messages are repressed at initiation and either degraded or stored, whereas stalled messages may be taken somewhere else or left at the ER.

Regardless of which mechanism is correct or whether both are, the proteins involved in executing miRNA-mediated repression are more well established. The core of the RISC complex consists of a miRNA-loaded Ago and a TNRC6 (Trinucleotide repeat-containing gene 6; a.k.a. GW182) protein. There are three functionally redundant TNRC6 proteins – TNRC6A, B, C – which critical for miRNA function (Baillat and Shiekhattar 2009). TNR6CA-C act to recruit further downstream effectors, including those involved in de-capping (DCP1 and DCP2) and deadenylation (CNOT1) (Behm-Ansmant et al. 2006).

The RISC complex, consisting minimally of Ago and the mature miRNA, finds its target mRNA most likely by a scanning mechanism, as suggested by kinetic analyses (Chandradoss et al. 2015). Once found, RISC, through TNRC6, recruits the various repressive machinery leading to the negative regulation of translation (Figure 1.7E) (Wilczynska and Bushell 2015).

1.4.3 Non-canonical roles of Drosha and Dicer

Several miRNA-independent roles have been described for Drosha and Dicer. Drosha has been implicated in both transcriptional activation (Gromak et al. 2013) and termination (Dhir et al. 2015). Furthermore, Drosha has been demonstrated to regulate transcript levels of its cofactor DGCR8, amongst other genes, by binding and cleaving hairpin structures in untranslated regions of these transcripts (Han et al. 2009; Knuckles et al. 2012). Dicer was shown to dice dsRNA in the nucleus, to prevent its accumulation from converging transcription events (E. White et al. 2014). Both of these have also been demonstrated to participate in DNA repair (discussed below, section 1.5.2).

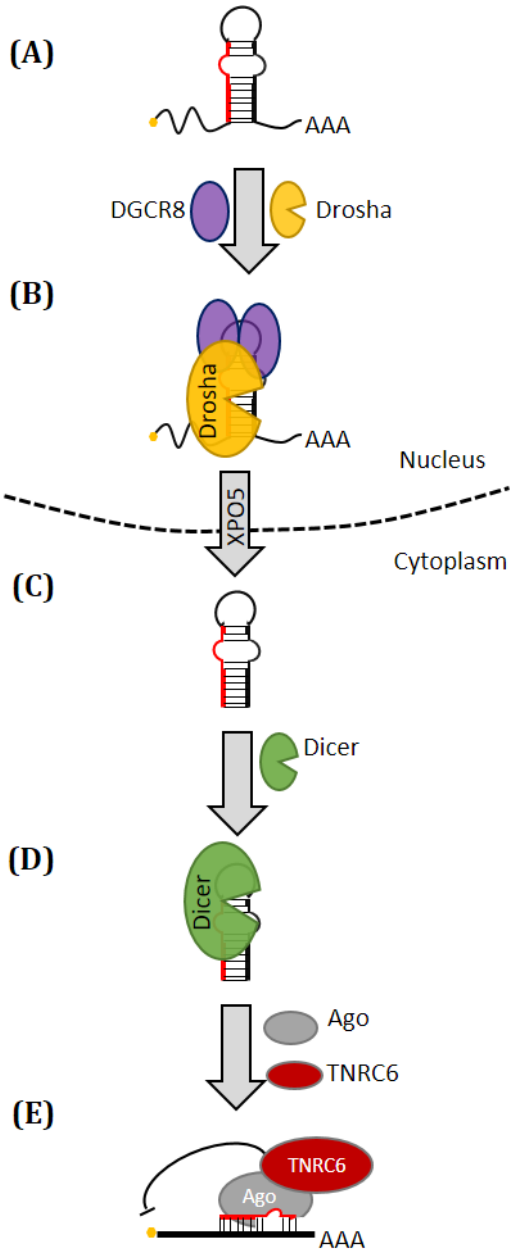


Figure 1.7. The microRNA pathway.

(A) The primary miRNA is a long precursor with a hairpin structure containing the mature miRNA sequence. **(B)** Drosha cleaves the stem of the pri-miRNA, measuring based on interactions with the basal ssRNA. A DGCR8 dimer orientates Drosha and increases affinity for pri-miRNA. **(C)** The resulting precursor miRNA hairpin is exported to the cytoplasm by XPO5. **(D)** Dicer cleaves the stem, removing the apical loop. **(E)** One strand of the miRNA duplex is incorporated into an Ago protein. TNRC6 is an adaptor that couples the miRNA-loaded Ago to the repression complexes to mediate miRNA effect.

1.5 RNA meets DNA repair

Over the past decade there have been a number of high-throughput proteomic studies looking at proteins involved in many aspects of DNA damage repair (Matsuoka et al. 2007; Paulsen et al. 2009; Hurov, Cotta-Ramusino, and Elledge 2010; Slabicki et al. 2010; Cotta-Ramusino et al. 2011; Adamson et al. 2012). These large siRNA screens (1.6-2.1x10^4 siRNA) have multiple caveats: namely the lack of knockdown validation for each gene means there can be many false positive/negatives and missing of factors altogether due to siRNA specificity and off-target effects; of two studies by independent groups using the same HR reporter, only 7 enriched genes overlapped between them (Slabicki et al. 2010; Adamson et al. 2012). However, in all six of these studies, RNA processing and binding proteins were among the top enriched groups. Furthermore, validations confirmed a direct role in DNA repair for many of these RNA factors, by either recruitment to foci and laser stripes or using repair reporters. The overarching theme from these studies is that RNA may be a critical component of DNA repair, a concept that is now gaining traction in the literature.

1.5.1 Transcription status of a locus affects its repair

A series of recent publications from the Legube lab has further fuelled the idea that RNA participates in DNA repair. This group established a system that allows the induction of ~100 DSBs in the endogenous genome via the endonuclease AsiSI, allowing repair dynamics to be researched at the nucleotide level without the need of artificial reporter systems (Iacovoni et al. 2010). Further characterisation of this system by chromatin immunoprecipitation following by sequencing (ChIP-seq) revealed that certain loci preferentially recruit RAD51 or XRCC4, indicating a preference for HR or NHEJ respectively (Aymard et al. 2014). Fascinatingly, the HR-prone sites were shown to be cut sites within transcriptionally active parts of the genome; conversely, NHEJ-prone sites were deficient in H3K36me3, a marker of active transcription. Critically, transcriptional activation of the GBP1 gene by IFN-γ switched the repair of a cut within that locus from NHEJ to HR (Aymard et al. 2014). A similar experiment was performed a year earlier by an independent group, where a cut within a doxycycline-inducible reporter construct switched association of the locus from 53BP1-bound to BRCA1-bound (i.e. NHEJ to HR) when transcription was turned on before damage induction (Tang et al. 2013). These data clearly implicate the prior

transcriptional status of a damaged locus in repair pathway decision and, in particular, that transcribed loci are preferentially repaired by the error-free pathway.

A follow-up study by the same group utilised a method known as high-throughput chromosomal conformation capture (Hi-C; described in (Belton et al. 2012)) which provides data on genome-wide chromatin organisation, in particular those sequences that associate in 3-dimensional space. Hi-C in the AsiSI system found that those transcriptionally active, HR-prone cut sites clustered together during G1 phase of the cell cycle when HR cannot take place (see section 1.2.6) and that this is an active process (Aymard et al. 2017). The authors suggest that this process prevents NHEJ and retains the DSBs for error-free repair in G2 phase. This data further alludes to a role for active transcription in DNA repair.

1.5.2 Does RNA participate in DNA repair?

Previous research of RNA in DNA repair has typically focused on transcriptional and post-transcriptional network changes because of the DNA damage stimulus, a famous example of which is the transcriptional activation of growth arrest or apoptosis genes by p53 (Meek 2004). However, most of these changes occur at a much slower rate than the DNA damage response, and as such are only relevant to hard-to-fix DSBs or cells in a particularly damage-prone environment. Some gene expression network changes can occur faster; for example, recent data demonstrated that pools of dephosphorylated and thus inactive miR-34 can be phosphorylated in response to DNA damage to help control cell cycle arrest (Salzman et al. 2016). Nevertheless, all of these examples still occur on the time scale of hours post-damage, clearly indicating that the role for RNA alluded to above is not due to gene expression changes.

More direct roles for RNA in DNA damage have been published. For example, 53BP1 was shown to require RNA for its association with chromatin (Pryde 2005). Immunoprecipitation of 53BP1 and analysis of those RNA species revealed these were gene products. Very recently, UV-C irradiation was shown to induce very rapid N6-methyladenosine (m6A) modifications of RNA at DNA damage sites and this was required for recruitment of DNA polymerase κ (NER polymerase, section 1.2.8.3) to those loci to aid in repair (Xiang et al. 2017). These data show a role for existing RNA as a scaffold or to participate in repair signalling at break sites. In contrast, several publications have suggested that RNA may be synthesised as a result of DNA damage

around those break sites to participate in the DDR (Hawley et al. 2017). This was first observed in *Neurospora crassa*, where the addition of replication stalling agents such as histidine cause DNA damage. The addition of histidine for 40 hours led to the identification of ~21 nt small RNA associated with the Argonaute homologue QDE-2, originating mostly from the rDNA repeat locus (Lee et al. 2009). These required an RNA-dependent RNA polymerase (RdRP) and a Dicer-like protein (DCL); the loss of which increased sensitivity to damage. The hypothesised model for these damage-induced small RNAs was to reduce aberrant RNA (aRNA) expression, however how that may be involved in the damage response *per se* was unclear. Later publications also reported the existence of *de novo* products. In *Drosophila*, the transfection of linearised but not circularised plasmids was found to repress corresponding reporter transcripts *in trans*, and sequencing revealed the production of small RNA (Michalik, Böttcher, and Förstemann 2012). Here again, it was proposed that these "endo-siRNA" would degrade truncated and aberrant transcripts. In *Arabidopsis*, RdRP-dependent small RNA were shown to be produced around breaks in a transgene reporter system (W. Wei et al. 2012). The same group also showed that small RNA could be produced from a GFP repair reporter system in human cells (W. Wei et al. 2012). Additionally, small RNA were observed in human cells around a highly repetitive but transcriptionally inactive integrated reporter locus (Francia et al. 2012). A recent follow-up from this group reported the production of small RNA from the $(TTAGGG)_n$ telomeric repeat locus (Rossiello et al. 2017). These later publications propose a model where *de novo* transcribed small RNA products are used to direct repair factors to damage sites; the exact point in the repair response at which this occurs is disputed however (Francia et al. 2012; Gao et al. 2014).

1.5.3 Transcription is locally inhibited on DNA damage

Whilst it is clear transcription has some role in DNA repair, it is also very well-established that transcription is inhibited following DNA damage induction. Many publications have demonstrated this phenomenon, showing that it is an active process orchestrated by the early DDR signalling proteins. ATM was found to inhibit both RNA Pol I and Pol II when DSBs are induced within their genes (Kruhlak et al. 2007; Shanbhag et al. 2010). Furthermore, DNA-PK was also shown to repress Pol II elongation; interestingly, by inhibiting DNA-PK it was also demonstrated that Pol II

could actually circumvent a DSB to allow transcription to continue past a break (Pankotai et al. 2012). This was also dependent on Ku binding, indicating end synapsing (see section 1.2.5) allows a polymerase to skip over a gap in DNA. As expected, resection around a DSB also inhibits transcription, as resection proceeds 5'-3' (i.e., towards a TSS) (Manfrini et al. 2015). Somewhat less obvious however was the observation that transcription was also repressed on the other strand (where the TSS would begin on the 5' end of a break) suggesting transcription is inhibited *in vivo* at ssDNA.

The chromatin architecture following DNA damage induction also inhibits transcription. The phosphorylation of H2AX (γH2AX) is an important modification that occurs very early at all DSBs (see section 1.2.3). γH2AX foci formation was shown to coincide with reduced transcription at those foci, which was later reaffirmed by ChIP experiments (Solovjeva et al. 2007; Iacovoni et al. 2010). Various repressive complexes are also recruited to DSBs, such as the NuRD (nucleosome remodelling deacetylase) complex, through PARP and ZMYND8 (Chou et al. 2010; Gong et al. 2015).

Finally, a recent study utilised the AsiSI system to investigate transcription genome-wide following DNA damage (Iannelli et al. 2017). This confirms the previously discussed studies that in all instances, transcription is inhibited locally (<1 Mb around DSBs) following DNA damage induction and that this is ATM signalling dependent, and the presence of γH2AX prevents Pol II association.

1.5.4 The shadow genome: Can RNA be a template for DNA repair?

The observations that RNA binding and processing proteins participate in DNA repair, and that RNA in many ways has been implicated suggest an RNA-mediated mechanism may exist. However, the vast quantity of literature showing that transcription is turned off following damage induction would suggest that instead of *de novo* transcription, there may be a role for the pre-existing transcripts, a concept backed up from the work out of the Legube lab. The most exciting possibility is that the RNA itself may be utilised as a template for the repair of its damaged gene. This activity would require an RNA-dependent DNA polymerase (RdDP; a.k.a. reverse transcriptase, RT) which, outside of telomere elongation, is not known to exist in eukaryotes (Yamaguchi et al. 2005). Nevertheless, there have been some compelling pieces of evidence of the past

decade that has suggested that an RNA-mediated recombination pathway may exist in eukaryotes.

In yeast, it was found that transformation with RNA oligonucleotides homologous to both sides of an inducible DSB increased precise repair efficiency greater than 500-fold compared to no oligonucleotides (Storici et al. 2007). Whilst this was almost 40,000-fold less efficient than the use of a comparable DNA oligonucleotide, it was the first direct evidence that RNA could be used at all. A follow up paper later confirmed this mechanism in yeast, and crucially showed that endogenous transcript RNA could be utilised as a template for recombination repair (Keskin et al. 2014). This mechanism depended on the RNA to hybridise with the duplex DNA *in cis*, as the frequency of RNA-templated repair events was increased in RNase H-deficient strain where DNA:RNA turnover is reduced. The observation that this DNA:RNA strand exchange could be catalysed by the recombinase Rad52 also suggested this may be an active process. The same group later verified this activity of Rad52 *in vitro,* showing that this is a novel inverse strand exchange mechanism where dsDNA is first bound by Rad52, then RNA is brought in to invade the duplex (Mazina et al. 2017).

This work also showed that the human RAD52 could carry out DNA:RNA inverse strand exchange, suggesting this mechanism may occur in human cells. To date, there are two publications that allude to a RNA-templated repair role in humans. The first utilised an integrated reporter system that could have transcription activated or inactivated and damage induced (L. Wei et al. 2015). This confirmed that DNA damage inhibits transcription locally, and further validated the previously discussed findings that transcriptionally active loci preferentially recruit HR proteins. Importantly, they demonstrated that RAD52 was recruited and this was dependent on active transcription and RNA molecules; this was also possible in G0/G1 phase of the cell cycle when resection and HR is canonically inhibited. The second study showed that classical NHEJ factors may also participate in an RNA-templated repair pathway (Chakraborty et al. 2016). However, both publications (as well as the discussed yeast experiments) relied heavily on the use of integrated reporter systems. Much more work needs to be done to determine whether this mechanism can occur as a general feature of DNA damage at endogenous loci. Nevertheless, these many pieces of evidence point towards a role for active transcription, nascent RNA, and RNA binding

and processing factors in DNA repair. Whether all these pieces are connected, or represent parts of many ways in which RNA can be involved in DNA repair, is yet to be fully understood.

1.6 Project aims

The overall aim of this PhD project was to further our understanding of the role of RNA in DNA damage repair. Originally, this was to investigate the involvement of the miRNA biogenesis machinery and address the outstanding questions of damage-induced small RNA as mentioned at the end of section 1.5.2.

Specific aims:

I. To determine which components of the miRNA biogenesis machinery are involved in DNA repair outside of their roles in the miRNA-mediated repression pathway.
II. To ascertain at which point in the repair pathway those factors are involved.
III. Investigate whether damage-induced small RNAs are a general feature of DSBs within endogenous genomic loci, and to research their biogenesis and function.

2 Chapter 2: Materials and methods

2.1 Materials & methods

2.1.1 *In vivo*-based methods

2.1.1.1 General tissue culture

U2OS cells (ATCC, HTB-96) and A549 cells (ATCC, CCL-185) were maintained in Dulbecco's Modified Eagle Medium (DMEM) (Gibco, 41966) supplemented with 10% v/v foetal bovine serum (FBS) and 2mM L-glutamine (Gibco, 25030081). U2OS cells stably expressing HA-ER-AsiSI (U2OS-AsiSI) were a gift from Gaelle Legube at the University of Toulouse, France and are described in (Iacovoni et al. 2010). U2OS-AsiSI were sustained in DMEM with 10% v/v charcoal-stripped FBS to reduce the presence of oestrogen analogues that may activate the ER-tagged endonuclease, plus 2mM L-glutamine and 1 µg/ml puromycin (Invivogen, ant-pr-1) to maintain selection. U2OS-HR and U2OS-NHEJ cells (described in sections 2.1.5) were sustained in DMEM with the addition of 1 mg/ml G418 (Sigma, A1720) to maintain selection. All cells were cultured at 37 °C with 5% CO_2. Early passage cell stocks were frozen in a 90% FBS 10% DMSO (Sigma, D2438) and kept in liquid nitrogen storage. The clonal U2OS cells were typically cultured for around 10 passages before a new stock was thawed. All media and reagents were typically pre-warmed to 37 °C for 30 minutes before addition to live cells.

2.1.1.2 Transfection of siRNA

siRNA reverse transfections were performed using Dharmafect 4 reagent for U2OS cells and Dharmafect 1 for A549 cells (Dharmacon, T-2004 & T-2001). Cells were grown to achieve 70-80% confluency on day of transfection, then detached from flasks by trypsin (Gibco, 15090046), centrifuged, and resuspended in fresh DMEM (for clonal U2OS, no selection agent was used). Meanwhile, the appropriate transfection reagent was diluted (1 µl/24-well, 2 µl/12-well, 4 µl/6-well) in Opti-MEM (Gibco, 11058021) and incubated for 5 minutes at room temperature. siRNA was similarly diluted in Opti-MEM to a 30 µM final concentration. Dharmafect solution was then added to siRNA solution for 20 minutes at room temperature and added to a fresh plate well. Resuspended cells were counted during incubations and plated on top of the reagent to reach experimental confluency 48 hours' time, unless otherwise stated.

The siRNA used in this thesis are found in Table 8-1.

2.1.1.3 Transfection of plasmids

DNA plasmids were typically forward transfected into U2OS cells using Dharmafect DUO reagent (Dharmacon, T-2010). The transfection reaction protocol was the same as that of siRNA by Dharmafect 1/4 (section 2.1.1.2), however the final solution was added dropwise onto pre-plated cells.

For the transfection of pCI-AsiSI plasmid (section 4.3) electroporation was used to ensure reproducible and maximal transfection efficiencies. U2OS cells were electroporated using the Amaxa Cell Line Nucleofector Kit V solution and following the manufacturer's U2OS optimised protocol (Lonza, VCA-1003). Briefly, $1x10^6$ U2OS cells were resuspended in 100 µl solution V with 2 µg plasmid DNA. Electroporation was performed using programme X-001 in a Nucleofector II device. Cells were then pooled across multiple identical reactions and plated to reach experimental confluency 16 hours later.

A list of plasmids is found below in table Table 8-3.

2.1.1.4 Generation of chromosomally integrated cell lines

For repair reporter assays in section 2.1.5 cell lines were generated by prolonged selection of single cells. HR and NHEJ reporters as described in (Seluanov, Mao, and Gorbunova 2010) were digested *in vitro* by NheI-HF (NEB, R3131) in 1X CutSmart buffer (NEB, B7204) for 6 hours at 37°C in a thermocycler. NheI restriction sites are located at the beginning and end of the reporter/neomycin resistance cassette meaning integrations are more likely to occur as intended. Digested fragments were gel extracted from 1% agarose gels and subject to an additional purification using a QIAquick PCR purification kit (Qiagen, 28106) before forward transfection into U2OS cells. Cells were grown for 24 hours before sparse splitting and selection in 1 mg/ml G418 for 1 week. Individual cells were then plated in 96-well plates and supplemented with 1:1 conditioned U2OS media to fresh media with 1 mg/ml G418. When wells reached confluency, they were split into progressively larger well formats until enough cells could be frozen. Cell lines were validated for correct integration of the repair reporter by transfection of the I-SceI plasmid and measuring GFP induction by flow cytometry (see section 2.1.5 for reporter assay protocol). The top candidate cell lines

were then further validated by positive control conditions; the cell line for each with the greatest response was then used in all further experiments.

2.1.2 Induction of DNA damage

Two main methods were used to induce DNA damage in this thesis: ionising radiation (IR) in A549 cells and AsiSI-induced DSBs.

An Xstrahl RS320 set to 195 kV/10 mA was used to irradiate A549 cells to a dose of 5 Gy IR. Cells were briefly outside standard incubation conditions during this time, so this was minimised as much as possible. A timer was set immediately following completion of the exposure. After 2 or 6 hours depending on the experiment, cells were placed on ice and washed once with cold PBS for harvesting of protein, RNA, or fixing of coverslips for IF as appropriate.

For induction of damage in U2OS-AsiSI cells, fresh charcoal-stripped DMEM was prepared with either 300 nM final 4-OHT (Sigma, T176) or an equal volume of DMSO for undamaged samples. This media was used to replace the existing cell media and cells returned to incubation. A timer was set following addition of 4-OHT to all cells. Following the required length of time, cells were processed as above as required.

2.1.3 Immunofluorescence

Preparation of slides for immunofluorescence (IF) studies was carried out utilising a pre-extraction method that increases resolution of DNA damage foci (Britton, Coates, and Jackson 2013). Following damage induction in cells seeded upon glass coverslips, cells were washed in PBS and pre-extracted using cytoskeleton (CSK) buffer (100 mM NaCl, 300 mM sucrose, 3mM $MgCl_3$, 10 mM PIPES pH 6.8 (Melford, B2004), 10 mM β-glycerol phosphate, 50 mM NaF, 1 mM EGTA, 5 mM sodium orthovanadate, 0.5% v/v Triton X-100 (all chemicals from Sigma unless otherwise specified)). CSK buffer was added for 2 minutes, followed by an additional 1 minute incubation in fresh CSK buffer and 10 minute fixation in 4% v/v paraformaldehyde in PBS (Alfa Aesar, 43368). Coverslips were washed in PBS and blocked in 10% v/v goat serum (Sigma, G9023) in PBS for 1 hour at room temperature, or for longer than 1 hour at 4°C.

Incubation with primary antibodies (see Table 8-2 for all antibody information) was carried out overnight at 4 °C. After several PBS washes, fluorescent secondary antibodies were added for 1 hour at room temperature. Coverslips were mounted on glass slides in Vectashield mounting medium with DAPI (Vector Labs, H1200) following

additional PBS washes. All slides were blinded by another member of the lab to ensure all image acquisition and data analysis was free from any biasing.

Image acquisition was performed using a Zeiss LSM501 confocal microscope with 63X objectives. Focusing was performed using the γH2AX-stained channel (typically 546 nm/red channel) to obtain clear and distinct foci. Laser power, gain, and digital offset settings were calibrated to no damage controls per experimental set. Images were acquired with a Plan-Apochromat 63X/1.4 Oil DIC lens with a pinhole size set to 1 Airy Unit for each laser. Image size was set to 1020x1024 pixels with a dwell of 1.26 µs/pixel; each line was scanned 4 times and averaged to reduce noise. For IF, at least 70 cells were counted per condition per experiment. Image quantification was performed by scoring foci positive cells by manual counting a cell was determined to be foci positive if there were greater than a threshold number of foci; this was dependent on the protein and is within the figure legends. Samples were only unblended after quantification was complete.

2.1.4 Cell cycle analysis by PI staining

Cells were trypsinised, centrifuged and resuspended in PBS (200 µl/1x10^6 cells). Ice cold 70% ethanol/30% PBS was added quickly whilst centrifuging and then left at least overnight at 4 °C to fix the cells. Cells were pelleted and resuspended in PBS; this was repeated another time as washes. After the second pelleting, cells were resuspended in PBS containing 10 µg/ml propidium iodide (PI) (Thermo, P1304MP) and 0.1 mg/ml RNase A (Sigma, R4875) and incubated at 37 °C for 30 minutes and 4 °C overnight. Analysis was performed using a BD FACSCanto II controlled by FACSDiva software (v8.0.1, BD). Debris and any large clumps of cells were removed from analysis by first measuring side and forward scatter and gating only the single cells. PI incorporation was measured by excitation with a 488 nM laser and collection of emission at 585 nM. Cell cycle populations determined by gating the first peak (DNA content *n*) corresponding to G1 and the second peak (DNA content 2*n*) corresponding to G2. S phase was determined as the region between these peaks (DNA content $1 < n < 2$).

2.1.5 HR and NHEJ repair reporter assays

Each reporter is cleaved by the expression of I-SceI endonuclease; the subsequent repair by the intended pathway results in reconstruction of a functional GFP gene and this can be measured by flow cytometry. Reporter cell lines were reverse transfected

with 30 nM siRNAs and seeded into 6-well format for 48 hours before transfection of 2 µg I-SceI plasmid to ensure complete ablation of the proteins of interest before any activation of the reporter. Media was replaced the day after each transfection. For RNase H1 over-expression experiments (section 5.7), cells were forward transfected with RNase H1 plasmid followed by forward transfection of I-SceI 6 hours later (Table 8-3). 48 hours following transfection of I-SceI, cells were trypsinised and resuspended in 800 µl phenol red-free DMEM (Gibco, 31053).

Control cells with no transfection and with over-expression of GFP were plated in parallel and used to optimise gating conditions on a FACSCanto II flow cytometer (BD). Non-transfected cells were used to gate single cells based on side and forward scatter, and to set the GFP-positive gate by excluding non-GFP green auto-fluorescence. GFP over-expression allowed optimisation of green-red compensation percentages and to confirm gating of GFP vs. non-GFP cells. At least $3x10^4$ (typically $5x10^4$) total cells were counted per condition. The proportion of GFP-positive cells in each condition was normalized to that of the control condition following subtraction of the proportion in the non-transfected control (Repair efficiency = $(GFP_{condition} – GFP_{null})/(GFP_{control} – GFP_{null})$). The mean of at least three biological replicates was plotted as repair efficiency where values lower than control have diminished HR or NHEJ repair and larger values have increased efficiency. Statistical significance was determined using the Wilcoxon signed rank non-parametric test comparing the mean normalised values to the control value of 1.0.

2.1.6 Laser microirradiation

$4.5x10^4$ U2OS cells were seeded into Lab-Tek chambered coverglass wells (Thermo Scientific, #155383). These plates have confocal slide glass as their base, and are designed to culture live cells for microscopy. The following day, 200 ng mCherry-RNase $H1^{mut}$ or an mCherry control plasmid was forward transfected into each well (Table 8-3). 24 hours later, media was replaced with phenol-red free DMEM following a PBS wash. Prior to microirradiation of a sample, 20 µM Hoescht 33342 (Thermo Scientific #62249) was added to pre-sensitise the cells to 532 nm laser damage. Cells were maintained at 37 °C and 5% CO_2 for the duration of the experiment in a confocal stage top environmental control system (okolab). Image acquisition was performed using a spinning disk confocal system (3i) with a 100X objective. mCherry fluorescence was

visualised with the 561 nm laser channel with 500 ms exposure at full power, and this was captured with a CMOS camera (Hamamatsu) at 2048x2048 pixels (0.124 µm/pixel) with 2x2 binning to improve the signal to noise ratio. Laser track lines were drawn across mCherry-positive nuclei using the 3i Slidebook software. Images in the 561 nm channel were taken every 500 ms and microirradiation was induced along the stripe after the fifth timepoint using a 532 nm pulsed laser (3i Ablate! Model 3iL13) at 10% power. Autofocus using the Definite Focus.2 channel (Zeiss) was utilised every fifth timepoint to account for any focal drift during the time-lapse.

To quantify relocation to the laser stripe, intensity along the laser stripe was normalised to the background intensity of a second undamaged stripe within the same nucleus and the first timepoint set to 1.

2.2 RNA methods

2.2.1 RNA extraction

Total RNA was harvested from cells using TRIzol (Invitrogen, 15596018) with a slightly adapted protocol to better enrich small RNA. Cultured cells were washed with PBS and 1 ml TRIzol added directly to the well. Cells were scraped in TRIzol which was then moved to a standard Eppendorf tube to homogenise for 5 minutes at room temperature. Per 1 ml TRIzol, 200 µl chloroform (Honeywell, C2432) was added, shaken, and centrifuged at 11,000x*g* for 15 minutes and 4 °C. The resulting RNA-containing upper aqueous phase was moved to a fresh tube. RNA was precipitated overnight at -20 °C following the addition of 750 µl isopropanol (Fisher, P/7500/17) and 20 µg glycogen (Invitrogen, 10814010). The next day, this was centrifuged at 17,000x*g* at 4 °C for 30 minutes and the resulting pellet washed with 75% ethanol (Fisher, E/0600DF/17). To dry the pellet, all liquid was aspirated followed by a brief centrifugation to collect any remaining liquid which was then also removed. As this allows the removal of nearly all ethanol, air drying was only needed for 1 minute. RNA pellets were resuspended in 100 µl water for further purification by acid phenol:chloroform extraction. 100 µl acid phenol:chloroform (Ambion, AM9722) was added, followed by shaking and centrifugation at 11,000x*g* for 10 minutes at 4°C. The upper aqueous phase was transferred to a fresh tube, and the remaining lower phases were back-extracted by the addition of 100 µl water and repeating the centrifugation

steps. This aqueous phase was added to the previous to maximise RNA recovery. A final overnight precipitation was carried out by adding 20 µg glycogen, 0.1 volumes 3 M sodium acetate pH 5.2 (Sigma, S2889), and 2.5 volumes 100% ethanol. Washing and drying of the pellet was performed as before, and then resuspended in water. RNA concentration and quality was quantified using a NanoDrop 2000 spectrophotometer (Thermo Scientific), aiming for a 260/280 nm ratio >1.8 and a 260/230 nm ratio >2.

2.2.2 TaqMan qPCR assay

For assessment of miRNA abundance, TaqMan real time PCR small RNA assays were used. Total RNA was isolated as above. 100 ng total RNA was reverse transcribed in a 15 µl reaction using the TaqMan miRNA reverse transcription (RT) kit (Applied Biosystems, 4366597) and either hsa-miR-21 RT primer (TaqMan Assay ID 000397) or the control snRNA U6 RT primer (TaqMan Assay ID 001973). An additional reaction with water instead of RNA was also set up as a no template control (NTC). RT reaction was performed in a thermocycler programmed to incubate at 16 °C for 30 minutes, 42 °C for 30 minutes, 85 °C for 5 minutes. Following RT, cDNA was diluted 5-fold in water. For the qPCR reaction, 5 µl of diluted cDNA used in a 20 µl reaction containing TaqMan Universal PCR master mix (Applied Biosystems, 4367846) and 1 µl of the appropriate gene-specific assay probe as supplied alongside the RT primer. The reaction was carried out in technical triplicates on an Applied Biosystems 7500 Fast RT-PCR system using the built-in TaqMan assay reaction settings (95 °C 10 minutes; 40 cycles of 95 °C 15 seconds, 60 °C 1 minute). miR-21 abundance was represented as ΔC_T relative to the paired snRNA U6: values were calculated by subtracting the mean paired U6 C_T values from the mean miR-21 C_T values for each condition. These abundances were then normalised across conditions to obtain a $\Delta\Delta C_T$ by subtracting the control siRNA ΔC_T from each knockdown sample; this logarithmic value was then plotted as $2^{-\Delta\Delta CT}$.

2.2.3 5’-triphosphate small RNA isolation

The method to isolate 5’-triphosphate RNA was developed in section 4.6; the final method used on samples for library preparation is described below. Total RNA was first depleted of 5’-monophosphate RNA (miRNA, rRNA) by Terminator 5’-phosphate-dependent exonuclease (epicentre, TER51020). 5 µg total RNA was incubated at 30 °C for 30 minutes in a 20 µl reaction containing 1 U Terminator exonuclease, 1X Terminator reaction buffer A, and 0.5 µl RNasin Plus (Promega,

N261B). RNA was purified by acid phenol:chloroform extraction and overnight ethanol precipitation as in section 2.2.1.

To make 5'-triphosphate small RNA amenable to standard small RNA library preparation, Terminator-treated RNA was then treated with 5'-polyphosphatase (Epicenter, RP8092H) to remove the γ and β phosphates, leaving them as 5'-monophosphates. The reaction containing 20 U 5'-polyphosphatase, 1X supplied buffer, and 0.5 µl RNasin was incubated at 37 °C for 30 minutes followed by acid phenol:chloroform extraction, ethanol precipitation, and resuspension in nuclease-free water.

2.2.4 Small RNA sequencing library preparation

For the timecourse small RNA-seq experiment in section 4.3, total RNA was prepared using the TruSeq small RNA library preparation kit (Illumina, RS-200-0012) following the manufacturer protocol, with 1 µg total RNA used as input per condition. Briefly, 5' and 3' adapters were ligated onto small RNA, which were then reverse transcribed using a primer against a sequence in the 3' adapter. This cDNA was PCR amplified by 13 cycles using a universal primer and a primer containing a unique barcode for each sample.

For the first small RNA-seq experiment (section 4.2) and the 5'-triphosphate small RNA-seq (section 4.6), an adapted, non-commercial protocol was used (Xu et al. 2015). All adapters and primers were ordered from Sigma with PAGE purification. 600 ng total RNA per condition was first ligated to 500 nM pre-adenylated 3' adapter in a reaction containing 400 U T4 RNA ligase 2 (truncated K227Q; NEB, M0351), 0.75 µl RNasin Plus, 10% PEG-8000 (NEB), and 1X T4 ligase 2 buffer for 2 hours at 26 °C. Following 3' adapter ligation, an RNA cleaning and concentration kit (Zymo, R1015) was used to remove the ligation components, followed by removal of excess 3' adapter by deadenylation following by degradation. The deadenylation reaction was carried out at 30 °C for 30 minutes with 10 U 5' adenylase (epicentre, DA1101K) in 1X supplied buffer with the addition of 5 mM DTT and 0.5 µl RNasin. This was quenched by the addition of EDTA (5 mM final). As the 3' adapter is DNA, this was selectively decayed by 10 U RecJ exonuclease (Epicentre, RJ411250) in 33.3 mM Tris-HCl pH 9.0 and 12 mM $MgCl_3$ at 37 °C for 30 minutes. The 5' adapter was then ligated using 10 U T4 RNA ligase (Epicentre, LR5010) in 0.25X supplied buffer, 50% PEG-8000, and 250 µM ATP for 2

hours at 26 °C. This was again cleaned using an RNA cleaning and concentration kit. The 5' and 3' adapter ligated RNA was then reverse transcribed at 37 °C in a 40 µl reaction containing 200 U MMLV RT (Epicentre, RT80110K) in 1X supplied buffer, 500 µM dNTPs, 500 µM DTT, and 500 nM RT primer. 20 minutes later the reaction was terminated by incubating at 85 °C for 15 minutes. This was then PCR amplified to add unique indexes for each sample. A 20 µl PCR reaction contained 1 U Phusion-HF DNA polymerase (Thermo, F-530) in 1X supplied buffer, 250 µM dNTPs, 500 nM RP-1 primer, 500 nM index primer, and 4 µl cDNA from the previous step; this was run for 13 cycles of [10 seconds 98 °C, 30 seconds 60 °C, 15 seconds 72 °C] following an initial 30 seconds 98 °C denaturation and ended with a 10 minute final extension at 72 °C. For both small RNA library preparation methods, indexed and PCR amplified libraries were run on TBE-polyacrylamide gels for purification and size selection. The band at ~150 bp (corresponding to 21-23 nt RNA ligated to the two adapters and additional index/RP-1 PCR primer sequences) was excised and fragmented by centrifuging through a gel breaker tube (0.5 ml Eppendorf with 21 G needle pierced holes). Libraries were eluted in a 50 mM NaCl, 10 mM Tris-HCl pH 8.0 buffer overnight at 4 °C. Library concentrations were determined by Bioanalyzer DNA high sensitivity chip and pooled at equimolar concentrations, then ethanol precipitated overnight and resuspended in 11 µl. 1 µl of this was used for a final TBE-polyacrylamide gel or Bioanalyzer chip to ensure adequate library concentration was achieved for NGS (≥10 nM in 10 µl).

2.2.5 4-thiouridine enrichment of nascent RNA

For metabolic labelling of nascent transcripts (Rädle et al. 2013), $3x10^6$ cells were first seeded into six 15 cm plates (19 ml media) and reverse transfected with either control or Drosha siRNA (30 nM final). 48 hours later, 14 ml media was aspirated and 9 ml of this discarded. The remaining 5 ml was added to a tube containing 50 µl of 100 mM 4sU and 100 µl of either DMSO or 100 µM 4-OHT to mix and added back on cells; plates were then returned to the incubator for 1 hour. This was scaled per the three plates per condition and staggered by 15 minutes per condition to ensure the 1 hour timepoint was strictly adhered to. To harvest, cells were placed on ice and PBS washed once in cold PBS. Cells were scraped quickly in cold PBS and transferred to a falcon tube. This was centrifuged at 600x*g* for 5 minutes at 4 °C to pellet cells. PBS was

aspirated and pellet resuspended in TRIzol (2 ml for three plates, ~40x10^6 cells). RNA was harvested following the protocol as above (section 2.2.1), with the omission of the second optional round of purification. Total RNA was subject to a biotinylation reaction to covalently link biotin to the thiol moiety of 4sU nucleotides. Per 1 µg RNA, 2 µg EZ-Link biotin-HPDP (Thermo, 21341) was used in a 10 µl reaction containing 10 mM Tris pH 7.4 and 1 mM EDTA. RNA was denatured at 70 °C for 5 minutes prior to reaction, which was carried out at room temperature for 90 minutes with rotation. Excess biotin was removed by double chloroform extraction, and RNA was ethanol precipitated.

To enrich 4sU-containing RNA, streptavidin-coated magnetic beads (Dynabeads M-280; Thermo, 11205D) were used to affinity purify due to its strong affinity for biotin. These were prepared for RNA following the manufacturer protocol. 50 µl beads were used per 25 µg biotinylated RNA. RNA was denatured at 70 °C for 5 minutes and placed on ice. Beads were resuspended in an equal volume to the RNA in 2X BW buffer (10 mM Tris pH 7.5, 1 mM EDTA, 2 M NaCl). RNA and beads were incubated at room temperature for 15 minutes with agitation. Beads were washed 3 times with BW buffer. To elute, 200 µl 100 mM DTT was added to beads for 5 minutes to reduce the biotin-4sU disulphide bond. Beads were separated a final time and supernatant moved to a fresh tube. This RNA was acid phenol:chloroform extracted and ethanol precipitated. The pellet was resuspended in 20 µl nuclease-free water and RNA concentration quantified. Between 2.5-3.0 µg RNA was typically recovered from an initial 60-70 µg input. 2.5 µg was sent to the University of Cambridge Biochemistry department for ribosomal RNA depletion by Ribo-Zero (Illumina) and strand-specific library preparation using the TruSeq Stranded RNA preparation kit (Illumina). This was sequenced on an Illumina NextSeq500 for 2x75bp cycles on high-output mode, generating 8x10^8 reads.

2.3 DNA methods

2.3.1 Agarose gel

For a 1% w/v agarose gel (which is sufficient to separate DNA <10 kb), 1g agarose (Melford, MB1200) was added to 100 ml TAE and boiled in a microwave with regular swirling to mix. When dissolved, 10 µl SYBR Safe (Thermo, S33102) was added. This

was added to a gel cast and allowed to cool to set. Samples were loaded in loading buffer (5% glycerol, 0.04% bromophenol blue) alongside a DNA ladder (Hyperladder 1kb; Bioline, BIO-33053). Gels were run at 120 V in 1X TAE and visualised using a UV gel imager (GeneFlash).

2.3.2 TBE-polyacrylamide gel

TBE-PAGE was used to run small RNA libraries for size selection. For two 8% TBE-polyacrylamide gels, 4ml 40% polyacrylamide/Bis solution 19:1 (Bio-Rad, 161-0144), 2 ml 5X TBE, and 14 ml water was mixed and 200 µl APS and 12 µl APS was added to initiate polymerisation. All casting plates were extensively cleaned prior to use with IMS and RNaseZap (Thermo, AM9780). Samples were loaded with 1X Novex loading dye final (Thermo, LC6678) alongside 5 µl 20 bp DNA ladder (Jena Bioscience, M-212). TBE-PAGE was run for ~2.5 hours at 120V in 0.5X TBE until the upper xylene cyanol band (component of the loading dye) was at the bottom of the gel. Following electrophoresis, gels were washed in 0.5X TBE containing 1 µl SYBRGold (Thermo, S11494). DNA bands were then visualised on a Phosphorimager (Typhoon FLA 700, GE) using a glass stage. The resulting image was printed at 1:1 ratio and used as a template to excise gel fragments.

2.3.3 DNA:RNA IP using S9.6 antibody

The DNA:RNA IP to enrich for R-loop-prone genomic fragments was performed by Wei-Ting Lu (Appendix 8.1). U2OS-AsiSI cells were washed and lysed in a cytoplasmic lysis buffer (10 mM HEPES pH7.9, 10 mM KCl_2, 1.5 mM $MgCl_2$, 0.34 M sucrose, 0.1% Triton X-100, 10% glycerol, 1 mM DTT). The nuclear pellet was obtained by centrifugation at 1500 RCF. This was further refined to a chromatin fraction by incubating the nuclear pellet with a buffer containing 50 mM Tris pH 8.0, 5 mM EDTA, 1% SDS and Proteinase K at 50 °C for 2 hours, and centrifugation after addition of KOAc to the final concentration of 1 M. Genomic DNA containing R-loops was then precipitated overnight from the chromatin fraction by the addition of ethanol. Genomic DNA was washed with 70% ethanol, re-suspended in water.

For the second DRIP (section 5.4), this was then digested with AsiSI *in vitro* for 2 hours at 37 °C. The DNA was ethanol precipitated again, re-suspended in 400 µl IP buffer (16.7 mM Tris pH 8.0, 1.2 mM EDTA, 167 mM NaCl, 0.01% SDS, 1.1% Triton X-100) and sonicated for 10 minutes (Diagenode Bioruptor, Medium setting, 30s on/30s off

interval). Half of the genomic DNA was treated with RNase H1 *in vitro* (NEB, M0297S, 3 units per µg DNA input, 37 °C 3 hours) as for the negative control. For DRIP, 1 µg S9.6 antibody (Kerafast, ENH001) conjugated with ChIP grade protein A/G magnetic beads (Thermo Scientific, #26162) was used for every 2 µg genomic DNA input. The resulting IP was carried out overnight at 4 °C. The beads were washed once in IP buffer, once with IP buffer with 500 mM NaCl, and once with LiCl wash buffer (10 mM Tris pH8.0, 250 mM LiCl, 1 mM EDTA, 1% NP-40), and once with TE with 50 mM LiCl. The magnetic beads and their respective input genomic DNA, were re-suspended in 400 µl nuclease-free water, and treated with RNase 1 at 37 °C (Invitrogen, AM2295). The samples were then subjected to 2 rounds of Proteinase K treatment at 55 °C, and DNA extracted with phenol:chloroform:isoamylalcohol (pH 8.0, Sigma P2069), followed by ethanol precipitation.

2.3.4 SYBR Green qPCR

Quantitative PCR analysis was performed on DRIP-ed fragments and their respective inputs by Wei-Ting Lu (Appendix 8.1). This was performed using equal volumes on input and IP DNA to obtain a measure of enrichment. The 20 µl qPCR reaction contained 1X Fast SYBR Green master mix (Thermo, 4385612) and 300 nM each of forward and reverse primers (Table 8-4). This was set up in technical triplicate and run on an Applied Biosystems 7500 thermocycler with the cycling parameters: 95 °C 20 seconds, 40 cycles of [95 °C 3 seconds, 60 °C 30 seconds]. Mean C_T cycles across technical replicates for DRIP samples was normalised to the C_T for their respective inputs.

2.3.5 DRIP-seq library preparation

The DRIP-ed DNA fragments were used for library preparation using the NEBNext Ultra II kit and multiplex oligos (NEB, E7645S & E7335S) following the manufacturer protocol with slight adjustments described here. For the first DRIP-seq, 750 ng was used as input; the second DRIP used 1 ng. DRIP fragments were first end repaired in 60 µl reaction containing the supplied End Prep enzyme (3 µl) and master mix (7 µl); this was incubated at 20 °C for 30 minutes followed by 65 °C for another 30 minutes. This step makes the fragments blunt ended and adds an A overhang for adapter ligation. The supplied adapters are hairpins that require ligation to the end-prepared fragments. The supplied adapters were diluted 1:30 fold to reach a working concentration of 5

µM. 2.5 µl of this diluted adapter was added to the end prepared fragments with 30 µl supplied ligation master mix and 1 µl enhancer. Ligation reaction was carried out at 20 °C for 15 minutes. The hairpin adapters contain a uridine nucleoside in the apical loop. Incubation with 3 µl supplied USER enzyme (37 °C 15 minutes) cleaves this, resulting in a linear dsDNA adapter-ligated fragment. At this point the samples were cleaned using the supplied SPRIselect beads (Beckman Coulter; supplied by NEB). Adapter-ligated fragments were PCR amplified with 5 µl each universal primer and a unique index adapter for each sample, in a reaction with 25 µl supplied Q5 master mix. The cycling conditions used were 98 °C 30 seconds, *n* cycles of [98 °C 10 seconds, 65 °C 75 seconds], and 65 °C 5 minutes. For the first DRIP, 11 cycles were used; this was determined to be higher than required so the second DRIP used 8 cycles. These reactions were cleaned again using the supplied beads. Concentration of libraries was quantified using a Bioanalyzer high sensitivity DNA chip. Libraries were pooled and subject to an additional round of clean up using SPRIselect beads, added at 0.85X volume to select away ~<200 bp primer-dimers and eluted in 11 µl water. 1 µl was used to validate adequately concentrated libraries and the remainder sequenced on an Illumina NextSeq500 (1x75 bp/2x75 bp high output for SE/PE respectively) at the University of Cambridge Biochemistry Department generating $8x10^8$ reads.

2.3.6 Cloning techniques

To clone the AsiSI construct used in this thesis, the pBABE-HA-ER-AsiSI plasmid was sub-cloned into pCI-Neo (Table 8-3). Primers (Table 8-4) were used with 5'-uncompatible ends containing restriction sites (SacII and SalI). PCR using these primers was used to generate fragments, which were digested using SacII and SalI-HF (NEB, R0157 and R3138) in 1X CutSmart buffer at 37 °C for 3 hours. Prior to this, parental plasmids were degraded using DpnI (NEB, R0176) which targets methylated bacterial-grown DNA. Digested DNA was run on a 1% agarose gel and the appropriately sized band excised. DNA was eluted from the gel fragment using the Zymoclean gel elution kit (Zymo, D4008). The AsiSI fragment was ligated into the pCI vector using 1 U T4 DNA ligase (Promega, M1801) in LigaFast buffer (Promega, M8221) for 5 minutes at room temperature and immediately transformed into competent DH5α *E. coli*. DNA was extracted from bacterial culture by miniprep (Wizard Plus system; Promega,

A1460). Successful clones were validated by Sanger sequencing, this was then scaled up for DNA extraction by maxiprep (Machery-Nagel NucleoBond Xtra, 740414).

2.4 Protein methods

2.4.1 Protein extraction

For protein isolation, PBS-washed cells were lysed on ice in a high salt radioimmunoprecipitation assay (RIPA) buffer (400 mM NaCl, 1.5% Igepal CA-630 (Sigma, I8896), 0.5% Sodium deoxycholate (Sigma, D6750), 0.1% SDS, 50 mM Tris pH 8.0, 1x complete protease inhibitor cocktail (Roche, 11836170001), 1x phosphatase inhibitor cocktail (J63907, Alfa Aesar)). Where all cells in a well were to be used for protein preparation, RIPA was added directly to the plate and cells scraped; otherwise RIPA was added to cell pellets. 80 µl RIPA was added per $1x10^6$ cells. Genomic DNA was sheared by sonication within a water bath-submerged sonicator on the highest setting for 5 minutes. Lysis occurred on ice until 30 minutes following addition of RIPA. Lysate was clarified by the centrifugation of non-soluble material (nucleic acids, cell membranes) at maximum speed for 5 minutes at 4 °C and transfer of the supernatant to a new tube.

Protein content was determined by Bradford assay. An aliquot of each sample was diluted 1:25 with water and this was added in duplicate to a cuvette. 950 µl 1X protein assay dye reagent (BIO-RAD, 500-0006) was then added to each cuvette. An equal volume of protein standards (BSA at 50, 100, 200, 400, 500 µg/ml) was also measured alongside samples. Protein concentration was determined by measuring A_{595nm} in an Eppendorf BioPhotometer D30 which calculates a standard curve and outputs direct concentrations. All samples within an experimental replicate were diluted in RIPA buffer to match the lowest concentration. Finally, an equal volume of Laemmli loading buffer (187 mM Tris pH 6.8, 30% glycerol, 6% SDS, 0.03% bromophenol blue) with 7.5% v/v β-mercaptoethanol (Sigma, M3148) was added to each sample. Samples were fully denatured by boiling at 95 °C for 5 minutes and stored long term at -20 °C.

2.4.2 Western blotting

For validations of damage induction and siRNA induction, extracted protein lysates were separated on SDS-polyacrylamide gels, transferred to nitrocellulose membranes, and immunoblotted with appropriate antibodies. SDS-polyacrylamide gels were

prepared in the laboratory using diluted 30% polyacrylamide (National Diagnostics, EC-890) containing 0.1% SDS. Resolving gel was cast first and contained 375 mM Tris pH 8.8; stacking gel on top of this was buffered with 125 mM Tris pH 6.8. Gels were set with 0.05% APS final and 30 µl TEMED. Protein lysates in Laemmli loading buffer were loaded alongside protein ladder (Bio-Rad, 161-0374) and run at 120 V for ~2 hours, until the dye front reached the bottom of the gel. Separated proteins were wet transferred from SDS-polyacrylamide gels onto 0.45 µm nitrocellulose membranes (GE, 1060002) to ensure complete transfer of the larger proteins studied in this thesis. Wet transfer buffer (25 mM Tris pH 8.3, 192 mM glycine, 20% methanol) was pre-chilled and transfer was carried out for 90 minutes set to 100 V/0.8 A.

Membranes were checked for transfer by staining with Ponceau S for 30 seconds; this was washed away with TBST (0.1% Tween (Sigma, P1379)). The membrane was then blocked in 5% milk (Marvel) in TBST at room temperature for 1 hour. Primary antibodies diluted in 5% milk-TBST were incubated overnight at 4 °C with rotation (see Table 8-2 for antibody information). The following day, membranes were washed 3 times with TBST and incubated with appropriate fluorescent secondary antibodies in 5% milk-TBST for 1 hour. Membranes were washed 3 times before visualisation on a LI-COR Odyssey scanner and analysed using Image Studio software (v2.1, LI-COR).

2.4.3 Luciferase-based assay for miRNA activity

Luciferase assay is described in (Meijer et al. 2013). U2OS or A549 cells were seeded and reverse transfected with siRNA in six wells of a 48-well plate per siRNA condition. 2 days later, pGL3 Firefly luciferase construct was transfected to all wells. pRLSV Renilla luciferase was also transfected in the same reaction; three of the six wells with the 0 let7 site construct whereas the remaining three had the 2X let7 sites construct. The following morning, cells were PBS washed and lysed on plate with 50 µl 1X passive lysis buffer (Promega, E194A) for 10 minutes at room temperature with agitation. 5 µl of each lysate were transferred to a well within a 96-well assay plate (Applied Biosystems, 4346907) which was then placed in an automated luminometer (Glomax, Promega). Assay was carried out using the built-in dual-reporter assay programme, using LarII and Stop&Glo (Promega, E1980) for the luminescence of Firefly and Renilla luciferases. Wells containing just passive lysis buffer were used as a blank measurement. Signal was averaged across each triplicate and normalised to Firefly for

each sample. Relative luciferase intensity for three biological replicates of the no site and 2X-site constructs was then plotted.

2.5 Bioinformatic approaches

For the NGS experiments in Chapters 4 and 5, all downstream analyses were performed using a combination of the command line (Ubuntu 14.04) and the R language (v3.3.3). The latest versions of programmes were typically used and updated when available; the version number at time of writing is cited. Where published programmes and packages were used, the specific commands will be described in this section (command line code will begin with **$** whereas R will begin **>**). For custom *ad hoc* scripts, there will be an explanation of what was done however due to word limit constraints the actual code will not be shown in this document; this can be made available on request.

2.5.1 Terminology

FASTQ files are the output from sequencing runs; these contain the read sequence, quality of call per base, and metadata about the run. **Paired-end** (**PE**) refers to a sequencing reaction where both ends of the same fragment are sequenced and this information is recorded in the FASTQ metadata; **single-end** (**SE**) is more common where only one end of a fragment is sequenced. **BAM** files are the output from alignment of the FASTQ files against a genome/reference and contain additional information about the mapping. **BED** files contain just the coordinates (chr, start, stop) of features (e.g. reads) rather than the sequence itself and allow quicker analysis post-alignment. **BEDPE** is the paired-end equivalent of BED. **Trimming** is the removal of the known adapter sequences for runs where the read length is longer than the fragment (i.e. small RNA). **Coverage** as used in this thesis refers to the number of times a base is read in the sequencing reaction. **Mappability** refers to how readily a given sequence can be mapped (e.g. repetitive sequences are difficult to map to). **Concordant** mates are mapped PE reads that align on the same chromosome, conversely, **discordant** are those mates that appear to map on different chromosomes and are usually due to repetitive sequences or may reflect real chromosomal/gene translocations.

2.5.2 Small RNA sequencing read processing

The 3' adapter sequence was trimmed from small RNA reads using cutadapt v1.11 (Martin 2011):

```
$ cutadapt –a TGGAATTCTCGGGTGCCAAGG –m 17 –M 30 –o <OUTFILE>.fastq
<INFILE>.fastq
```

For small RNA library preparation using adapters with degenerate 4 nt sequences on the ligating edges (sections 4.2 and 4.7), trimmed FASTQ reads were filtered using an in-house script to only keep identical reads if they had a unique 4 nt barcode. Following this, the 4 nt was indiscriminately removed from both ends of the read using an additional call to cutadapt following adapter trimming:

```
$ cutadapt –u 4 –u -4 –o <OUTFILE>.fastq <INFILE>.fastq
```

Trimmed reads were aligned to the human genome (latest Ensembl release GRCh38 (Yates et al. 2016)) using bowtie2 v2.2.9 (Langmead and Salzberg 2012):

```
$ bowtie2 –p 8 –x <GENOME_INDEX> -U <TRIMMED_READS>.fastq -S
<OUTFILE>.sam
```

SAM files were converted to binary SAM (BAM) and sorted by samtools v1.3.1 (Li et al. 2009):

```
$ samtools view –bS <INFILE>.sam | samtools sort –o
<OUTFILE>.srtd.bam
```

To remove miRNA reads from downstream analyses, genome-aligned small RNA were then mapped to miRNA hairpins (miRBase sequence database release 21; (Kozomara and Griffiths-Jones 2014)), again using bowtie2. The unaligned reads from this were then re-aligned to the genome to obtain a BAM file without miRNA. To remove other annotated small RNA such as snoRNA, snRNA and tRNA, a custom BED file containing the genomic coordinates for these was constructed using the ensemble reference release 85 (Yates et al. 2016) and the DASHR small non-coding RNA database (Leung et al. 2016). Samtools was used to remove all reads mapping to this BED file:

```
$ samtools view –h –L small_rna.bed –U <OUTFILE>.bam <INFILE>.bam
```

This BAM file contains all remaining unannotated small RNA and was used for further analyses. To fractionate this file based on sizes of small RNA, samtools and awk were used:

```
$ samtools view –h <INFILE>.bam | awk '{if ($1 ~ "@")print $0; else
if ($1 !~ "@" && length($10) <= 23 && length($10) >= 21)print $0}' |
samtools view –hbS - > <OUTFILE>.21-23.bam
```

This was repeated changing the parameters to also extract small RNA 19-20 nt and 24-26 nt.

To calculate size distribution of reads (Figure 4.8), the BAM files before and after miRNA hairpin alignment were read by samtools and piped to an in-house awk script to count reads at each length (17-30 nt). These values were divided by the total number of reads to obtain a proportion which was then plotted using the ggplot2 package in R (Wickham 2009).

2.5.3 miRNA differential expression (DE)

For section 4.4, DE of miRNAs within the dataset were analysed. To do this, the genome aligned BAM files were used and miRNA reads were counted using HTSeq-count v0.6.1p1 (Anders, Pyl, and Huber 2015) using the miRBase release 21:

```
$ htseq-count --format=bam --type=miRNA --idattr=Name --quiet
<INFILE>.bam <MIRBASE_REF>.gff3 > <COUNTS_FILE>
```

These raw counts were imported into R for use with the DESeq2 package (v1.12.4; (Love, Huber, and Anders 2014)) where `sampleTable` is a data frame containing metadata about conditions (e.g. timepoint and replicates):

```
> dds <- DESeqDataSetFromHTSeq(sampleTable = sampleTable, directory
= dir, design = ~ time)
```

DESeq2 fits a generalised linear model for each gene by modelling count data as a negative binomial (NB) distribution. This model effectively assumes most genes should not change, and thus normalises and calculates dispersion estimates between replicates with this as a fundamental component. The outcome of this is that DE at very low and high counts can be more reliably tested. DE was performed to look for both changes over time (the previous command applied this as a design formula) and changes at any time point by using a likelihood ratio lest (LRT) and applying a reduced model.

```
> dds.time <- DESeq(dds)
> dds.LRT <- DESeq(dds, test = "LRT", reduced = ~ 1)
> res.time <- results(dds.time)
> res.LRT <- results(dds.LRT)
```

miRNAs that were DE with a false discovery rate (FDR; also referred to as an adjusted p-value, p-adj) < 0.1 were extracted for KEGG pathway analysis (section 2.5.7). The log2(fold change) and –log10(unadjusted p-value) were used to generate volcano plots

for each time point against undamaged; with conditional highlighting of data points with FDR < 0.1.

2.5.4 Small RNA around AsiSI sites

To search for damage-induced *de novo* small RNA around DNA break sites, the non-miRNA/sncRNA BAM files were used. For fitting to a NB distribution and DE of counts within AsiSI-proximal regions, DESeq2 was similarly used as before. All 1220 AsiSI motifs were analysed; as 1121 of these are not cut, there should be no change and thus the DESeq2 GLM should identify even small changes around those loci that are cut. A list of the 99 cut AsiSI site is in Table 8-5 (Aymard et al. 2014).
For plotting coverage around AsiSI loci, these BAM files were converted to BED files and position sorted as the coordinates of a mapped read are sufficient for these analyses.

```
$ bedtools bamtobed –i <INFILE>.bam | sort –k1,1 –k2,2nr >
<OUTFILE>.bed
```

Read coverage was then calculated at AsiSI loci, which were annotated within a BED file. For small RNA, coverage was determined in 10 bp windows. This was performed using bedtools v2.26.0 (Quinlan and Hall 2010), where `CHR_LENGTHS` is a 2-column text file containing chromosome number and length of the chromosome as needed for faster analysis by bedtools:

```
$ bedtools coverage –s –sorted –g <CHR_LENGTHS> -a <AsiSI_BINS> -b
<READS>.bed > <PLUS_STRAND_COV>
$ bedtools coverage –S –sorted –g <CHR_LENGTHS> -a <AsiSI_BINS> -b
<READS>.bed > <MINUS_STRAND_COV>
```

Here the `–s/–S` options select for reads on the plus/minus strand respectively. Coverage values were normalised to total number of mapped reads per million, and then each site averaged per window. Conditions were pasted together and this data frame was imported into R for graphing using ggplot2. For all coverage graphs, positive/negative coverage refers to coverage on the plus/minus strand respectively. For plotting representative transcriptional activity control sites, this was repeated the same way. Those control sites were produced from 4sU-seq data in section 4.9.

2.5.5 4sU-seq analysis

Paired-end FASTQ files from 4sU-enriched RNA was aligned to the human genome (GRCh38) using HISAT2, an alignment programme that can also account for spliced RNA reads (D. Kim, Langmead, and Salzberg 2015):

```
$ hisat2 –p 8 --dta --rna-strandness RF –x <GENOME_INDEX> -1
<MATE_1>.fastq -2 <MATE_2>.fastq –S <OUTFILE>.sam
```

The aligned SAM was converted to BAM and sorted by samtools as before (section 2.5.2). The 4sU-seq libraries were generated with a strand-specific protocol and PE sequenced. The second read of each pair represents the 5' end and strand of the original RNA fragment. Only this mate was used for further analyses by extracting only the mate 2 from the paired BAM by selecting for appropriate flags:

```
$ samtools view –F 260 –f 128 <PE_BAM>.bam > <MATE_2>.bam
```

A BED file was also made from this SE BAM file using bedtools as before. SE BAM was used for counting in AsiSI loci using htseq-count; BED was used to plot coverage.

2.5.6 DE of 4sU-enriched genes

The pseudo-alignment programme kallisto (Bray et al. 2016) was also used to align and count 4sU reads mapping to the human transcriptome (Ensembl release 85 (Yates et al. 2016)) to obtain transcript per million (TPM) for each gene using 10 bootstraps:

```
$ kallisto quant –i <INDEX> -b 10 –rf-stranded –t 8 –o <OUTPUT>
<INFILES>.fastq
```

The TPM values per gene were used as counts for DE analysis by DESeq2. To import this dataset, the tximport package and DESeq2 function was used:

```
> txi <- tximport(files, type = "kallisto")
> dds <- DESeqDataSetFromTximport(txi, colData = samples, design = ~
condition)
```

Following import of the kallisto-counted table, differential expression was performed as above (section 2.5.3). Contrasts were extracted to obtain DE tables for each appropriate comparison and these were used for pathway analysis.

2.5.7 KEGG and GO analysis of DE genes

The list of differentially expressed miRNA were uploaded onto the DIANA-mirPath v.3 website (Vlachos, Zagganas, et al. 2015). KEGG analysis was performed using TarBase v7.0 validated miRNA targets (Vlachos, Paraskevopoulou, et al. 2015). This analysis finds all the validated miRNA targets and performs pathway analysis on these, which provides a read out for the pathways controlled by those miRNAs.

For 4sU-seq, DE genes were subject to gene ontology (GO) analysis on the GSEA Molecular Signatures Database (MSigDB) v6.0 (Subramanian, Tamayo, and Mootha 2014). For TF motif analysis, the MSigDB was also used.

2.5.8 DRIP-seq analysis

DRIP in section 5.3.2 was single-ended (SE), whereas the second DRIP (section 5.4) was paired-ended (PE). For both runs, FASTQ files were aligned to the human genome (GRCh38) using bowtie2 with appropriate SE/PE settings. Output SAM was converted to BAM, with a filtered applied to remove reads with a mapping quality (MAPQ) < 30, as typically done for chromatin IP sequencing as these reads are normally erroneous or originate from highly repetitive loci and thus have many mappings. For PE BAM, if one read had MAPQ<30 but the mate did not, then this was not filtered away as long as the reads were concordantly mapped (one read in a repetitive region with the mate in a unique region would be a sensible mapping to keep with PE sequencing). These filtered BAM were converted to BED. For PE DRIP, this was a BEDPE file; the full sequenced fragment information was then inferred by taking the most 5' and 3' ends of the 5' mapping and 3' mapping read respectively. This was used to make a new BED file with fragments instead of reads. Interchromosomal pairs and extremely long (>2500 bp) intrachromosomal pairs were ignored for these analyses.

Coverage of reads or fragments was calculated around the 99 cut AsiSI sites (Table 8-5) using bedtools, with the omission of stranded counting options. This was normalised to RPM or FPM as appropriate. This was repeated for the 99 control sites, as well as the 1121 remaining uncut sites. For Figure 5.6, 4 random subsets of 99 from the 1121 were extracted unbiasedly using the command line:

```
$ shuf –n 99 1121_sites.bed | sort –k1,1 –k2,2n > random_1.bed
```

For Figure 5.8 the control undamaged 4sU-seq data was used to separate the 99 AsiSI sites in to groups of 33 by transcriptional activity. The mean TPM across the three biological replicates was used to order all genes by transcriptional activity. AsiSI cut genes were then assigned low, medium or high by rank in this list. AsiSI sites >500 bp away from a gene were assigned a value of 0 TPM and thus in the low TPM group.

3 Chapter 3: Establishing a role for the miRNA biogenesis machinery in DNA repair.

3.1 Introduction to chapter.

Genomic DNA is constantly bombarded by both endogenous and environmental agents that result in a diverse range of damage to genomic DNA. DNA can be damaged during routine replication, during metabolism by reactive oxygen species (ROS), ionising radiation (IR) from X-rays and radiotherapy, UV from sunlight, and from ingested chemicals (section 1.2.1) (Hoeijmakers 2009). These factors and others result in a multitude of different modifications to the DNA helix with the most deleterious form being double strand breaks (DSBs) which, due to the nature of the lesion, can easily result in erroneous repair and chromosomal translocations (Ciccia and Elledge 2010). Ultimately, this will result in either cell death or transformation.

Following DSB induction and DNA damage recognition there is an activation of the early signalling cascade that leads to chromatin remodelling. A crucial part of this is the activation of ATM and subsequent phosphorylation of histone H2AX, a histone modification present at all DSBs. This is followed by various repair factors being recruited to the damage site to initiate repair pathway choice (section 1.2.3).

In the many decades of DSB repair research, studies have primarily focused on the involvement of proteins in detecting breaks, signal transduction of the damage signal, and the resolution of damage. It was not until relatively recently that RNA has been implicated in playing some role in DNA repair (reviewed in section 1.5). For example, several high-throughput proteomic studies have identified RNA processing enzymes as being important for the repair process (Matsuoka et al. 2007; Paulsen et al. 2009; Hurov, Cotta-Ramusino, and Elledge 2010; Slabicki et al. 2010; Cotta-Ramusino et al. 2011; Adamson et al. 2012), whilst other bodies of work have demonstrated that active transcription influences the repair of a break within that locus (Tang et al. 2013; Aymard et al. 2014).

Of particular interest to this project is the involvement of the miRNA biogenesis enzymes in DNA repair. The miRNA pathway involves the transcription of a pri-miRNA gene, liberation of the pre-miRNA hairpin in the nucleus by Drosha, and further processing to a mature 21-23 nt miRNA in the cytoplasm by Dicer (section 1.4.2). The

mature miRNA is then incorporated into Argonaute (Ago) which assists in locating a target mRNA by base pairing with the miRNA. This leads to translation repression, sequestration, or degradation of the target mRNA, dependent on several factors such as the TNRC6 adaptor proteins and the CCR4-NOT complex. Recently, the biogenesis factors Drosha and Dicer have been implicated in DNA repair in plants, *Drosophila,* and human cells (Francia et al. 2012; W. Wei et al. 2012; Michalik, Böttcher, and Förstemann 2012). Intriguingly, these reports also observed new small RNA species derived from the sequence around an induced DSB. These were proposed to have functions in either repair signalling, recruitment of repair factors to sites of damage, or to degrade potentially aberrant transcripts. These publications had several caveats however, namely the extensive use of exogenously integrated reporters with abnormally high expression levels or highly repetitive sequences. Furthermore, the stage within the DDR that Drosha and Dicer act was disputed among these, as was the proposed mechanism of action. The aims of this PhD project were to reproduce these findings and determine whether this was a general feature of DSB repair in endogenous loci, and to consolidate the previously discussed (section 1.5) lines of evidence for a role of RNA in DNA repair. This chapter will focus on establishing which miRNA components participate in DSB repair and to further characterise this. To do this, various approaches were undertaken. Firstly, immunofluorescence (IF) was used to investigate the recruitment of various repair factors to sites of damage and how loss of Drosha and Dicer affects this; secondly, repair outcome was also studied. Two damage induction models were also utilised. Ionising radiation (IR) is a biologically relevant inducer of DSBs and as such is a gold standard for DNA repair studies. An additional system utilises an inducible endonuclease that cleaves at ~100 loci within the endogenous genome (Iacovoni et al. 2010; Aymard et al. 2014). Establishing this system was also a requirement for the subsequent chapters which involve nucleotide-level methodologies.

3.2 Drosha and Dicer, but not TNRC6A-B, are required for 53BP1 recruitment to sites of DNA damage.

To investigate the role of the miRNA biogenesis enzymes and the downstream function of the miRNA pathway in DNA damage repair, a DSB inducing system needed to be

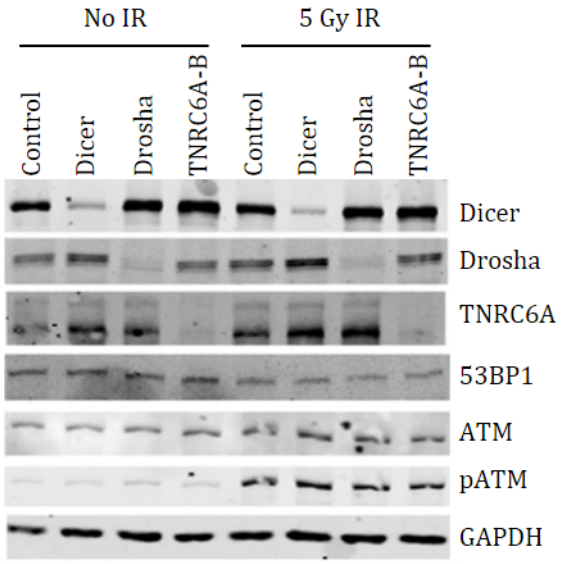


Figure 3.1. Treatment with siRNA for 48 hours is sufficient to knockdown Drosha, Dicer, and TNRC6A-B and this does not affect pATM activation.

A549 cells treated with the indicated siRNA for 48 hours were subject to 5 Gy of IR. Immunoblotting with the appropriate antibodies confirms knockdown of these proteins, and induction of damage as indicated by phosphorylated ATM on Ser1981. Blot is representative of three biological replicates.

established. The standard model for inducing DNA DSBs to study repair mechanics is via ionising radiation (IR), with doses of IR up to 5 Gy commonly used to induce up to 200 DSBs per cell.

Drosha, Dicer, and TNRC6A-B were depleted by siRNA for 48 hours, which was sufficient to deplete protein levels (Figure 3.1). DSBs were induced by exposing siRNA-treated A549 cells to 5 Gy IR. 6 hours later, protein was harvested and induction of DNA damage was validated by immunoblotting for phosphorylated ATM at Ser1981 (pATM), as this is a marker of activated ATM and one of the earliest events in the DDR (Figure 3.1). Levels of total and phosphorylated ATM do not appear to be affected by depletion of Dicer, Drosha, or TNRC6A-B suggesting that the miRNA pathway components have no effect on DSB detection or repair signalling.

Following this is repair pathway selection. A critical protein in this, and a major focus in this chapter, is 53BP1 which limits resection to both promote NHEJ and prevent more deleterious minor pathways such as SSA and alt-NHEJ (section 1.2.7) (Panier and Boulton 2013). Immunofluorescence was also carried out (alongside Figure 3.1) to observe 53BP1 recruitment to IR-induced foci (IRIF). The knockdown of Dicer and Drosha appeared to reduce the number of 53BP1 foci per nucleus, however loss of TNRC6A-B did not (Figure 3.2A). Quantification of this showed that this was statistically significant (Figure 3.2B). As a control, phosphorylated histone H2AX at Ser139 (γH2AX) was also probed for to confirm that observed foci are indeed sites of DSBs. The number of γH2AX foci did not change upon removal of Drosha or Dicer, confirming that DSBs were indeed induced in nuclei where 53BP1 was not recruited (Figure 3.2AB). Together, these data (Figures 3.1 & 3.2) suggest that Drosha and Dicer are not required for DDR signal transduction as no effects on the induction of pATM or γH2AX could be observed, but are implicated in 53BP1 recruitment to DSBs.

The lack of an effect of TNRC6A-B depletion on 53BP1 recruitment indicates that this is not due to the canonical activity of miRNAs. As Drosha and Dicer are the sole biogenesis enzymes known to process miRNAs to their mature form, it is reasonable that their removal would affect miRNA levels and thus perturb the many cellular pathways that miRNAs control. One such pathway is the cell cycle, which is also a major determinant of DNA repair outcome. As HR is only active in S/G2 phase, cells in these phases of the cell cycle will have factors competing with 53BP1. To test that

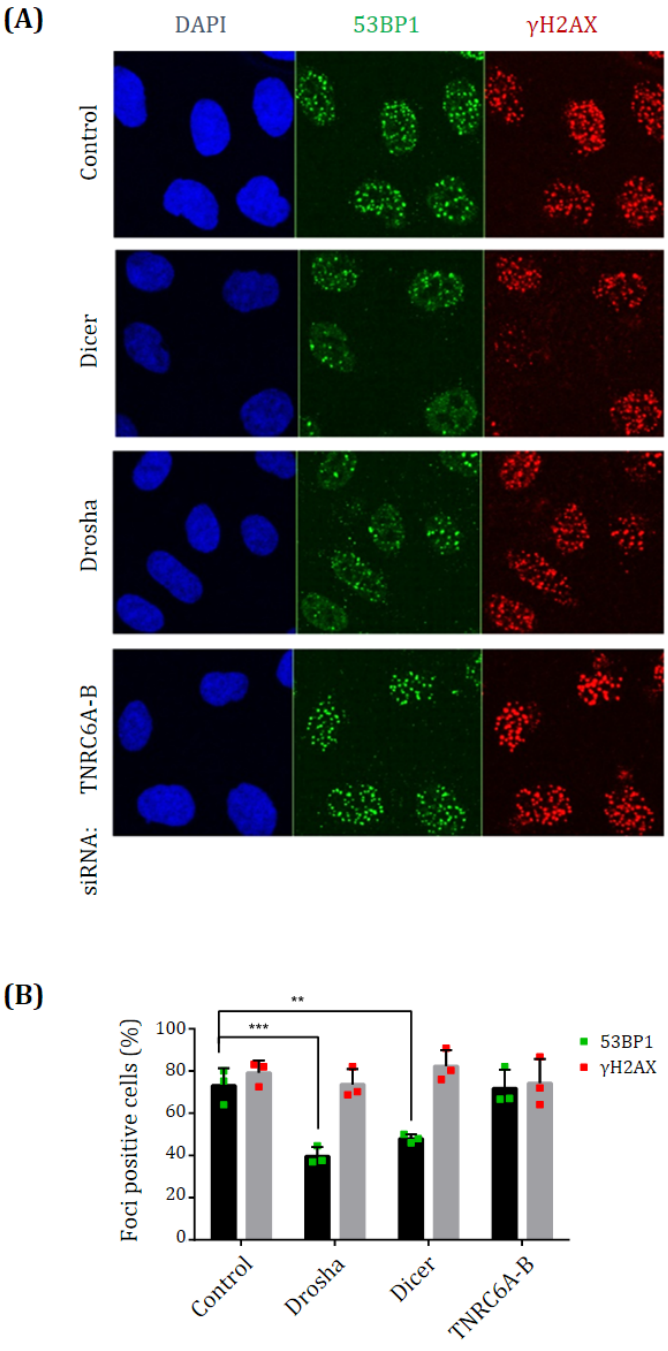


Figure 3.2. 53BP1 recruitment to IRIFs is dependent on Drosha and Dicer, but not TNRC6A-B in A549 cell lines.

(A) Representative IF images of A549 cells treated with the indicated siRNA. Cells were fixed by paraformaldehyde 6 hours following 5 Gy of IR, and 53BP1 (green) and γH2AX (red) foci were visualised, with nuclei stained by DAPI (blue). **(B)** Quantification of IF images as in A. Cells were scored positive with ≥10 53BP1 foci or ≥15 γH2AX foci. ≥70 cells scored per condition; N=3. ***p<0.001, **p<0.01, ANOVA with testing for multiple corrections. Error=SD.

depletion of Drosha, Dicer, and TNRC6A-B for 48 hours does not affect the cell cycle and thus 53BP1 recruitment, the population of cells in G1, S, or G2 phase of the cell cycle in these conditions was determined using propidium iodide staining and measuring by flow cytometry (Figure 3.3A). There was no significant change to the distribution of cells at any of these phases, indicating that the differences observed to 53BP1 foci was not due to cell cycle perturbance.

To address the possibility that additional pathways may be affected by perturbing miRNA biogenesis, the relative abundance of miRNA miR-21 was quantified by TaqMan qPCR analysis in cells deprived of Drosha, Dicer, and TNRC6A-B for 48 hours (Figure 3.3B). Abundance was calculated relative to the snoRNA U6, a small RNA that is not processed by the miRNA biogenesis enzymes and thus should be unaffected. Only siRNA depletion of Dicer appeared to have a minor effect on miR-21 levels, however this was insignificant when tested (Figure 3.3B).

It is difficult to determine whether a minor change to miRNA levels would have an impact on the cellular pathways they control. To further investigate the role of 48 hour siRNA depletion of Drosha, Dicer, or TNRC6A-B on miRNA activity within the cell, a reporter assay was utilised as depicted in Figure 3.3C (Meijer et al. 2013). This consists of a *Renilla* luciferase ORF with or without 2 sites for the miRNA let-7. Endogenous let-7 within RISC recognises these target sites leading to repression of the translation or decay of the mRNA. If miRNA levels or activity are perturbed, repression is prevented causing a de-repression of the mRNA and increased translational efficiency of a functional luciferase enzyme. Luciferase activity can be easily quantified and therefore used as a measure of miRNA activity within the cell. When Drosha and Dicer were knocked down, no changes to the miRNA activity of the cell was observed (Figure 3.3D). In contrast, removal of TNRC6A-B completely de-repressed luciferase activity, highlighting the critical role of these adaptor proteins in miRNA-mediated repression of target transcripts (Figure 3.3D). These data show that depletion of Drosha and Dicer at 2 days is not sufficient to significantly alter miRNA levels, and that any minor changes do not appear to perturb the canonical role of miRNA in orchestrating translational repression of at least the well-studied miRNAs miR-21 and let-7. As such, 48 hours of siRNA depletion of these factors was used for all subsequent experiments. These data also show that TNRC6A-B is an ideal control for observing the consequence

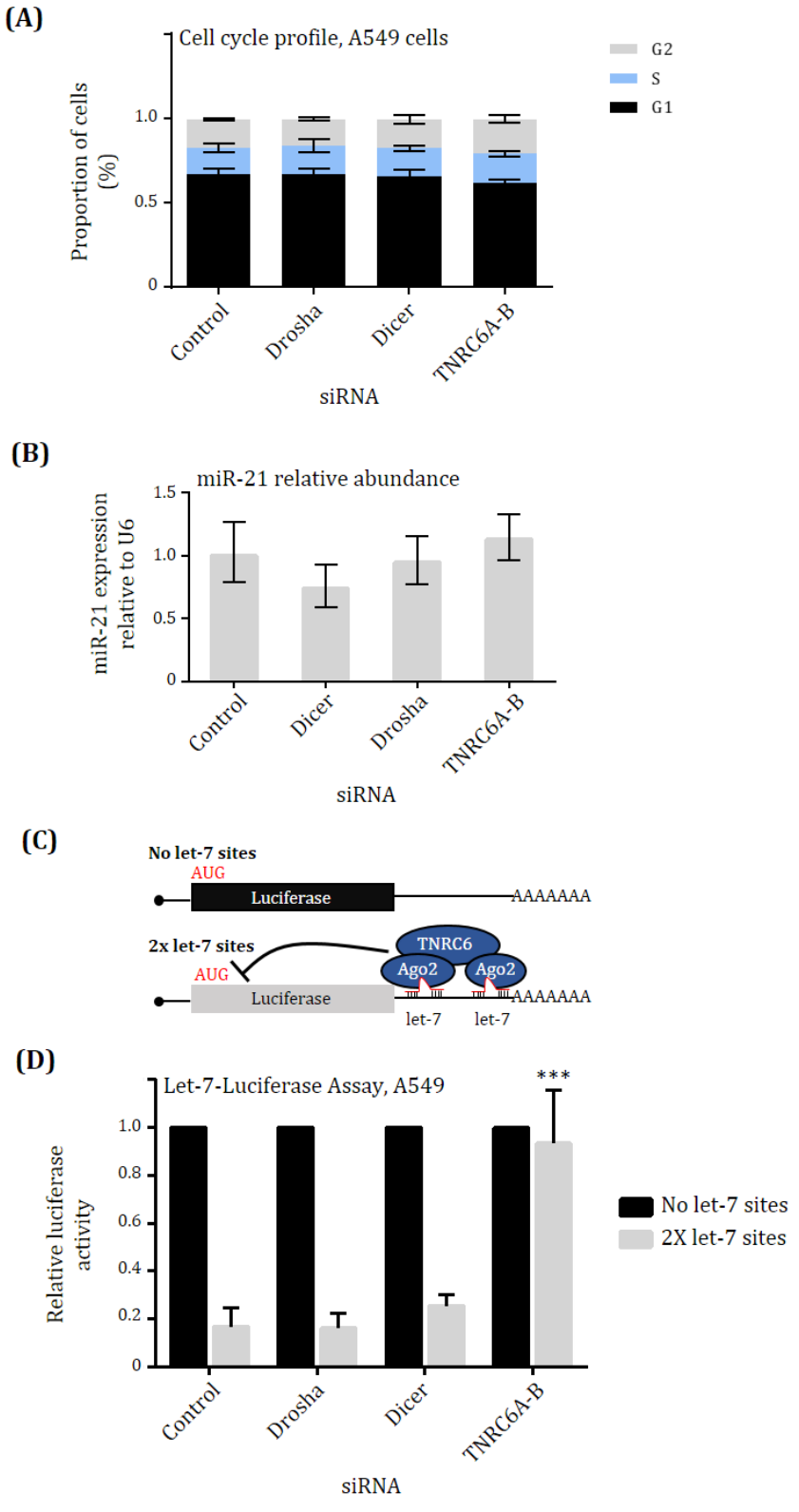


Figure 3.3. Two days of knockdown of Drosha and Dicer does not perturb the cell cycle or miRNA activity within the cell.

(A) FACS-PI analysis of cells treated with indicated siRNA show no significant changes to cell cycle populations. N=3, Student's t-test. **(B)** RNA was harvested

from A549 cells depleted of Drosha, Dicer, and TNRC6A-B. Taqman qPCR analysis of miR-21 show only minor but insignificant changes to expression levels relative to snoRNA U6. N=3, Student's t-test. **(C)** Schematic of the let7-luciferase assay. *Renilla* luciferase constructs with or without let-7 target sites were expressed in cells treated with siRNA. Endogenous let-7 represses the translation of the construct with sites; perturbation of this will de-repress translation and this can be measured by luminescence. **(D)** Quantitation of the experiment described in (C). De-repression of a luciferase ORF with 2 miRNA-let-7 seed sequences only occurs on knockdown of TNRC6A-B. A549 cells; N=3; ***p≤0.005, Student's t-test. Error=SD.

of Drosha and Dicer depletion, as any outcome as a result of reduced miRNA activity would also be seen in TNRC6 depleted samples. As TNRC6A-B had no effect on 53BP1 recruitment, but Drosha and Dicer did (Figure 3.2), this indicates that the effect of the miRNA biogenesis enzymes in DNA repair stems from an additional, non-canonical activity.

3.3 Early DDR signalling is not affected by Drosha depletion.

The presence of phosphorylated H2AX is a key part of the chromatin landscape at every DSB, acting as a platform for a feedback loop involving autophosphorylated ATM, MDC1, and the MRN complex which propagate the damage signal along the chromatin and also initiate other cellular responses to DSB such as cell cycle arrest (section 1.2.3). As was previously shown, γH2AX foci formation was unaffected when the miRNA biogenesis machinery was knocked down (Figure 3.2); furthermore, overall pATM levels were not perturbed either, as would be expected as it is the major H2AX kinase (Figure 3.1). This suggests the early DDR signalling cascade occurs prior to the point at which Drosha and Dicer may be involved. To test this hypothesis the recruitment of pATM and MDC1 to foci was examined in cells treated with Drosha siRNA for 48 hours. As the effect previously shown was strongest with Drosha depletion, from this point on Drosha knockdown was used (Figure 3.2). An earlier time point of 2 hours following 5 Gy IR was also used in order to better capture the recruitment and signalling from these early events within the DDR. Knockdown of Drosha did not affect the recruitment of pATM or MDC1 to IRIFs (Figure 3.4). γH2AX was again unaffected at the 2 hour time point, and confirms foci are indeed DSBs.

3.4 Investigation into the recruitment of repair factors downstream of 53BP1 to damage foci suggest that both major DSB repair pathways are impaired upon loss of Drosha.

Two major DNA damage repair pathways – HR and NHEJ – exist to resolve a DSB and the decision to commit to these pathways is dictated by a competition between various factors to decide which pathway is used. This ultimately revolves around resection of the broken DNA, where 53BP1 protects the dsDNA from digestion by DNA endonucleases such as EXO1. As NHEJ involves the direct ligation of the DNA ends, any resected ssDNA runs the risk of being removed, resulting in a deletion event. Thus,

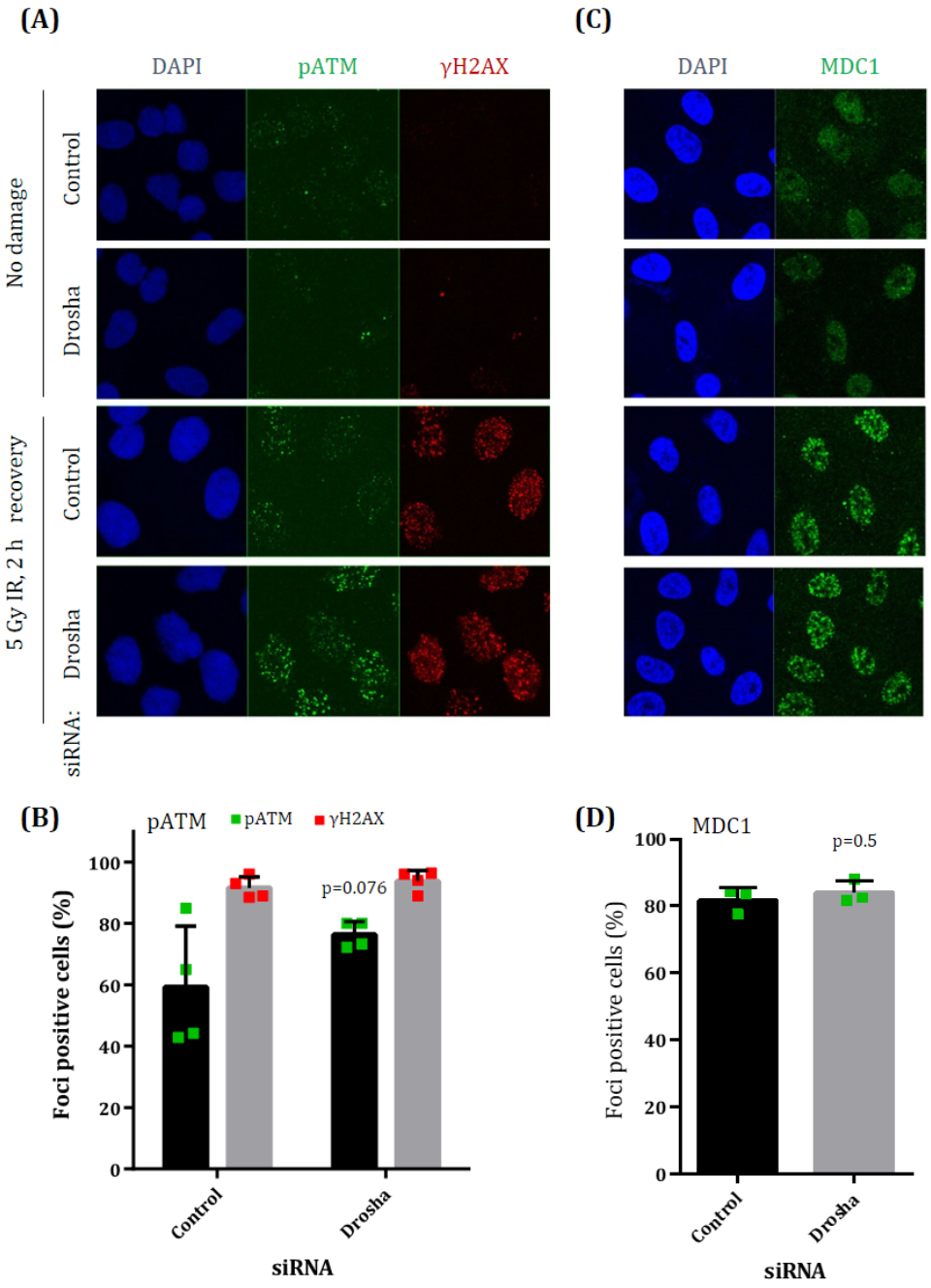


Figure 3.4. Early DDR signaling events are not dependent on Drosha in A549 cells.

(A) Representative IF image visualising 5 Gy IR-induced pATM foci in A549 cells, 2 hours after damage. **(B)** Quantification of (A). Cells were scored positive with ≥10 pATM foci or ≥15 γH2AX foci. At least 70 cells scored per condition; N=3, Student's t-test. Error=SD. **(C)** Representative IF as in (A), visualising MDC1 foci. **(D)** Quantification of (C) as before; positive MDC1 cells had ≥10 MDC1 foci.

when resection carries out beyond a certain point, the cell must commit to HR if a sister chromatid is available, or utilise SSA or alt-NHEJ (section 1.2.7). Drosha and Dicer depletion resulted in reduced 53BP1 recruitment, meaning that HR is potentially increased. To investigate this, immunofluorescence was used to evaluate the ability of two HR factors, BRCA1 and RAD51, to be recruited to DSBs following IR (Figure 3.5). As HR only occurs in S/G2 phase, the number of cells positive for HR factors should be much lower than for proteins such as 53BP1. To control for this, γH2AX was also examined to determine the full extent of DNA damage occurring within these cells. Quantification of BRCA1 and Rad51 IF showed around 50-60% foci positive cells, which dropped significantly upon Drosha depletion (Figure 3.5BD) with γH2AX maximally induced in all conditions. These data suggest that Drosha may be acting prior to 53BP1 in a way that would affect general repair factor recruitment to DSBs. This in conjunction with data presented in Figure 3.4 suggests that the likely stage at which Drosha is acting upon is the chromatin remodelling phase downstream of early DDR signalling. This generates the hypothesis that Drosha may be needed to open up or modify the chromatin in some way for HR and NHEJ proteins to be recruited.

3.5 Establishing the AsiSI cell line as an alternative system to study DNA damage.

IR is a typically used model for inducing DSBs as it is convenient and is an environmentally relevant form of damage. However, the direct and indirect effects of IR on a cell is not limited to just DNA DSBs, causing damage to RNA, proteins, and cell membranes (Reisz et al. 2014). Furthermore, the DSBs induced by environmental events including IR are randomly introduced into the genome. This limits the ability to investigate the events at nucleotide resolution that are occurring around DSBs and makes investigating the involvement of RNA molecules hard as each damage site would be unique. As a major focus of the work in this project is to study the transcriptional and RNA landscape around DSBs themselves, defined breaks need to be introduced into the genome.

To overcome these issues, the established inducible endonuclease system AsiSI was utilised (Aymard et al. 2014). There are 1220 naturally occurring AsiSI sites within the genome, however, studies have shown that only around 100 are recognised within the

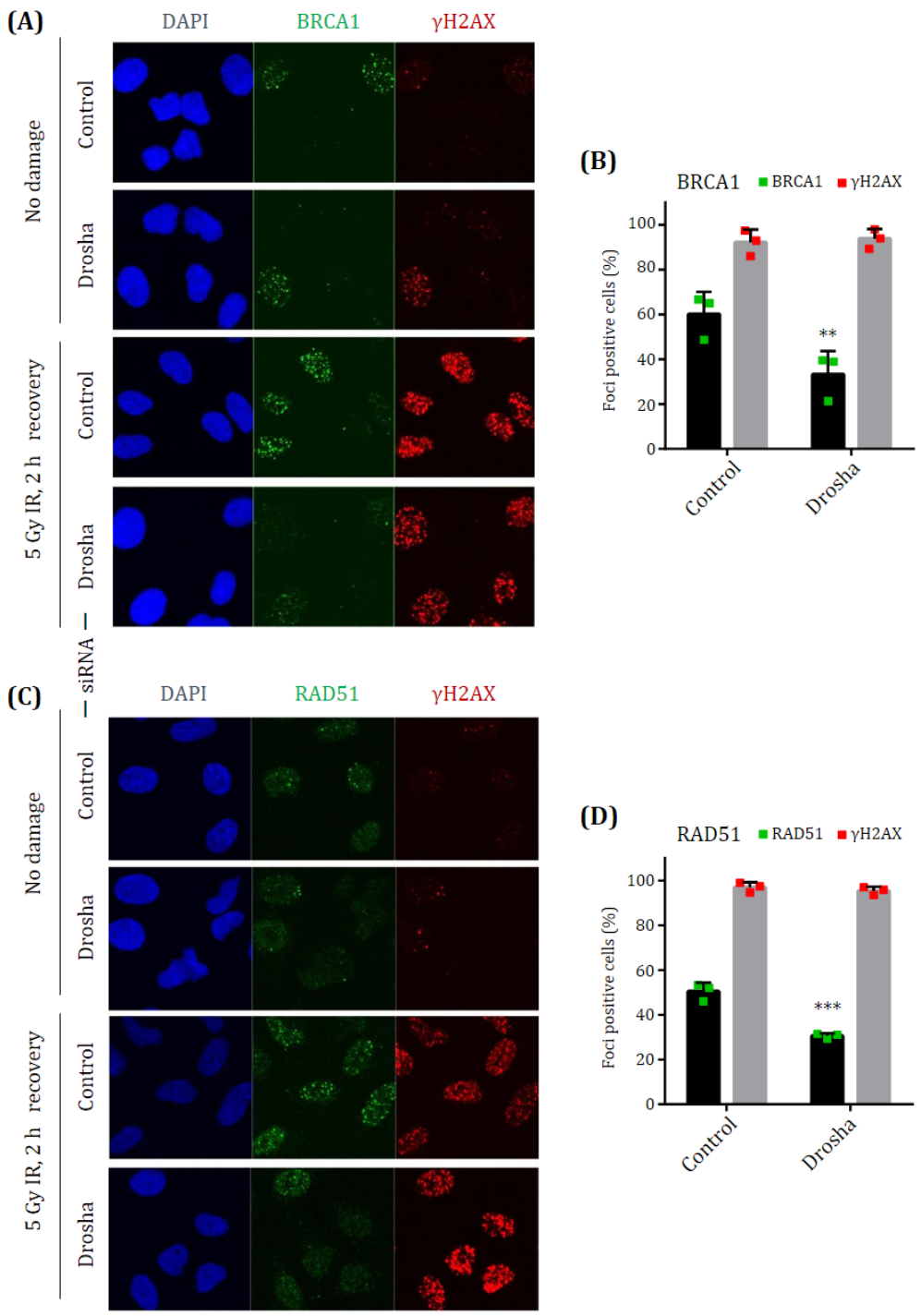


Figure 3.5. HR factor recruitment is impaired on Drosha knockdown.
(A) Representative IF image visualising 5 Gy IR-induced BRCA1 foci in A549 cells, 2 hours after damage. **(B)** Quantification of (A). Cells were scored positive with ≥10 BRCA1 foci or ≥15 γH2AX foci. At least 70 cells scored per condition; N=3. **p<0.01, Student's t-test. Error=SD. **(C)** Representative IF as in (A), visualising RAD51 foci. **(D)** Quantification of (C) as before; positive RAD51 cells had ≥5 RAD51 foci. ***p<0.001, Student's t-test.

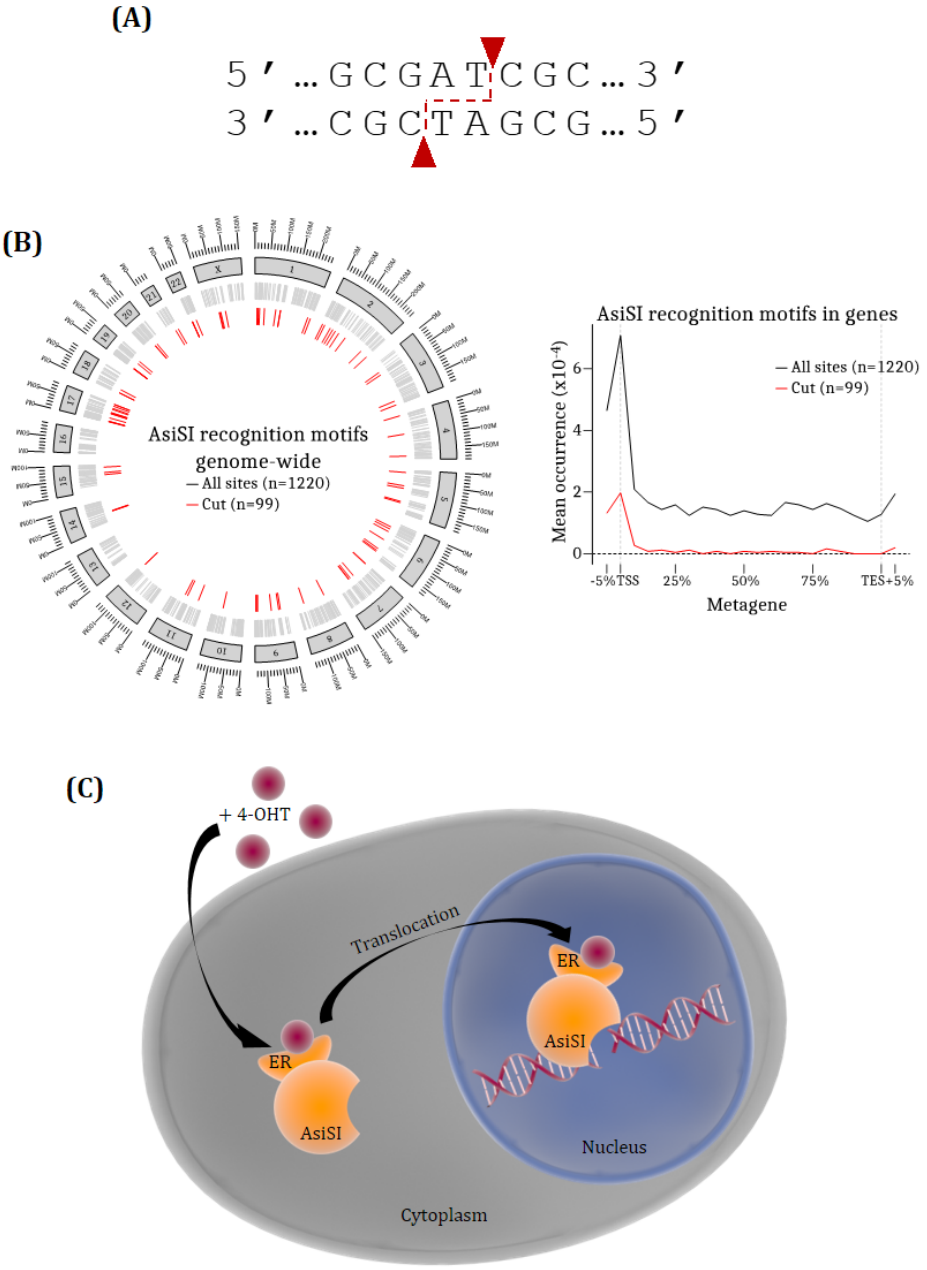


Figure 3.6. The inducible AsiSI system.

(A) The recognition motif of AsiSI. When cleaved, AsiSI leaves a 2 nt 5' overhang. **(B)** Genome-wide (left) and metagene (right) analyses of the occurrence of all 1220 AsiSI recognition motifs (grey). The 99 reliably cut sites are indicated in red. **(C)** 4-OHT is rapidly taken up by cells when added to media, where it binds the ER fused to AsiSI within the cytoplasm. This causes translocation to the nucleus where AsiSI can find the GCGATCGC motif and cleave it as in (A).

genome of U2OS cells (Figure 3.6AB) (Aymard et al. 2014). This is most likely due to CpG methylation within motifs, or the chromatin structure around some of these sites. Conveniently, this makes it comparable to the 100-200 DSBs induced by 5 Gy IR. Furthermore, the dsDNA endonucleolytic activity of the enzyme means only DSBs are produced. The addition of 4-hydroxytamoxifen (4-OHT) causes translocation of ER-tagged AsiSI from the cytoplasm to the nucleus where it can then find its recognition motif and cleave it, causing DSBs (Figure 3.6C).

U2OS cells stably expressing an ER- and HA-tagged AsiSI (U2OS-AsiSI) were a kind gift from the Legube lab. To test whether this system could cause DSBs, as well as investigate the speed at which damage could be induced after induction, a time course experiment was set up to collect total protein in parallel with fixed cells on coverslips for immunofluorescence. Cell media was replaced with fresh media containing 300 nM 4-OHT and cells harvested at 30, 60, 120, and 240 minutes alongside an undamaged control (time point 0). DSB induction was measured by quantifying occurrence of γH2AX foci by immunofluorescence, as well as overall levels by western blot (Figure 3.7 and 3.8). An increase in foci formation can be observed in as little as 30 minutes of 4-OHT treatment (Figure 3.7) with maximal induction seen at 1 hour (Figure 3.7 and 3.8A). pATM induction can be seen at the 30 minute timepoint by western blotting, however overall levels of γH2AX do not appear to peak until 2-4 hours (Figure 3.8B). This could be due to the difficulty in extracting histones with regular protein preparations due to their highly positive charge. Nevertheless, damage was readily observed by IF and pATM western blotting. By immunoblotting for the HA-tagged AsiSI, the translocation of the enzyme can also be observed at 30 minutes and is fully nuclear at 1 hour (Figure 3.7 and 3.8A). These data show that the AsiSI system can be used to induce DSBs quickly, and that AsiSI will cleave its target DNA as soon as it enters the nucleus.

3.6 A role for the biogenesis apparatus in DNA damage can also be observed in the AsiSI cell model.

To validate the previous findings using IR in A549 cells, similar experiments were carried out in the U2OS-AsiSI cell line. In all cases, damage induction was carried out by incubation with fresh media containing 300 nM 4-OHT or an equal volume of DMSO

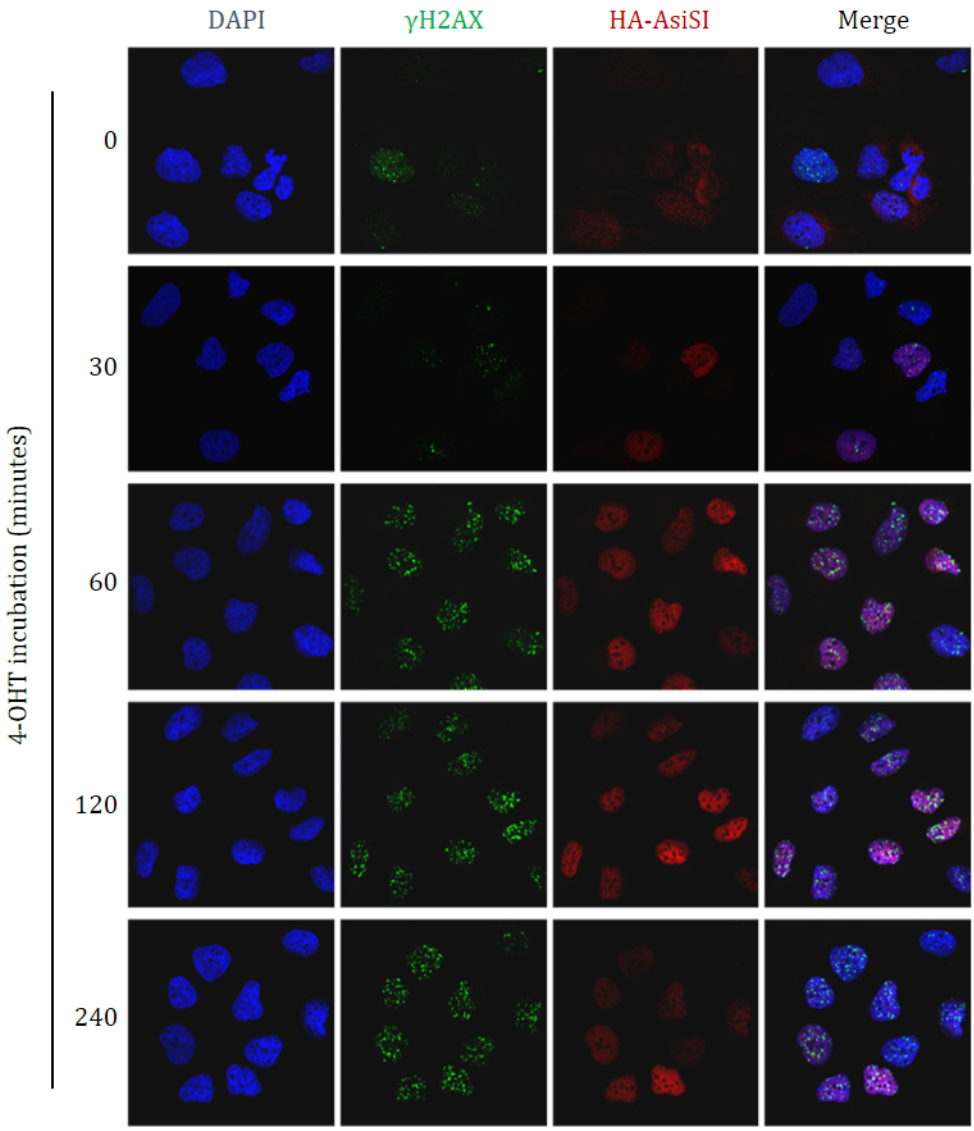


Figure 3.7. DNA damage can be reliably induced using the AsiSI endonuclease.

Representative IF images staining for damage induction (γH2AX; green) in U2OS-AsiSI cells by the addition of 4-OHT for the indicated lengths of time. In red, IF for the HA-tagged AsiSI enzyme shows 4-OHT induced nuclear re-localisation, where DAPI (blue) stains the nucleus.

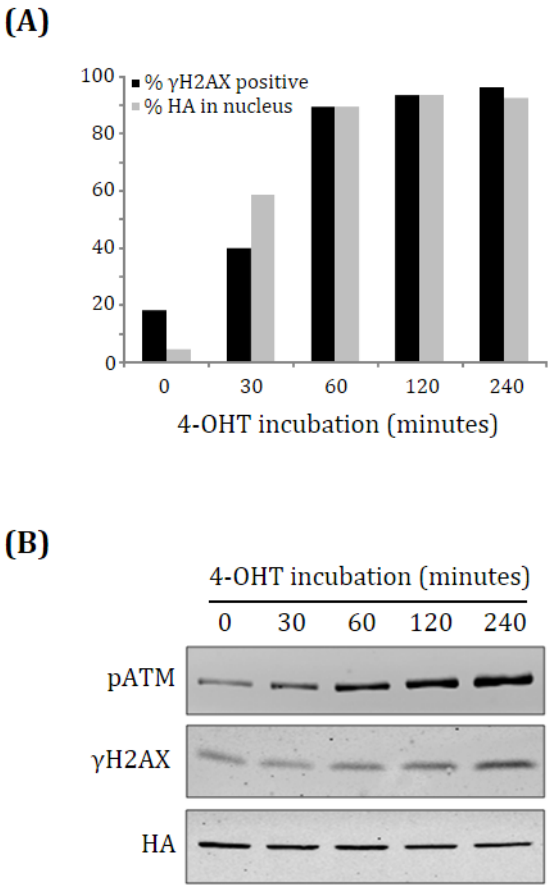


Figure 3.8. DNA damage can be reliably induced using the AsiSI endonuclease.

(A) Quantification of IF images in Figure 3.7. A γH2AX positive cell was counted as having ≥15 foci per nucleus. **(B)** Western blot showing DNA damage induction by 4-OHT incubation for the indicated times by immunoblotting for the specified proteins. HA blotting acts as a loading control in stably expressing cells.

for undamaged control. U2OS-AsiSI cells were treated with siRNA against Drosha, Dicer, or TNRC6A-B for 48 hours. DNA damage was induced by addition of 4-OHT for 6 hours, followed by harvesting of protein for western blot and fixing of cells on coverslips for IF. Knockdown of Drosha, Dicer, and TNRC6A-B was confirmed by western blotting, as was damage induction by blotting for pATM (Figure 3.9). As in A549 cells, no change to pATM levels could be observed when any of these factors were depleted, further showing that early DDR signalling is independent of Drosha and Dicer. The recruitment of 53BP1 and γH2AX to AsiSI-induced damage foci was observed via IF. As in A549 cells, loss of Drosha and Dicer reduced 53BP1 foci formation whereas TNRC6A-B did not (Figure 3.10AB). γH2AX remained unchanged in all cases. These results demonstrate that the role for Drosha and Dicer in the DDR is not specific to A549 cells or IR damage.

To confirm that this effect of Drosha and Dicer in the DDR in the U2OS-AsiSI system was similarly independent of the miRNA pathway, both the cell cycle population and miRNA activity was tested. The cell cycle profile of cells treated by Drosha, Dicer, or TNRC6A-B siRNA showed insignificant differences in G1/S/G2-phase populations, indicating that the effect seen on 53BP1 is not due to changes in the cell cycle (Figure 3.11A). The siRNA depletion of TNRC6A-B does appear to cause a slight shift from G1 to G2, however this is statistically insignificant. The role of Drosha and Dicer in 53BP1 recruitment is likely to be miRNA-independent as in A549 cells, but to test this the let-7 luciferase assay was carried out in the U2OS-AsiSI cells (Figure 3.11B). As before, at 48 hours of siRNA treatment, only TNRC6A-B knockdown caused any de-repression of miRNA targeted messages (Figure 3.11B). These results confirm a role for the miRNA biogenesis enzymes outside of miRNA-mediated repression, and that this is independent of cell type. Furthermore, this shows that this activity is specific to DSBs, as AsiSI can only cause dsDNA breaks.

3.7 Repair outcome is significantly diminished upon loss of Drosha.

The recruitment of DNA repair proteins to DSB foci is a powerful technique to study DNA damage repair, however observing reductions in foci formation does not necessarily reflect that repair of the break itself is perturbed. Repair outcome was therefore examined using GFP reporter assays which can provide a read out of repair efficiency for pathway sub-types. These systems are designed so that repair of specific

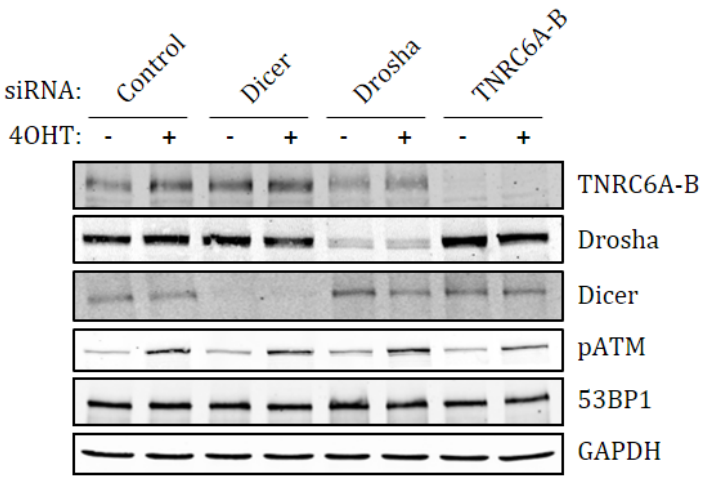


Figure 3.9. Drosha and Dicer, but not TNRC6A-B, are required for 53BP1 recruitment to sites of AsiSI-induced damage in U2OS cells.

Representative western blot confirming siRNA depletion of indicated proteins. Damage induction by AsiSI (induced by 4-OHT) as shown by pATM is unaffected, as is total levels of 53BP1. N=3.

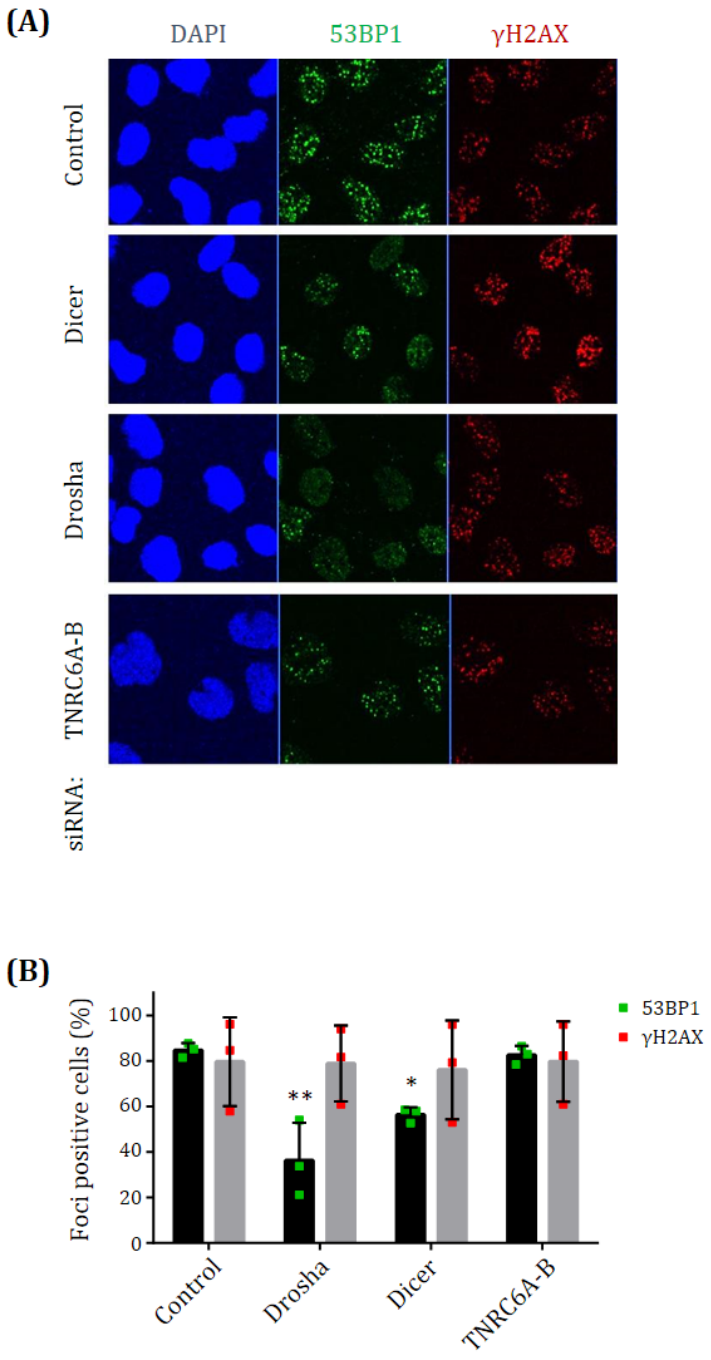


Figure 3.10. Drosha and Dicer, but not TNRC6A-B, are required for 53BP1 recruitment to sites of AsiSI-induced DSBs in U2OS cells.

(A) Representative IF staining for 53BP1 shown in green, γH2AX in red, and the nucleus (DAPI) in blue. Images are representative fields taken from the same experiment for each condition. **(B)** Quantification of IF images as in (A). Cells were scored positive with ≥10 53BP1 foci or ≥15 γH2AX foci. At least 70 cells scored per condition; N=3. **p<0.01, *p<0.05, ANOVA with testing for multiple corrections. Error=SD.

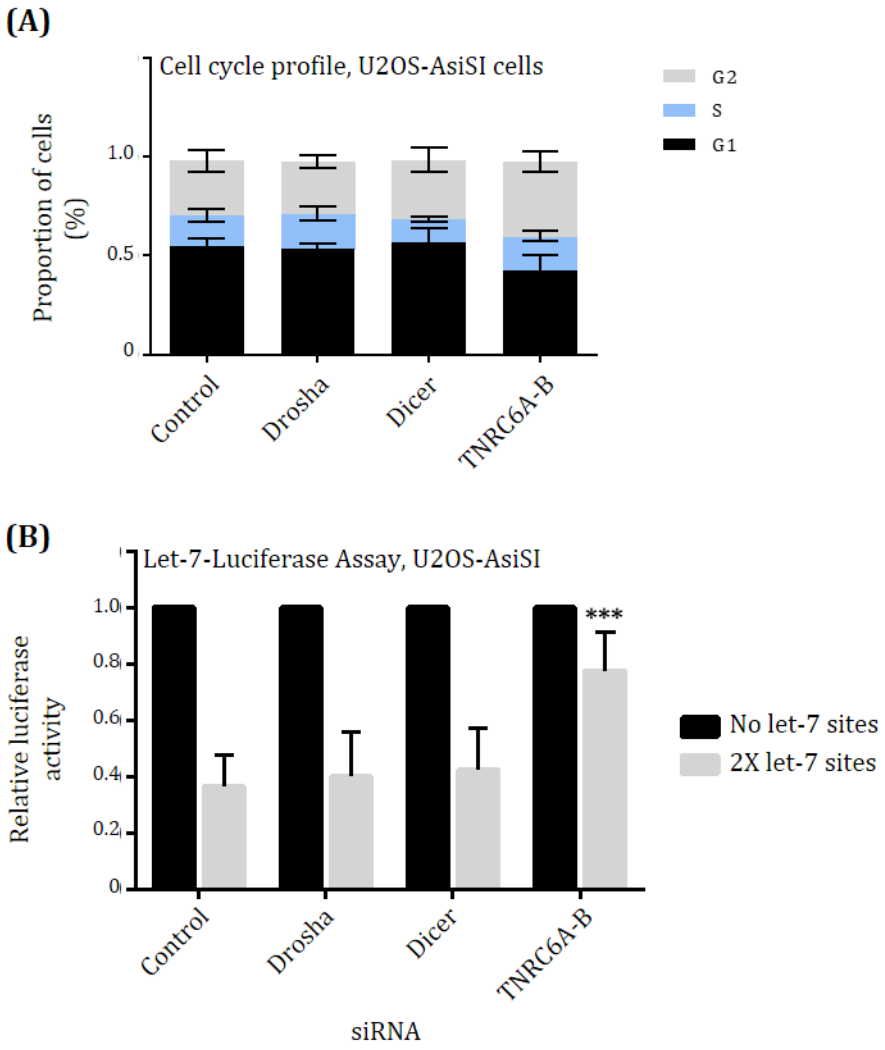


Figure 3.11. Two days of knockdown of Drosha and Dicer does not perturb the cell cycle or miRNA activity within U2OS-AsiSI cells.

(A) FACS-PI analysis of U2OS-AsiSI cells treated with indicated siRNA show no significant changes to cell cycle populations. N=3, Student's t-test. Error=SD.

(B) De-repression of a *Renilla* luciferase ORF with 2 let-7 seed sequences only occurs on knockdown of TNRC6A-B. U2OS-AsiSI cells; N=3; ***p≤0.001, Student's t-test. Error=SD.

breaks induced by the I-SceI endonuclease activates a functional GFP gene which can then be easily measured as a read out of repair efficiency by flow cytometry. The HR reporter contains a GFP ORF with two opposing I-SceI recognition motifs that causes a frame-shit rendering the translated product non-functional. Downstream of this is an incomplete GFP ORF with the correct genetic information (Figure 3.12A). Repair of I-SceI cleavage via NHEJ, SSA, or MMEJ does not result in a functional GFP however HR repair utilising the downstream GFP sequence as a template does. The NHEJ reporter consists of a GFP gene containing two introns and an exonic sequence from an adenoviral gene (Ad); splicing and translation results in a non-functional GFP protein (Figure 3.12B). Cleavage of opposing I-SceI recognition motifs within the introns either side of the Ad exon leads to its excision. End joining repair ligates the two introns together, which after splicing produces a functional GFP protein.

To ensure results were representative of actual DNA repair within a cell – which is heavily dependent on chromatin – stable integrated reporter cell lines were generated. U2OS cells were transfected with linearised reporter and selected for over several weeks as described in materials and methods (section 2.1.1.4). Individual colonies were isolated generating monoclonal lines. These were tested for successful incorporation of the reporter by transient transfection of I-SceI and measuring GFP expression (Figure 3.12C). To ensure those GFP-expressing clones were measuring the appropriate repair pathway as expected, BRCA1 was depleted prior to I-SceI expression (Figure 3.12C). As BRCA1 is required for proficient HR, a decrease in GFP expression should be observed. From the cells lines generated, the line which had the greatest differential in GFP activity before and after BRCA1 depletion was selected. The experimental set-up is described in section 2.1.5; briefly, proteins of interest were depleted for 48 hours prior to I-SceI expression to ensure repair occurs in a setting devoid of them. 48 hours later, cells were analysed for GFP expression by flow cytometry.

A panel of siRNAs was used to deplete individual components of the miRNA pathway, alongside the HR factor BRCA1 as a positive control and NHEJ factor 53BP1 as a negative control for the HR reporter, with the opposite being the case for the NHEJ reporter. Additional samples with no I-SceI transfection and with GFP transfection were used to gate GFP cells from non-fluorescing cells. These gating settings were then

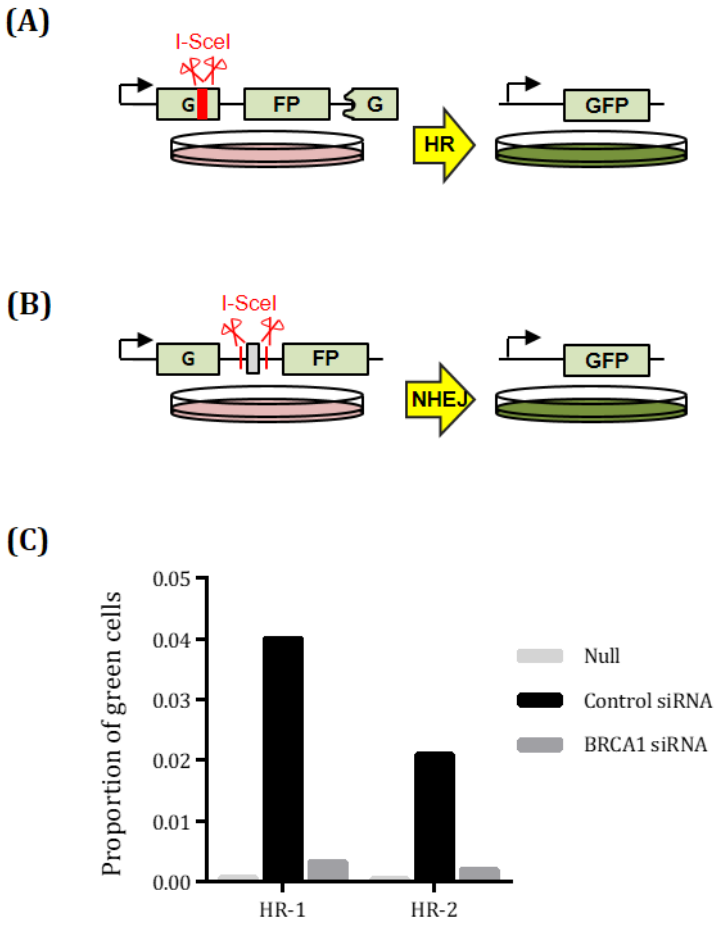


Figure 3.12. Development of repair reporter cell lines.
(A) Schematic of HR reporter. Two I-SceI sites frame-shift a GFP ORF, so no functional GFP protein is expressed. Following cleavage by I-SceI and repair by HR utilising the downstream corrected sequence as a template, a functional GFP ORF is reconstituted, which can be measured using flow cytometry. **(B)** Schematic of NHEJ reporter. The presence of an exonic sequence (grey rectangle) causes a non-functional GFP product. Excision of this exon by I-SceI expressed and subsequent NHEJ repair results in a fluorescing GFP which can be measured by flow cytometry. **(C)** The top GFP-expressing HR colonies were tested for HR specificity by BRCA1 siRNA depletion, and the best cell line used for all subsequent experiments.

used for all other samples (examples of this from an experiment using the HR reporter is shown is Figure 3.13A). The proportion of GFP-positive cells was normalised to the proportion of GFP-positive cells in the control siRNA sample as a read out of repair efficiency.

As expected, the BRCA1 depletion ablates HR repair (Figure 3.13B). 53BP1 removal did have a minor but insignificant effect on HR efficiency, reflecting its recently reported and somewhat paradoxical role in also promoting HR when the cell has already committed to resection (Ochs et al. 2016). The knockdown of Drosha and Dicer significantly reduced HR efficiency. Knockdown of Drosha cofactors DGCR8 and DDX17 had a similar effect to that of Dicer, however the opposite was observed for DDX5 where repair efficiency increased. As these are all cofactors of Drosha, it suggests that there may be an involvement of a similar complex to that of the miRNA biogenesis Microprocessor (section 1.4.2.1).

In the NHEJ reporter cell line, BRCA1 siRNA treatment had no effect whereas loss of 53BP1 decreased efficiency as anticipated, confirming the reporter measures end joining repair (Figure 3.13C). Drosha and Dicer depletion also reduced NHEJ repair, confirming the observations with IF that both major repair pathways are dependent on Drosha and Dicer (Figure 3.2 and Figure 3.5). Curiously, the effect on loss of Drosha was modest whereas loss of Dicer had as strong an affect as 53BP1 depletion. Crucially, removal of the miRNA effector components TNRC6A-C or Ago1-2 had no effect on the repair of either pathway. This further supports the previous data that the miRNA biogenesis apparatus has a role in DNA repair outside of miRNA mediated repression.

3.8 Chapter discussion

This chapter has established a non-canonical role for the miRNA biogenesis enzymes in DNA repair. Whilst this has been reported before (Francia et al. 2012; W. Wei et al. 2012), there were a number of discrepancies between publications such as the point in the repair response at which these proteins are required, and what components of the miRNA pathway are required (see section 1.5.2) (Francia et al. 2012; W. Wei et al. 2012; Michalik, Böttcher, and Förstemann 2012). The results presented here show that early DDR signalling is unaffected when a cell is depleted of Drosha (Figures 3.1, 3.2, 3.4, 3.9), meaning that the initial propagation of the damage signal is not dependent

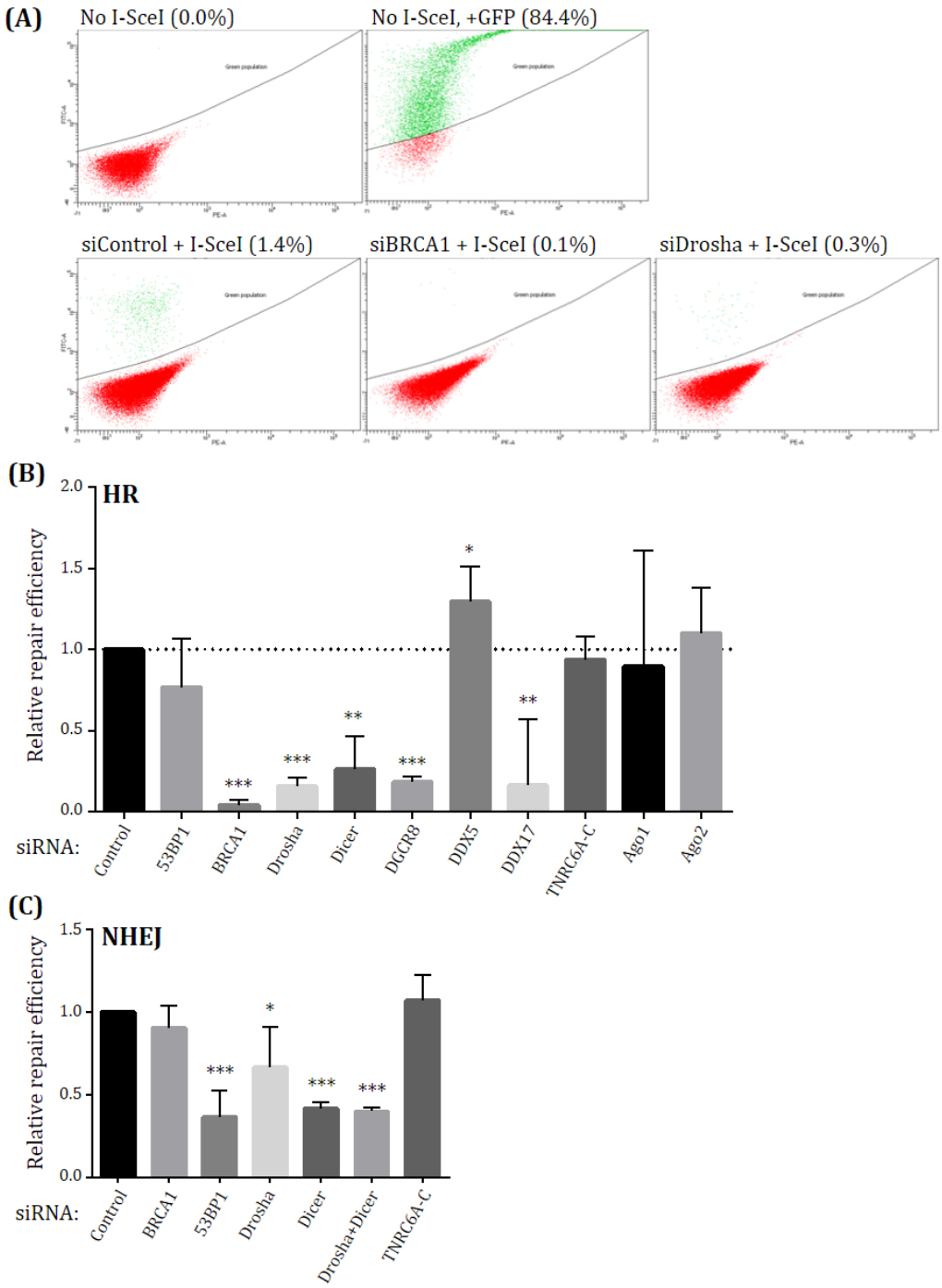


Figure 3.13. siRNA depletion of miRNA biogenesis enzymes, and several of their cofactors, are necessary for efficient repair.

(A) Example of how GFP-positive cell population was gated using the FACS software. **(B)** The proportion of GFP-positive HR reporter cells in each condition as measured by flow cytometry was normalised to the control sample as a measure of HR repair efficiency. N=4, ***p<0.001, **p<0.01, *p<0.05, Wilcoxon signed rank nonparametric test with multiple corrections. Error=SD. **(C)** As in (B), using the NHEJ reporter cell line.

on these RNA processing enzymes. This is in contrast to work by Francia et al., where the authors demonstrated a decrease in pATM recruitment to sites of damage in cells lacking Drosha or Dicer, but in concordance with Wei et al. (Francia et al. 2012; W. Wei et al. 2012). An early event that was consistently shown to be unaffected however was the phosphorylation of H2AX, which was again validated in these studies. The key H2AX kinase is autophosphorylated ATM, and whilst there is some redundancy with DNA-PK (section 1.2.3), a diminished pATM signalling pathway is consistent with reduced H2AX activation. Thus, the results presented here appear to fit more in line with the consensus in the DDR field, however identifying exactly why these discrepancies have arisen should be an important part of further study as that may help elucidate the function of Drosha in DNA repair.

Of particular interest is the reduced recruitment of both NHEJ and HR proteins upon loss of Drosha (Figures 3.2, 3.5, 3.10). As mentioned in sections 1.2.6.1 and 3.4, 53BP1 and BRCA1 have an antagonistic relationship where less recruitment of one ordinarily means an increase of the other. For both NHEJ and HR factors to be perturbed, a more general point prior to both must be affected. As this must be following early DDR signalling (Figure 3.4), the chromatin remodelling phase is the most plausible. During this phase, chromatin remodelling proteins unwind condensed DNA and deposit marks on histones that allow subsequent recruitment of repair factors (section 1.2.4). Drosha may have a part to play at this stage, meaning in its absence the chromatin is not easily accessible by factors such as 53BP1. These conclusions were further strengthened by the observations that repair outcome by both NHEJ and HR is diminished in Drosha and Dicer depleted cells (Figure 3.13BC)

The lack of a role for the TNRC6 proteins (Figures 3.2, 3.9, 3.13) demonstrates that the effect observed in the absence of Drosha is not due to the activity of miRNAs. This statement was further backed up by miRNA activity assays, showing that loss of TNRC6 completely ablates miRNA let-7 activity whereas at 48 hours, siRNA depletion of Drosha had no effect (Figure 3.3D & 3.11B). The mechanisms of miRNA turnover is still relatively uncharacterised, however recent data suggests that this is slow for many miRNAs (Marzi et al. 2016), suggesting there may be enough residual miRNAs to carry out their function in the short time that Drosha is depleted. While there are many miRNAs that do have short half-lives, these findings show that they are not responsible

for the effects seen as the same phenomena would be observed in TNRC6 depleted conditions.

A lack of a role for Ago1 or Ago2 provides further evidence that canonical miRNA activity is not involved (Figure 3.13B). This may also suggest that Drosha and Dicer may not be producing small RNA as their mode of action in the DNA damage response, unless damage-induced small RNAs are incorporated into a different protein complex. This would be an important experiment to investigate function of small RNA.

As an alternative to IR, the AsiSI endonuclease system was chosen for several reasons. As DSBs are introduced differently, it allows validation of the results observed previously; and as AsiSI only produces DSBs with no off-target effects, shows that these observations are specific to DNA damage. Most importantly however is that it is a very well-documented endonuclease system and causes many breaks in regular and endogenous genomic settings (Aymard et al. 2014; Aymard et al. 2017). As Drosha and Dicer are RNA processing enzymes, it is likely they are involved in processing an RNA at DSBs. The AsiSI system is therefore an excellent system to induce DNA damage, as RNA species can be determined by next generation sequencing approaches and mapped back to the known genomic loci. As this is the primary focus of the next two chapters, it was fundamental that the results in A549 cells were reproducible in U2OS-AsiSI cells. This system is not without its caveats however; damage induction is slower than IR, with maximal induction not seen until an hour after addition of 4-OHT (Figure 3.8A). Damage is clearly observable 30 minutes after induction, as is the translocation of AsiSI to the nucleus (Figure 3.8A). This suggests that the rate-limiting step of damage induction is the translocation, and that damage at different sites is occurring at different rates. Another issue could arise from any leakage of the enzyme into the nucleus during standard cell culture. This would induce DSBs and potentially slow cellular growth, and even lead to mutation and loss of some DSBs due to error-prone repair. Steps to prevent this were undertaken, by culturing in charcoal-treated media lacking hormones that may activate the ER, as well as keeping many early cell passage stocks to ensure maximal damage induction.

This chapter has demonstrated a role for Drosha and Dicer in DNA repair outside of the miRNA pathway. As these are RNA processing enzymes, the likely mechanism for these in the DDR is through RNA metabolism. As previously reported, damage within

reporter constructs and repetitive loci led to the production of small RNA from the adjacent sequences. In the next chapter, multiple NGS experiments were carried out to establish whether this is a feature of DSB repair at endogenous loci.

4 Chapter 4: An exhaustive RNA sequencing strategy to identify damage-induced transcription products

4.1 Introduction to chapter

In the previous chapter a non-canonical role for Drosha and Dicer in DNA repair was established. Knocking down Drosha had no effect on early DDR signalling but did prevent recruitment of many repair factors and reduced repair efficiency. This effect was shown to be distinct from that of the processing of miRNAs, as the knockdown of TNRC6A-B and thus abolishment of the miRNA pathway did not have the same consequence. However, the exact role of Drosha in DNA repair is still to be determined. Drosha has been demonstrated to participate in several other pathways such as transcriptional activation and termination splicing (Gromak et al. 2013; Dhir et al. 2015). In common between all these activities is that Drosha is binding to and processing in some way an RNA molecule. It is therefore most likely that the role for Drosha in DSB repair would be related to RNA processing. Indeed, several publications have reported the production of small RNA products from the sequences around DSBs (Francia et al. 2012; W. Wei et al. 2012; Michalik, Böttcher, and Förstemann 2012). This necessitates the initiation of RNA transcription and the production of new RNA products at the site of damage, and was proposed to require the miRNA biogenesis enzymes (Francia et al. 2012; W. Wei et al. 2012). This has been controversial, in part due to the consensus that transcription is inhibited around breaks (see section 1.5.3) (Shanbhag et al. 2010; Pankotai et al. 2012; Manfrini et al. 2015; Solovjeva et al. 2007; Iacovoni et al. 2010; Iannelli et al. 2017) and in part due to the artificial systems used to investigate these. In *Arabidopsis* and humans, studies have utilised exogenously integrated reporters that are either remarkably transcriptionally active or highly repetitive (W. Wei et al. 2012; Francia et al. 2012). Attempts to look at endogenous loci have been restricted to the ribosomal repeat locus in *Neurospora* (Lee et al. 2009), or the $(TTAGGG)_n$ telomeric repeat locus (Rossiello et al. 2017).

In this chapter, a comprehensive series of NGS experiments were conducted to attempt to ascertain whether new transcription and small RNA products are a feature of DSB repair at endogenous loci. This utilised the AsiSI endonuclease system (Aymard et al. 2014) as established in Chapter 3, which cuts at around 100 loci in the genome.

4.2 Small RNA sequencing within the U2OS-AsiSI cell line did not reveal any new small RNA species from around DSBs.

To investigate whether small RNAs can be produced from around DSB sites within the endogenous genomic setting, the U2OS-AsiSI system was utilised to induce defined breaks. A simple pilot experiment was performed first to establish whether these could be observed. U2OS-AsiSI cells were treated with either 300 nM 4-OHT or an equal volume of DMSO for 6 hours, then RNA harvested using TRIzol alongside protein extraction. DSB induction was validated by western blot (Figure 4.1A). Small RNA-seq libraries were then produced using an adapted library protocol (Xu et al. 2015). This method is detailed fully in section 2.2.4: critically this method utilises a degenerate 4 nucleotides at the ligating ends of each adapter which reduces ligation bias as a result of secondary structure. Following ligation and PCR amplification, libraries were run on TBE-polyacrylamide gels and libraries of the required size (with adapters, these are around 150 bp for 21 nt small RNA) were excised and eluted (Figure 4.1B). A final, pure library of the correct size was validated by a second gel and by Bioanalyzer analysis (Figure 4.1CD). These libraries were pooled at equal concentrations as determined by the Bioanalyzer (Figure 4.1D) and deep sequenced.

Small RNA reads were aligned to the human genome (latest Ensembl release GRCh38 (Yates et al. 2016)) using bowtie2 (Langmead and Salzberg 2012) (section 2.5.2). Aligned reads for each sample were extracted and coverage across the genome (binned into 10 kb windows) was calculated and plotted against each other (Figure 4.2). This analysis showed that there was very little variation in the small RNA profiles at 6 hours following DSB induction by AsiSI.

4.3 Setting up a plasmid based AsiSI system

Damage-induced small RNAs have been reported in several publications (Francia et al. 2012; W. Wei et al. 2012; Michalik, Böttcher, and Förstemann 2012). As these utilised exogenously integrated reporter systems and repetitive sequences, it may be that the increased copy number of DNA sequence or high RNA expression led to those observations. However, with the data that Drosha and Dicer are required for DNA repair (Chapter 3) it stands to reason that they may be required to process small RNA around DSBs. Why these were not detected in the pilot experiment could be due

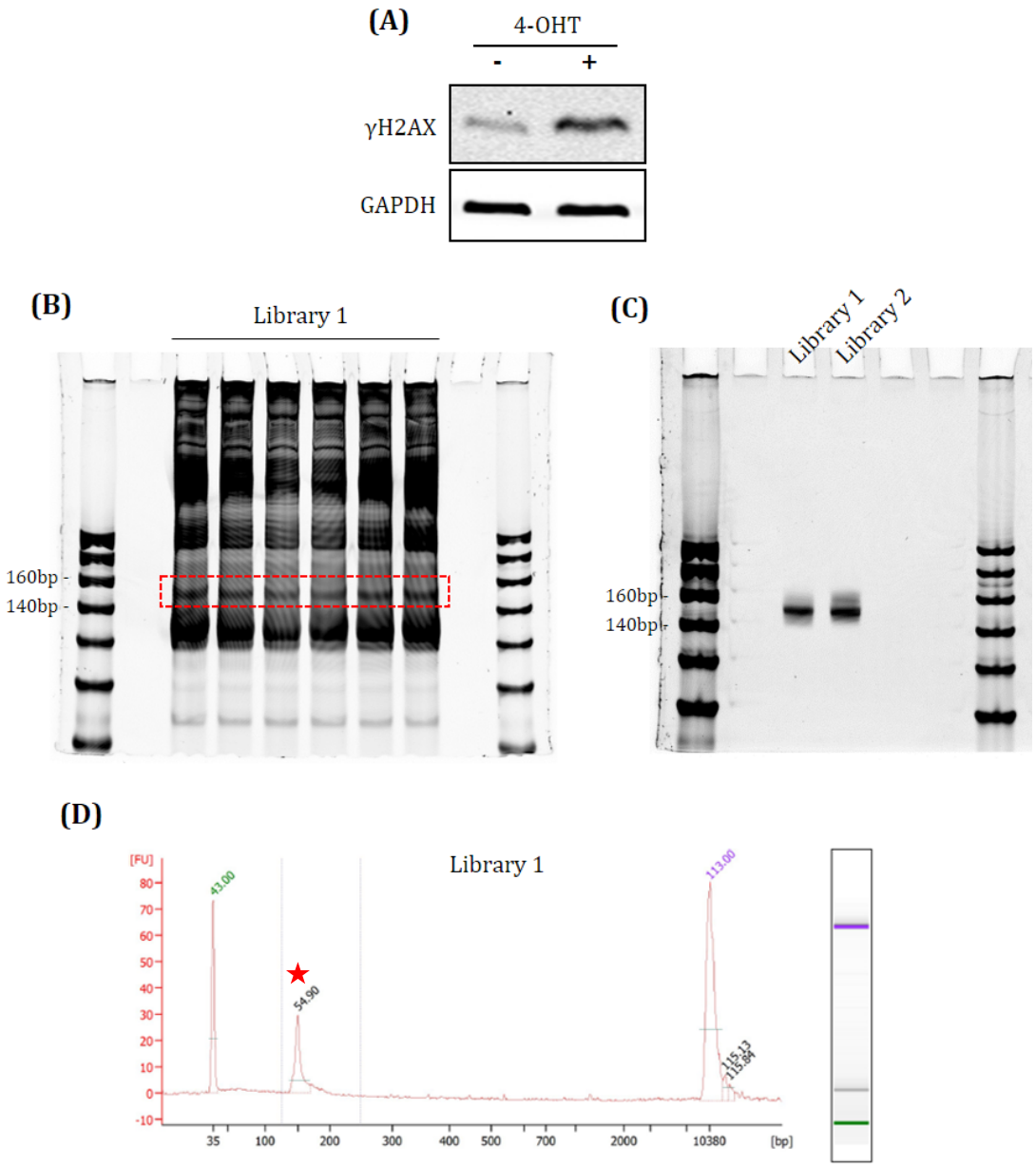


Figure 4.1. Small RNA-seq library preparation for pilot experiment.
(A) U2OS-AsiSI cells were treated with 300 nM 4-OHT to induce damage by AsiSI, or an equal volume DMSO as control. 6 hours later cells were harvested for RNA and protein. Immunoblotting with γH2AX confirms damage induction. **(B)** Following adapter ligation and PCR amplification, final libraries were split over multiple lanes in TBE-polyacrylamide gels to prevent overloading. The small RNA library band at ~150 bp was excised as the red box. Example of one library. **(C)** 1 µl of excised library was run on a TBE-polyacrylamide gel to show purification of final products. **(D)** High sensitivity Bioanalyzer trace of the libraries in (C) (red star). Libraries were pooled at equimolar concentrations based on this quantification.

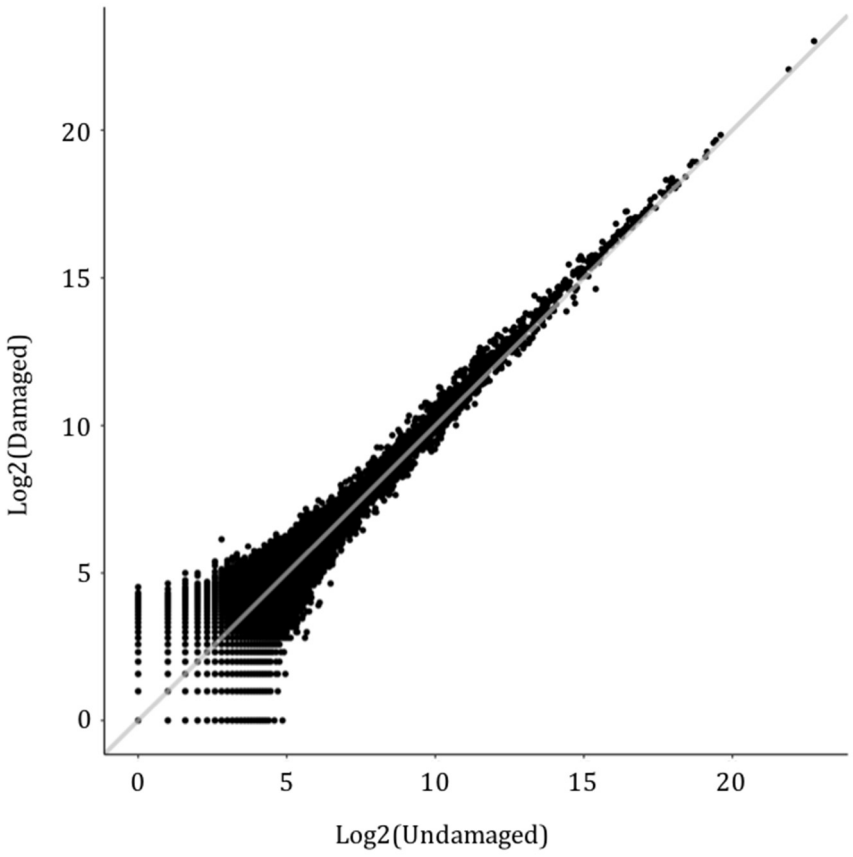


Figure 4.2. Pilot small RNA-seq data do not show any difference in mapped reads across the genome after damage.

The human genome (GRCh38) was binned into 10 kb windows and small RNA read coverage counted within each window. The log2 counts in the undamaged and damaged conditions were plotted against each other. Line x=y in light grey.

to the potential leaking of the ER-tagged AsiSI into the nucleus. As small RNAs can have very long half-lives (Marzi et al. 2016), it might be that background damage produces enough small RNAs that persist to mask small RNAs produced on 4-OHT induction. Alternatively, the 6 hour time point may be too late and any small RNAs produced may be rapidly turned over. Arguing against this possibility however is the time points utilised by the groups that had detected small RNA in their reporter systems, where the earliest time was 12 hours following damage induction (W. Wei et al. 2012; Francia et al. 2012; Michalik, Böttcher, and Förstemann 2012). To address all these possibilities, a follow-up small RNA-seq experiment was devised to contain many time points following 4-OHT (0, 1, 6, 12, 24 hours) and to also utilise a transfected high-expression ER-AsiSI vector (pCI-AsiSI) and an empty vector (EV) control in regular U2OS cells (Figure 4.3A). The EV control serves to provide an AsiSI-free background; as these cells were never exposed to the specific AsiSI DSBs, none of the small RNAs sequenced would be due to AsiSI cleavage.

U2OS cells were transfected with pCI-AsiSI or pCI-EV by electroporation. 18 hours later, cells were treated with 300 nM 4-OHT or an equal volume of DMSO for the EV and 0 hour controls. At each time point (Figure 4.3A) cells were harvested to obtain protein and RNA, and cells were also fixed on coverslips for IF. DNA damage induction was confirmed by western blotting for pATM (Figure 4.3B). This also confirmed that the 0 hour control (pCI-AsiSI transfected + DMSO) had a comparable amount of background DNA damage to cells devoid of the endonuclease. IF provided an alternative method of validating DNA damage induction by γH2AX staining, and to also observe AsiSI translocation to the nucleus (Figure 4.3C).

4.4 Analysis of time course

Two biological replicates of the transfection-based time course experiment were performed and small RNA-seq libraries were generated. Libraries of the correct size were gel extracted and pooled in equimolar amounts and deep sequenced (Figure 4.4). Small RNA reads were mapped to the human genome (GRCh38) using bowtie2. Mapped reads were extracted and then mapped to miRNA hairpins (miRBase release 21 (Kozomara and Griffiths-Jones 2014)) to identify any miRNAs and isomiRs . To test that DNA damage in this system caused general and expected changes to the transcriptional landscape, differential expression analysis of miRNAs over time was

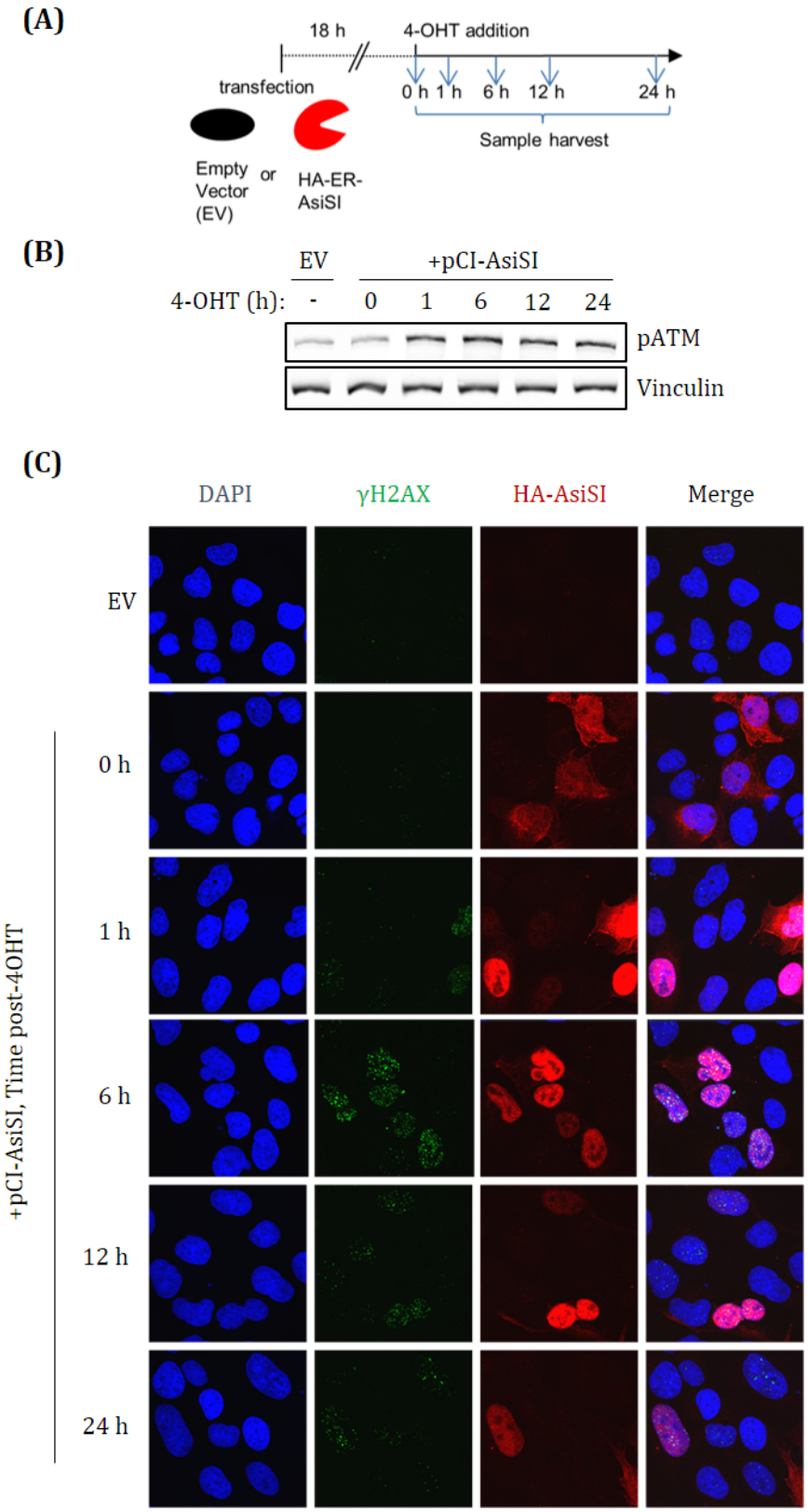


Figure 4.3. Electroporation of a mammalian expression AsiSI into U2OS cells can induce DNA damage.

(A) Schematic of experimental setup. 18 hours after electroporating AsiSI or an empty vector control (EV), 4-OHT is added and cells harvested at the

indicated time points. Both RNA and protein was harvested and in parallel cells on coverslips were processed for IF. **(B)** Representative western blot showing DNA damage induction by 4-OHT addition. **(C)** IF staining for γH2AX in green, HA-AsiSI in red, and nucleus (DAPI) in blue. The AsiSI enzyme is translocated into the nucleus within 1 hour and DNA damage foci can be seen.

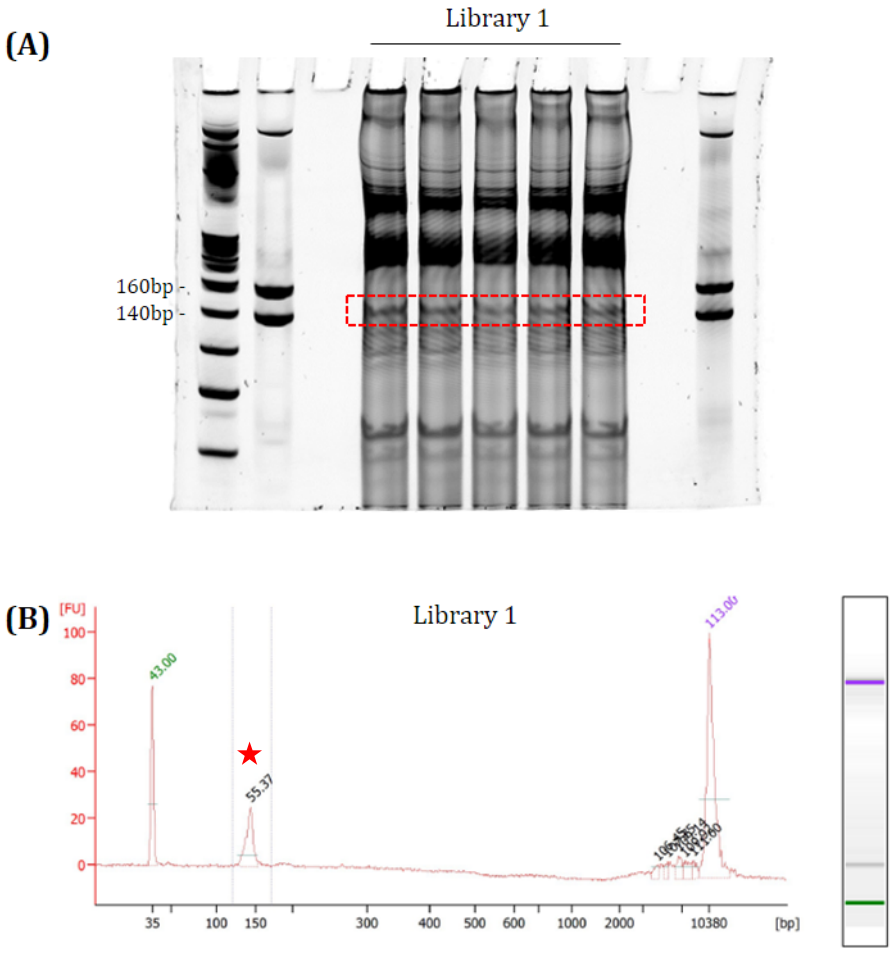


Figure 4.4. Small RNA library preparation for DNA damage time course experiment.

(A) A representative TBE-polyacrylamide gel image of a completed sRNA library split over several wells. The approximate slice cut out of the gel for elution is indicated by the red box (small RNA library ~150 bp). **(B)** Bioanalyzer trace of the same sample as in (A) showing a final library of the expected size (150 bp; indicated by red star). The concentration determined here was used to pool all samples at equal concentrations for multiplexed sequencing.

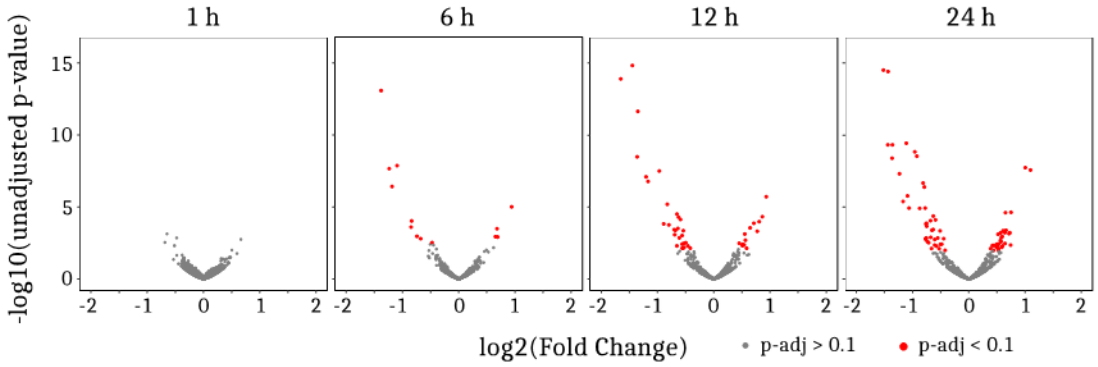
performed using DESeq2 (section 2.5.3) (Love, Huber, and Anders 2014). The default adjusted p-value (p-adj; false discovery rate (FDR)) of 0.1 was used for significance cut-off. This analysis found that no miRNAs were significantly differentially expressed 1 hour following 4-OHT addition, with 8 miRNAs passing the significance threshold at 6 hours, 42 at 12 hours, and 71 at 24 hours (Figure 4.5A). This also revealed that most differentially expressed miRNAs have decreased abundance following DNA damage. Performing KEGG pathway analysis (DIANA-MirPath v.3 (Vlachos, Zagganas, et al. 2015)) on the differentially expressed miRNAs at 24 hours showed an enrichment for pathways involved in cell proliferation (Figure 4.5B). This analysis takes all validated miRNA:mRNA interactions (as maintained by TarBase (Vlachos, Paraskevopoulou, et al. 2015)) and performs pathway analysis on the mRNA targets to identify network changes due to miRNA differential expression. Splitting this list into up- and down-regulated miRNAs showed similar pathway enrichment between the two groups; interestingly, analysis of up-regulated miRNAs revealed an enrichment of HR pathway and DNA replication genes.

This dose dependent effect of DNA damage on miRNA regulation (Figure 4.5A) highlights how miRNAs are not utilised as a rapid means of regulating gene expression. Figure 4.6 shows the fold change of a subset of differentially expressed miRNAs over time, which demonstrates how it takes several hours for noticeable differences in miRNA levels to occur. For example, a well-studied DNA damage response miRNA, miR-34a (Kato et al. 2009; Zhou et al. 2014), only reaches 2-fold expression 24 hours after DNA damage. The abundances of miRNAs in the EV samples was also used to compare to the undamaged control of a previously published dataset (Francia et al. 2012). As the compared dataset had detected damage-induced small RNA, this was done to determine whether this sequencing experiment had a similar degree of sensitivity. This showed that the ability to detect small RNA, and importantly low abundance miRNA, was at least as sensitive (Figure 4.7).

4.5 Analysis of non-miRNA small RNAs does not provide any evidence for small RNA generation around DNA breaks.

Having established that AsiSI-induced DNA damage can induce changes to miRNA expression after several hours, analysis was performed to identify damage-induced

(A)



(B)

KEGG pathway	p-value
MicroRNAs in cancer	7.99E-53
Cell cycle	7.61E-13
Proteoglycans in cancer	1.51E-11
Protein processing in endoplasmic reticulum	6.62E-11
Renal cell carcinoma	4.18E-10
Ubiquitin mediated proteolysis	4.82E-09
Adherens junction	1.63E-08
Endocytosis	1.80E-08
Pancreatic cancer	6.95E-07

Figure 4.5. Differentially expressed miRNA occurs many hours after DNA damage induction.

(A) miRNA reads were counted using HTSeq-count and tested for differential expression using DESeq2, with a design formula looking for changes over time. The log2 fold change at each timepoint following 4-OHT induction contrasted to the 0 hour time point in plotted against the negative log10 unadjusted p-value for this fold change. Conditional highlighting of points with an FDR/adjusted p-value of less than 0.1 show statistically significant differential expression. **(B)** The 71 significant miRNA at 24 hours were used for KEGG pathway analysis, using only validated miRNA-mRNA interactions. The top 10 enriched KEGG pathways are shown.

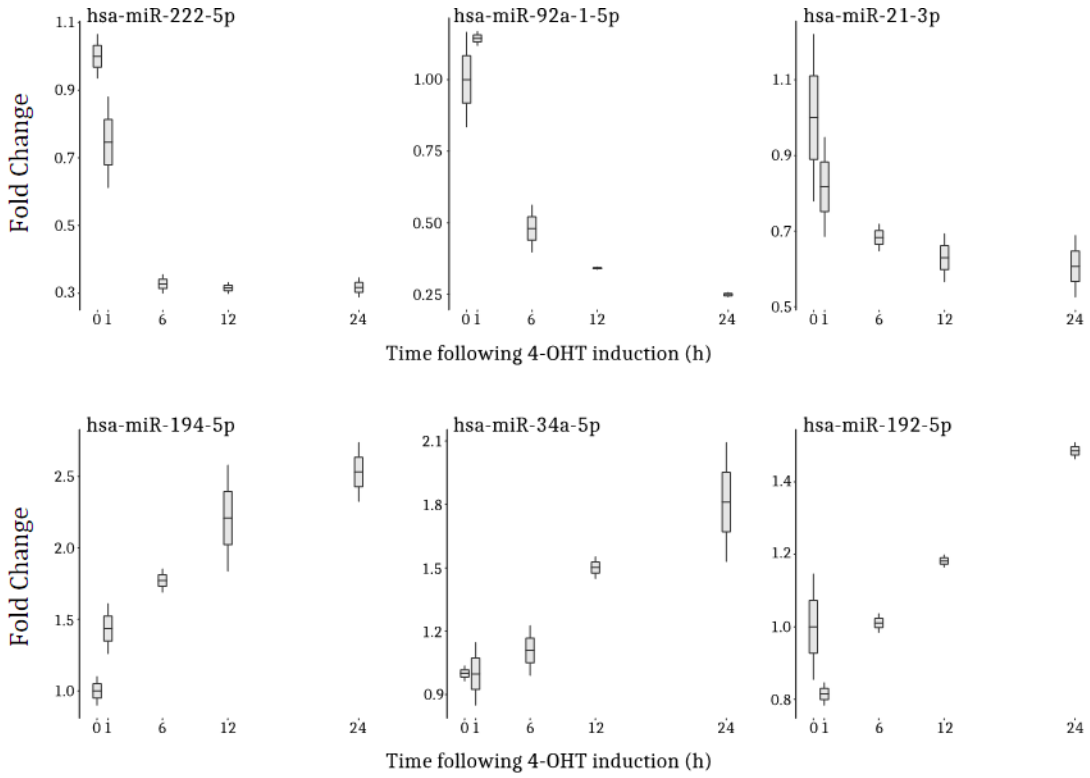


Figure 4.6. A subset of differentially expressed miRNAs show a dose-dependent effect of DNA damage on miRNA expression over time.
Normalised counts for the three most significant upregulated and downregulated miRNAs over time were extracted from the DESeq2 object. This was normalised to the counts at timepoint 0 (undamaged) and plotted on a continuous x-axis to highlight the dose-dependent effect of DNA damage of miRNA expression.

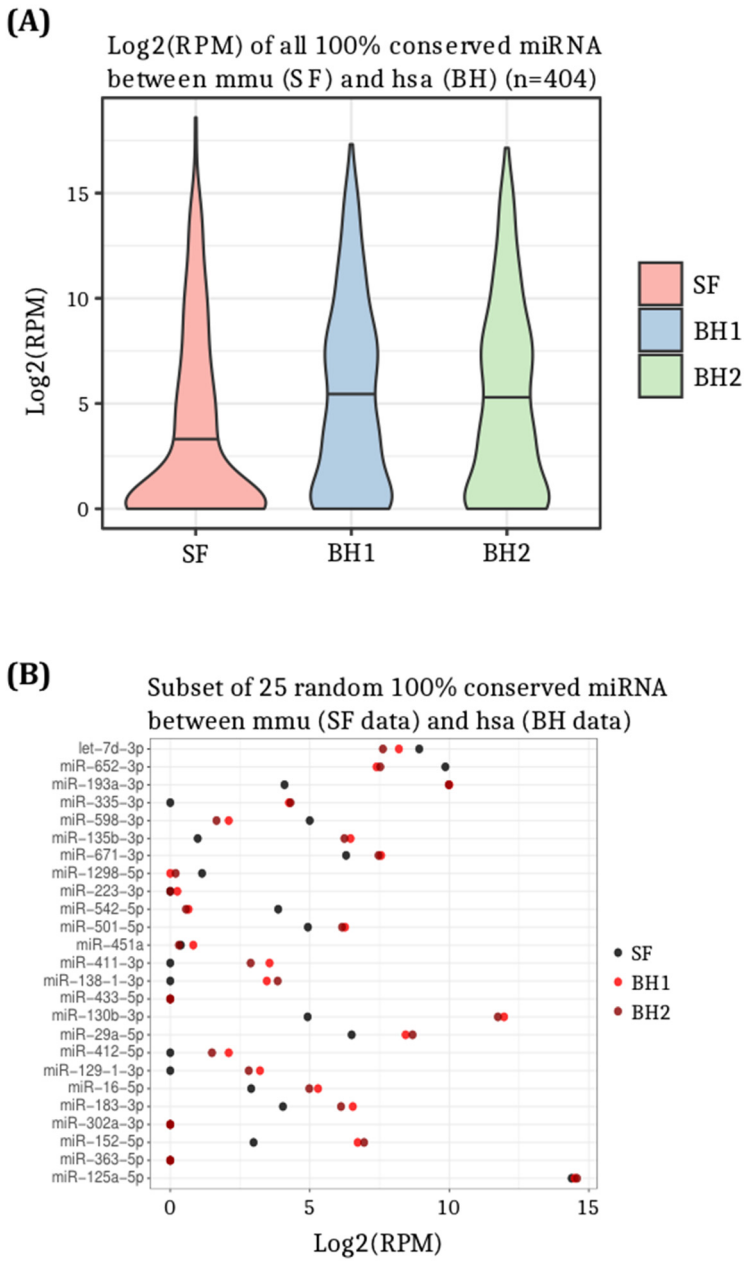


Figure 4.7. Comparison with a published dataset shows detection of low abundance miRNA is at least as sensitive.

(A) The single replicate from dataset in (Francia et al. 2012) was compared to the two biological replicates produced in this project (BH1 and BH2). As SF data was in mouse, the 100% conserved miRNA were compared. Log2 reads per million (RPM) of undamaged condition was plotted in a violin plot showing better coverage of low abundance miRNA. **(B)** A random subset of 25 miRNAs were extracted and RPM for each replicate was plotted.

small RNA around AsiSI cut sites. Small RNA reads that did not align to miRNA hairpins were then further filtered by removing other small RNAs such as snoRNA, tRNA, and YRNA as annotated by DASHR (Leung et al. 2016) and within Ensembl release GRCh38.85 (Yates et al. 2016). The remaining reads were therefore devoid of any known small RNA gene products and these were used in further analyses (Figure 4.8). Unlike with the pilot experiment (section 4.2), the presence of repeats and multiple conditions allows a very robust statistical model to be built, meaning even small but significant changes in the small RNA population occurring around AsiSI cut sites following induction of the endonuclease can be detected. A custom GTF file was created using the coordinates 5 kb upstream and downstream around every AsiSI motif; these 10 kb regions were approximately equal in size to the exogenously introduced loci in two previous publications (Francia et al. 2012; W. Wei et al. 2012). As the number of loci that are known to be cut is 99 (Aymard et al. 2014), this leaves over 1000 sites that can act as a control where no changes should be observed, which is important for the GLM in DESeq2 which assumes most units tested should not be differentially expressed (section 2.5.3). HTSeq-count (Anders, Pyl, and Huber 2015) was used to generate raw count tables for each of these loci which could then be used by DESeq2. This analysis was performed twice using different design formulae; first to determine differential expression over all time points as in section 4.4, and secondly using a likelihood ratio test and a reduced model to identify any changes at any timepoint regardless of the other timepoints. This was done to ensure any significant changes were not missed as no assumptions of whether there would be a dose dependent effect of time could be made. This was performed for all small RNA (17-30 nt), as well as those that were 21-23 nt as previously reported (Francia et al. 2012; W. Wei et al. 2012); for completion, other subset size classes were included. Both these analyses found no significant differences at any time point at any locus, across any size group tested (p-adj>0.1) (Figure 4.9).

To better visualise this, the region 1 kb upstream and downstream of each of the 99 cut AsiSI sites was binned into 10 bp windows and coverage per million reads calculated in each. The mean window coverage across all sites for the plus and minus strands was then plotted, comparing each time point alongside the EV condition (Figure 4.10, left). Whilst there does appear to be a build-up of reads around the AsiSI

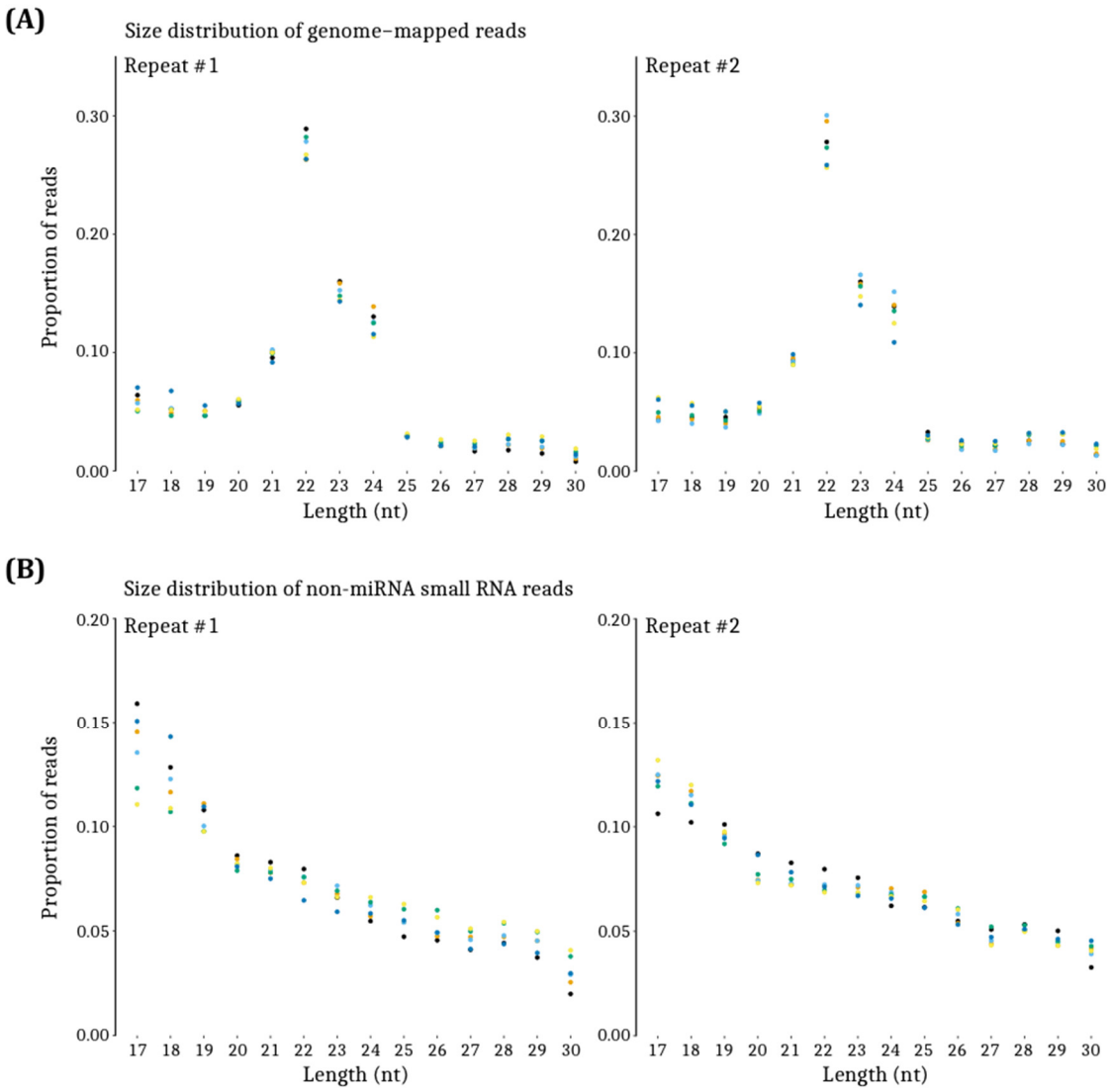


Figure 4.8. The size distribution of genome mapping reads before and after removal of miRNA reads show that miRNAs make up most of small RNA sequencing runs.

(A) Size distribution of all genome-mapping small RNA reads (17-30 nt) was calculated using custom scripts. Left, replicate 1; right, replicate 2. miRNA are typically 21-24 nt. Proportion calculated as (number of reads at size)/(total number of reads) **(B)** Size distribution of all genome-mapping reads following filtering of out of miRNA.

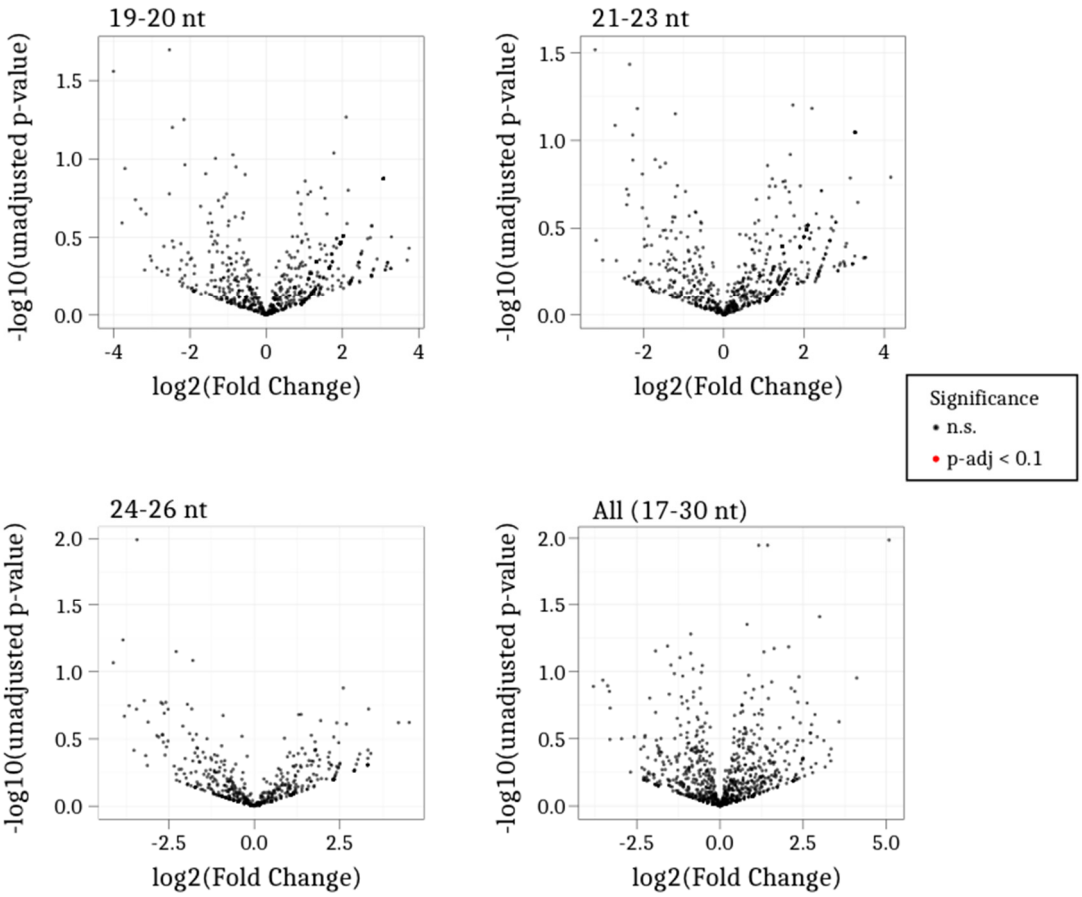


Figure 4.9. No changes to the small RNA population around AsiSI sites could be detected upon DNA damage using DESeq2.

Small RNA reads depleted of all sncRNA were binned into the indicated size groups. Those mapping in a 10 kb region around every AsiSI motif were tested for differential expression using DESeq2. This testing was designed to look for changes over time, or at any timepoints. Volcano plots were generated from the result of this analysis (showing 0 hour against 24 hour conditions), with conditional highlighting of any points with an FDR/p-adj<0.1. No sites were significant in any size class.

cut site, this also appears in the EV condition where the cells were never exposed to AsiSI. This is potentially due to the enrichment of AsiSI sites at TSSs (Figure 3.6B), indicating that the sequenced non-miRNA small RNA may be PASR/TSSaRNA (see section 1.4.1.1) or degradation products of pre-existing mRNA. Furthermore, there does not appear to be any pattern in the peaks over any time points indicating this is just variability. To address this, a set of random control sites in the same approximate position within gene units with similar transcriptional activity was created by utilising the 4sU-seq dataset discussed later and explained in detail in section 2.5.6 and listed in Table 8-6. Mean coverage was calculated and plotted as before (Figure 4.10, right). Like the AsiSI cut sites, these random sites have a build-up of reads in both EV and post-damage conditions, and there appears to be variability between time points rather than any pattern. These data demonstrate that DNA damage-induced small RNAs are not produced around DSBs within non-repetitive, endogenous genomic loci.

4.6 Development of a method to sequence 5'-triphosphate small RNA

DNA damaged-induced small RNAs were detected previously in a reporter system and repetitive loci in human cells, but also from reporter systems in fungi, plants, and Drosophila (see section 1.5.2). In these other organisms, there exists other small RNA pathways such as the RNA silencing pathway. A critical component of these pathways that does not exist in humans is the RNA-dependent RNA polymerase (RdRP), which can amplify many copies of small RNA from a single or small number of original transcripts (Dalmay et al. 2000; Lipardi and Paterson 2009). In *A. thaliana*, an RdRP was shown to be required for DNA damage-induced small RNA production (W. Wei et al. 2012). Certain mammalian polymerases have been demonstrated to possess minor RdRP activity (Wagner et al. 2013), and it could be possible that in a DNA damage context this activity may be utilised to generate small RNA, although this activity has not been demonstrated. One hallmark of these amplified products is the presence of a 5'-triphosphate instead of the 5'-monophosphate carried by miRNAs (Sijen et al. 2007; Pak and Fire 2007). In small RNA-seq library preparation, the 5'-monophosphate is exploited to specifically capture miRNAs which means alternatively capped small RNAs are not sequenced. It could be that previously reported small RNAs were fortunate to

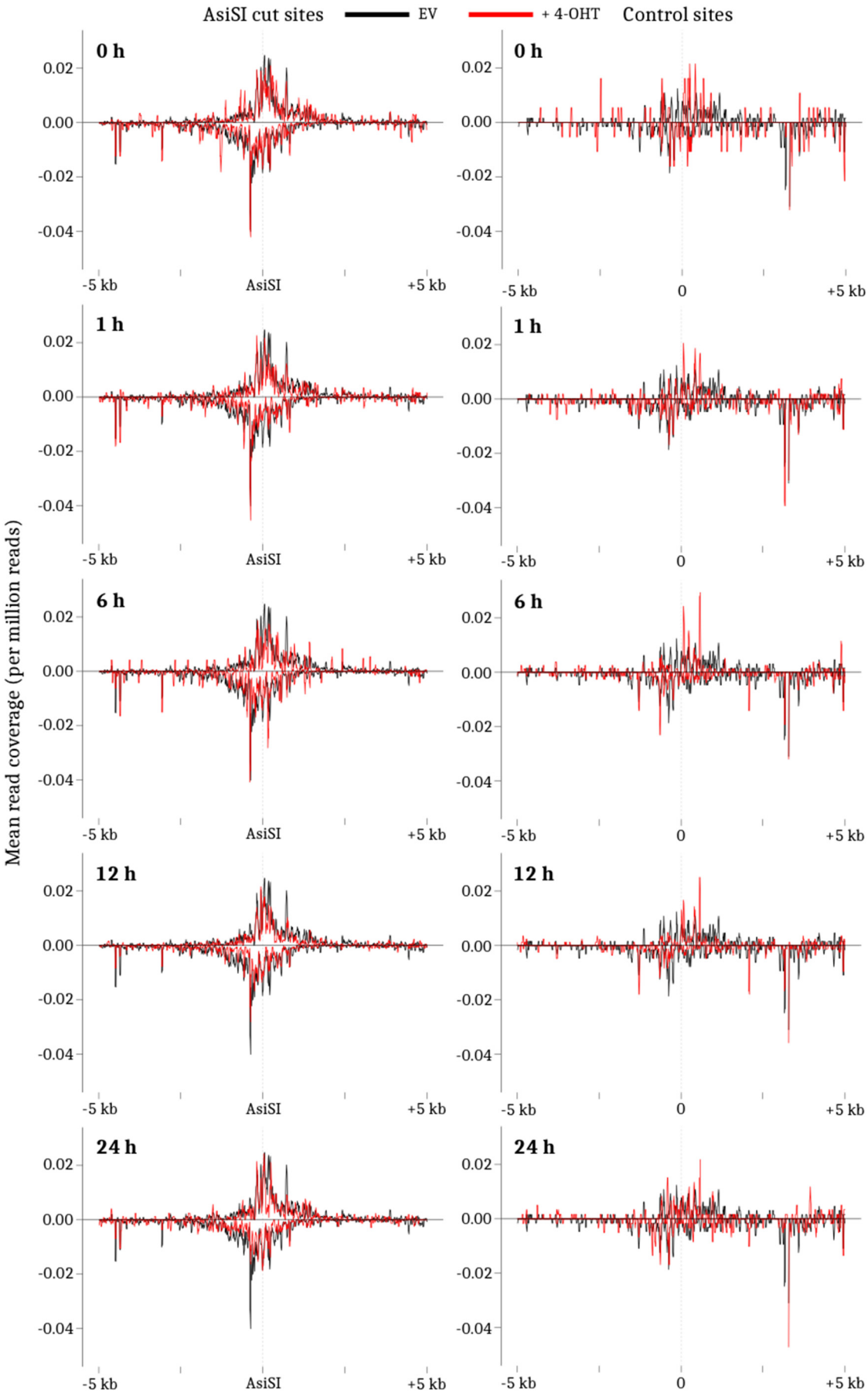

AsiSI cut sites
EV
+ 4-OHT
Control sites
Mean read coverage (per million reads)
0 h
1 h
6 h
12 h
24 h
0.02
0.00
-0.02
-0.04
-5 kb
AsiSI
+5 kb
0

Figure 4.10. No changes to small RNAs at any AsiSI cut site at any time point after damage.

Mean coverage in 10 bp bins of non-miRNA small RNA reads was calculated in the regions 5 kb either side of the 99 cut AsiSI sites (left) or the same number of transcriptionally matched control sites (right). Each graph by row is a different timepoint in the pCI-AsiSI transfected conditions compared to the empty vector (EV) control. Timepoint signal is in red, EV in black. As EV has never been exposed to AsiSI, peaks in this sample cannot be due to AsiSI-induced DSBs.

Positive values for read coverage are those reads that map to the plus strand of the genome, conversely, negative values are those that map to the minus strand. N=2.

capture these 5’-triphosphate RNAs through chance dephosphorylation because of cell lysis or another mechanism. To investigate the existence of an undescribed RNA amplification pathway in humans, a strategy to sequence 5’-triphosphate small RNA had to be devised (Figure 4.11). Conversion of any 5’-triphosphate moieties to 5’-monophosphate would allow capture of these using standard small RNA library preparation techniques. Dephosphorylation can be achieved through the action of many phosphatases, however 5’-polyphosphatase was chosen as this enzyme only removes the γ- and β-phosphates leaving the α-phosphate meaning an additional kinase step is not required. As 5’-triphosphate small RNAs may be of very low abundance, they could be lost if prepared alongside the abundant miRNAs. A 5’-monophosphate-specific RNA exonuclease was used to first remove miRNA from total RNA. The Terminator exonuclease does not affect levels of long or small capped RNA as determined by qPCR (Figure 4.12) at any concentration tested. However, it does efficiently process long or small RNA with 5’-monophosphates (Figure 4.12B). Calculating abundance of miR-21 relative to snoRNA U6 shows maximal degradation of the miRNA at even the lowest concentration of Terminator (Figure 4.12C). This method (described in section 2.2.3) of first degrading 5’-monophosphate miRNAs, followed by dephosphorylation of 5’-triphosphate RNAs (Figure 4.11) was then used to create small RNA libraries.

4.7 Analysis of triphosphate small RNA-seq does not reveal a human small RNA amplification pathway at DSBs

Total RNA samples as used for previous sequencing (sections 4.3, 4.4, 4.5) were processed using the method established in the previous section (Figure 4.11). Both biological replicates of the EV sample and 24 hour 4-OHT time point were processed and purified by ethanol precipitation. Small RNA libraries were then prepared as before, however at the point of gel extraction a band of the correct size could not be easily seen (Figure 4.13A). The approximate size range was processed regardless and a pure band of the correct size was observable by analysis on the Bioanalyzer (Figure 4.13B). These samples were pooled and subject to deep sequencing. Routine data quality analysis revealed a high degree of duplicity, potentially due to the low complexity of the sample suggesting that 5’-triphosphate small RNA may be very low

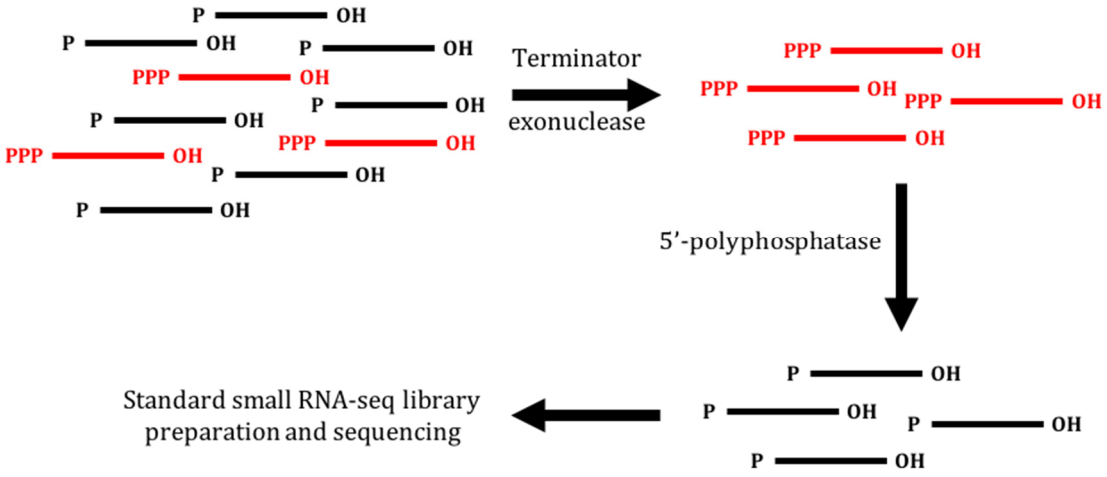


Figure 4.11. Schematic of 5'-triphosphate small RNA-seq strategy.
The Terminator exonuclease digests all RNA with 5'-monophosphate moieties (miRNA, rRNA), leaving any capped or 5'-triphosphate RNA intact. 5'-polyphosphatase removes γ and β phosphates, rendering the remaining 5'-triphosphates as 5'-monophosphates which can then be readily processed by standard small RNA library preparations.

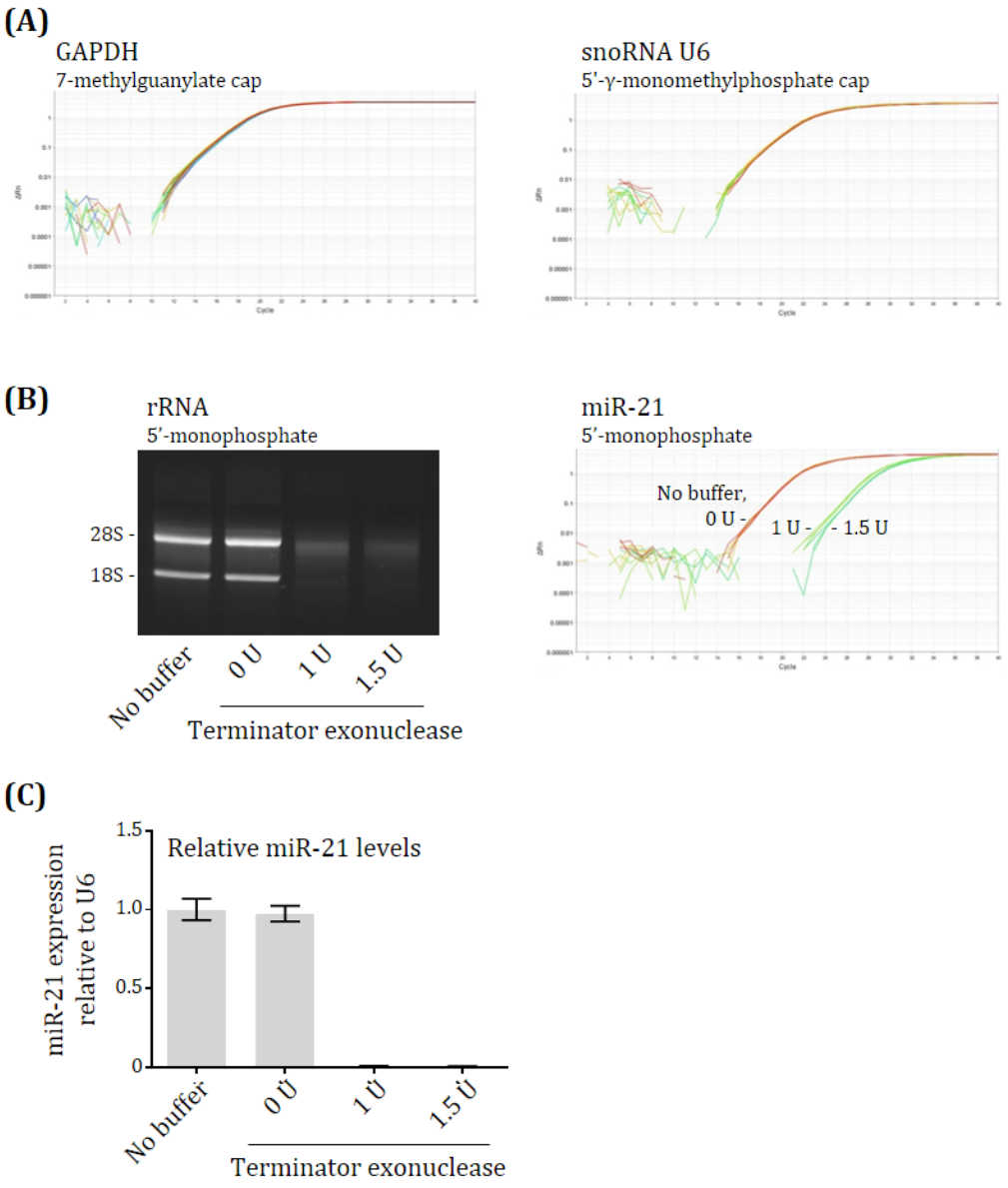


Figure 4.12. Optimisation of the 5'-monophosphate-dependent exonuclease Terminator shows complete and highly specific degradation of miRNAs.

(A) qPCR fluorescence against cycle number plots show capped RNA species (those without a 5'-monophosphate) are not degraded by Terminator exonuclease. **(B)** 5'-monophoshate RNA species are completely degraded by the exonuclease within 30 minutes. Left, 28S and 18S rRNA; right, miR-21 by TaqMan qPCR. **(C)** Quantification of TaqMan qPCR (as in B, right) for 5'-monophosphate miR-21 relative to 5'-capped snoRNA U6 shows complete depletion of 5'-monophoshate RNA species.

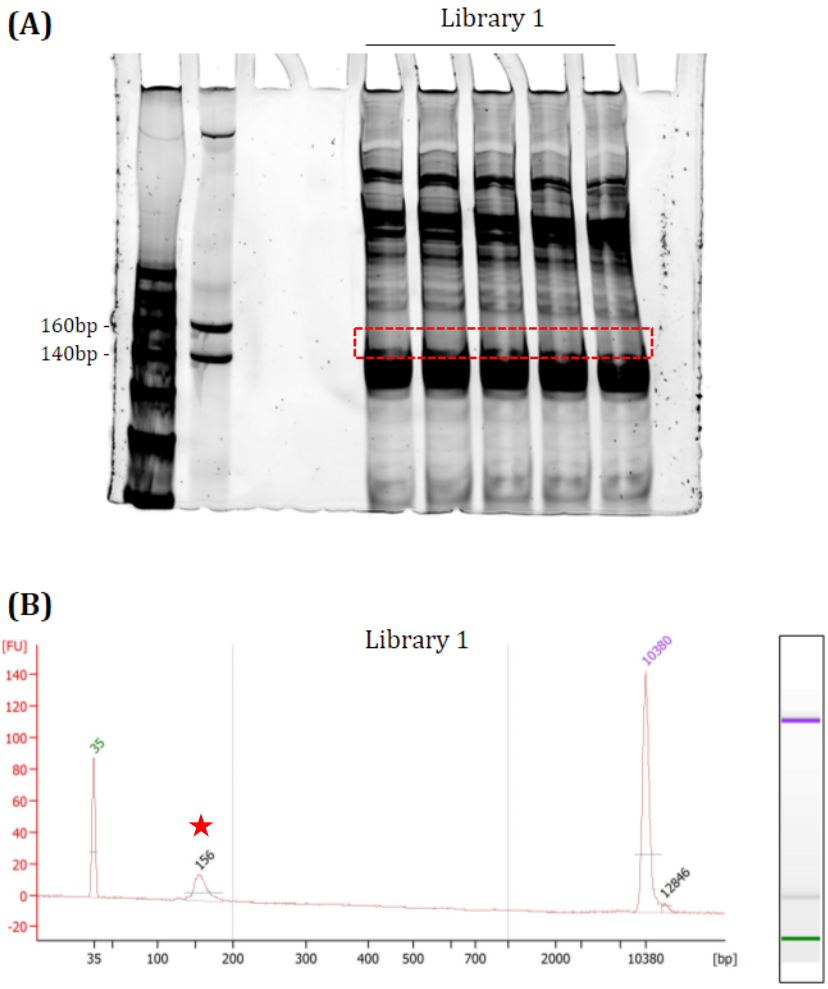


Figure 4.13. Small RNA library preparation for 5'-triphosphate small RNAs.

(A) Example TBE-polyacrylamide gel image of a 5'-triphosphate small RNA library split over several wells. A distinct band at the correct size is difficult to see unlike in Figures 4.1A and 4.4A. The red box indicates the size range cut out of the gel for library extraction. **(B)** Bioanalyzer trace of a completed library showing a library of the correct size (~150 bp, indicated by red star). Concentrations were determined from this peak and samples were pooled at equimolar concentration.

abundance within the cell. These duplicated reads were removed and the remaining reads aligned to the genome. Plotting the size distribution of mapped reads showed that no clear size class emerges (Figure 4.14A). This may suggest a lack of a distinct functional 5'-triphosphate RNA species. Reassuringly, removal of miRNA hairpin-mapping reads does not change this profile meaning that the 5'-monophophate degradation step prior to library preparation was very efficient and the majority of these reads are likely to be 5'-triphosphate small RNA (Figure 4.14B).

As 5'-triphosphate small RNA has not been described previously, a cursory analysis was performed to see where these may arise in the genome. Metagene analysis revealed that these may be primarily at the promoter/TSS of genes (Figure 4.15). This may suggest that this methodology may capture the 5'-most ends of newly synthesised RNA, as prior to addition of the 5' cap, mRNA transcripts have a 5'-triphosphate end (section 1.3.2).

To determine whether an increase in reads occurs around AsiSI break sites, the same type of DESeq2 analysis as in section 4.5 was performed. As secondary amplified products may be of variable sizes, this was performed with all reads (17-30 nt) as well as distinct size groups (19-20, 21-23, 24-26 nt) as before. As with the standard 5'-monophosphate small RNAs, no significant (FDR<0.1) changes could be detected (Figure 4.16).

Read coverage profiles around AsiSI sites were generated as in section 4.5 using the 5'-triphosphate reads (Figure 4.17A). This showed only a very small number of reads mapped to these regions, and that there does not appear to be any increase compared to the EV control. Like with the standard 5'-monophosphate small RNA-seq (Figure 4.10), a build-up can be seen around the AsiSI site even in the EV condition. Comparison to the transcriptionally matched control sites (Table 8-6) however demonstrate that this is a circumstance of the AsiSI sites being located close to TSSs (Figure 4.17B). These data provide no evidence that there is a RNA-dependent RNA polymerase activity at DSB sites in human cells.

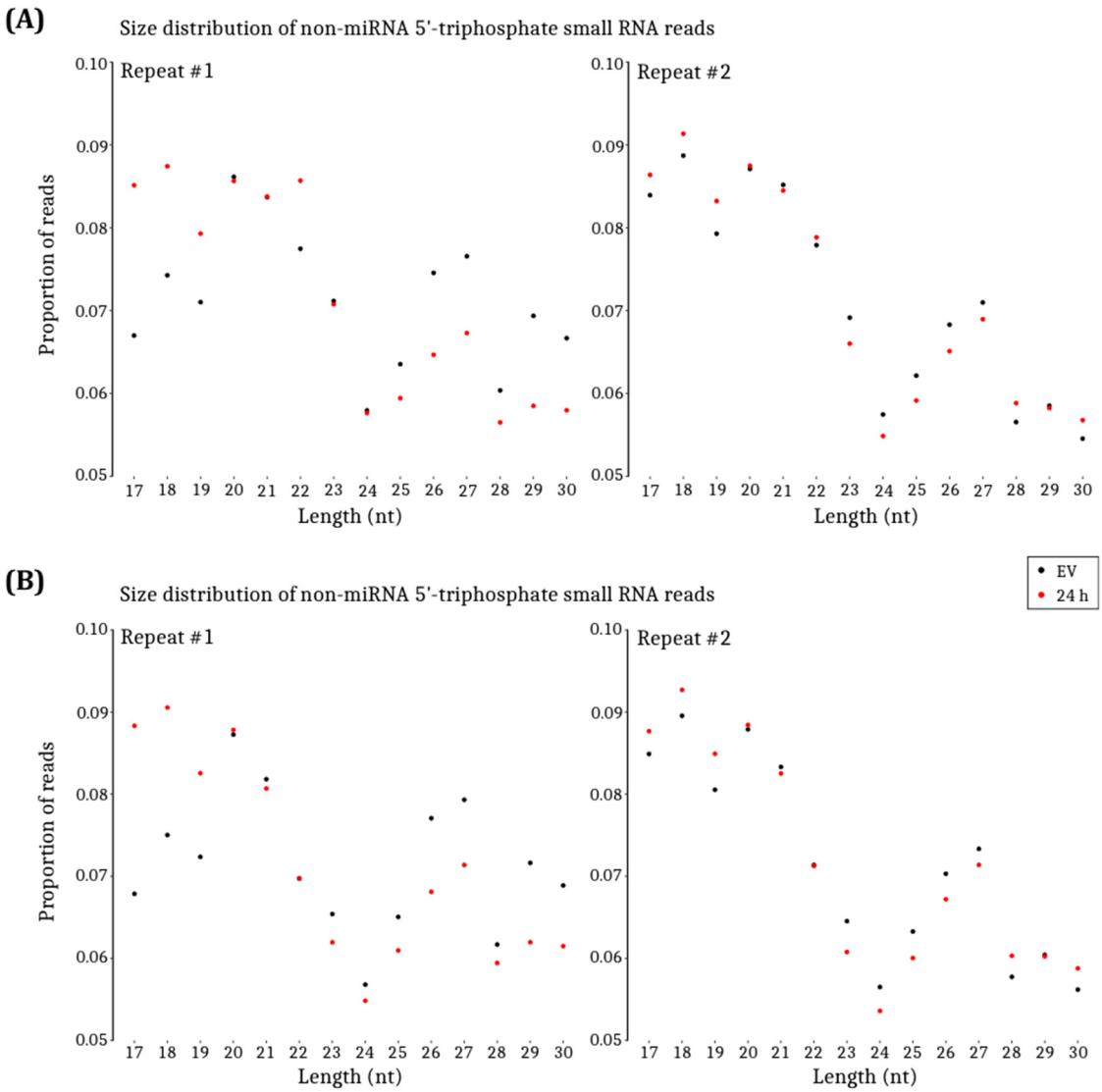


Figure 4.14. Size distribution of de-multiplexed 5'-triphosphate reads show no obvious size class indicating there may not be functional species of small RNA with 5'-triphoshphates.

(A) Size distribution of all genome-mapping reads 17-30 nt. Left, biological repeat 1; right is biological repeat 2. Proportion calculated as (number of reads at size)/(total number of reads) **(B)** As in A, using the reads that did not align to miRNA hairpins.

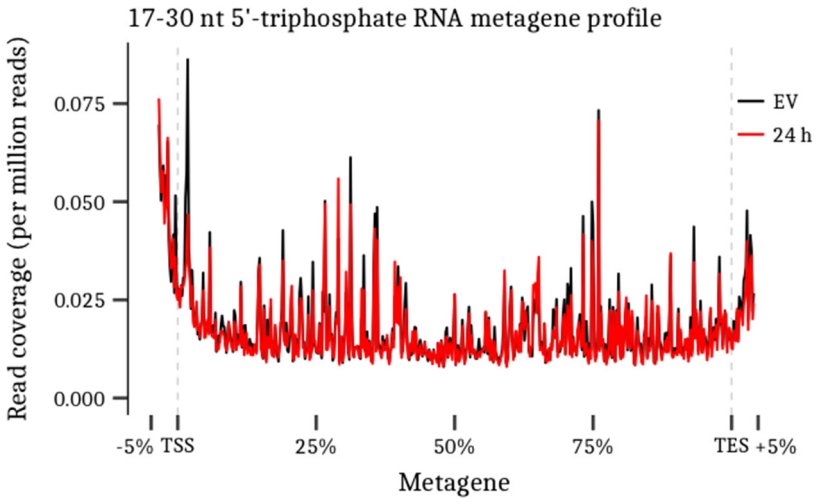


Figure 4.15. 5'-triphosphate small RNA reads are predominantly at the 5' end of transcripts.

All genes (>5 kb), plus 5% of the gene length upstream and downstream of the gene, was binned into 550 windows (i.e. 5 bins = 1% metagene). Coverage of 5'-triphosphate reads in each window was calculated and normalised to RPM. Each window was averaged across all genes to generate a metagene profile. Total number of genes was 22,467. Coverage is mean of two biological replicates per condition.

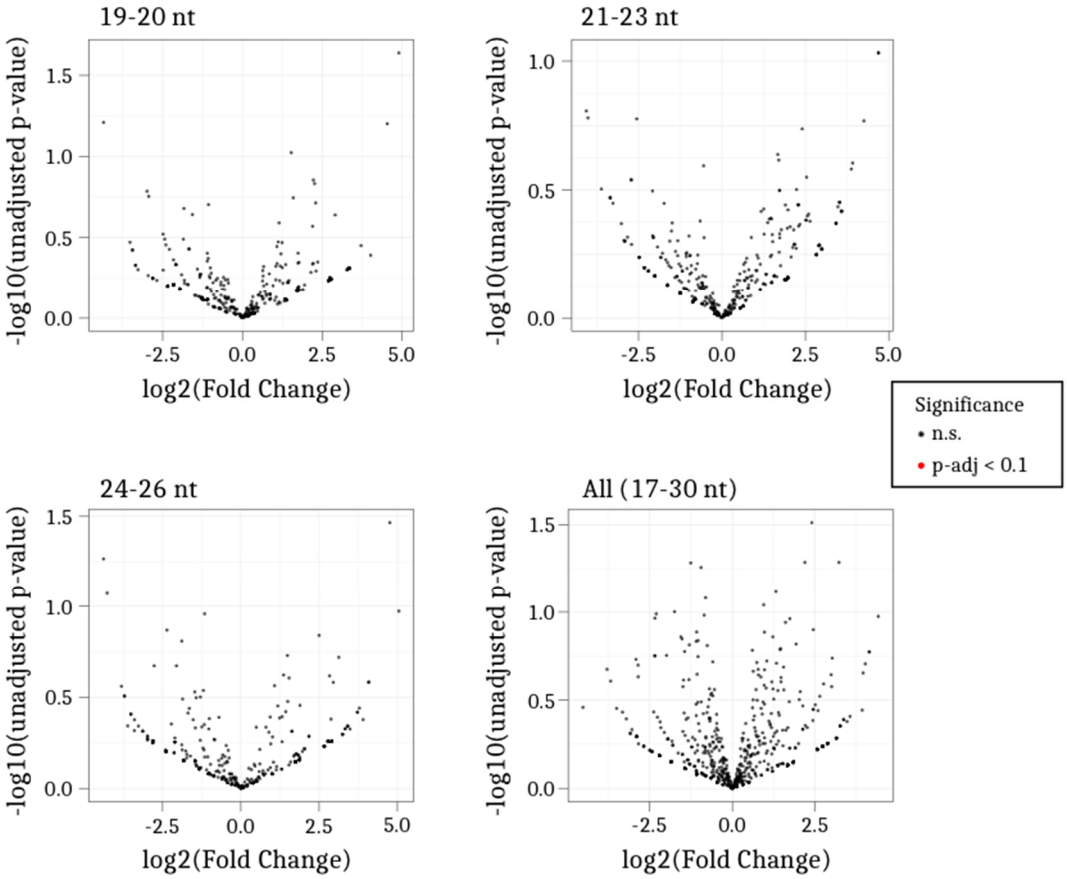


Figure 4.16. No significant changes to the 5'-triphosphate small RNA population mapping around AsiSI recognition motifs could be detected. 5'-triphosphate reads depleted of all sncRNA were binned into the indicated size groups. Those mapping in a 10 kb region around every AsiSI motif were tested for differential expression using DESeq2. Volcano plots were generated from the result of this analysis, with conditional highlighting of any points with an FDR/p-adj<0.1. No sites were significant in any size class.

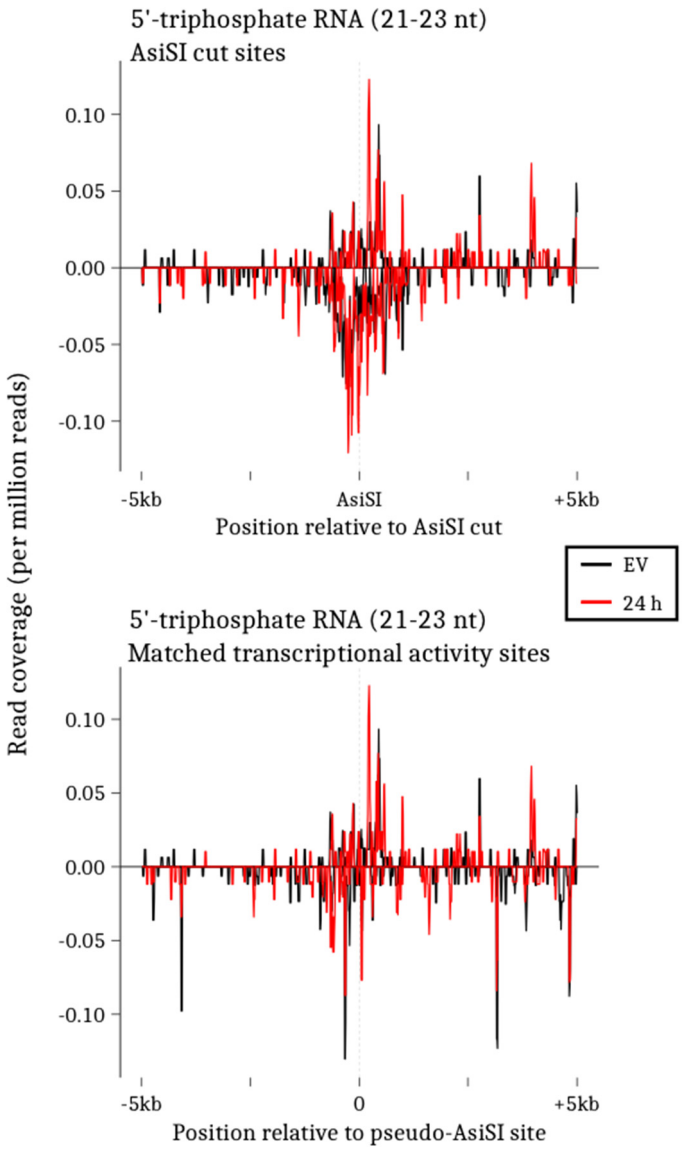


Figure 4.17. 5'-triphosphate RNA read density around AsiSI sites shows no changes.

Mean coverage in 10 bp bins of non-miRNA 5'-triphosphate small RNA reads was calculated in the regions 5 kb either side of the 99 cut AsiSI sites (top) or the same number of transcriptionally matched control sites (bottom). 24 hours post-4-OHT signal is in red, empty vector (EV) in black. As EV has never been exposed to AsiSI, peaks in this sample cannot be due to AsiSI-induced DSBs.

Positive values for read coverage are those reads that map to the plus strand of the genome, conversely, negative values are those that map to the minus strand. N=2.

4.8 Optimising 4sU incorporation and pull-down specificity

DNA damage-induced small RNAs were unable to be detected in the previous sections, however the clear role for Drosha in DNA repair (Chapter 3) would suggest some involvement of an RNA species. A final approach was undertaken to identify any newly transcribed RNA species derived from the sequences around DSBs. To maximise the possibility of detecting these hypothetical RNAs, a method to enrich for only the RNAs produced following DSB induction was used. Several methods have been established in the literature, however here 4-thiouridine (4sU) incorporation was chosen as it is rapidly taken up by cells and utilised in transcription, with studies using 4sU pulses for as little as 10 minutes (Dolken et al. 2008; Rädle et al. 2013). The second reason for choosing 4sU over other alternative nucleosides is the specificity of the enrichment process: BrU incorporation for example requires the use of an antibody, which means non-specific products will also be pulled down as washes cannot be too stringent. 4sU-containing RNA on the other hand is easily biotinylated and enriched via incubation with streptavidin-coated beads. The biotin-streptavidin interaction is one of the strongest biological interactions (Chivers et al. 2011) and so stringent washes can be used to dramatically reduce non-specific binding. The specificity of 4sU incorporation can be observed in Figure 4.18 where cells treated with or without 4sU for 1 hour were subject to the same biotinylation reaction and streptavidin enrichment showing that only 4sU-containing RNA can be enriched in this manner.

4.9 4sU-seq

Previous experiments found that the U2OS-AsiSI cell line required 1 hour of 4-OHT treatment to be maximally induced, with damage being observed within 30 minutes (Figure 3.7, 3.8). Furthermore, Drosha depletion affected the recruitment of factors such as 53BP1 (Figure 3.2), which is recruited within minutes of damage induction (Bekker-Jensen et al. 2005; Hansen et al. 2016). As 4sU is incorporated within a few minutes, and no assumptions about how early a DNA damage-induced RNA is transcribed can be made, both 4sU and 4-OHT were added at the same time. This was performed in control or Drosha siRNA background, to determine how Drosha may affect this. Cells were harvested 1 hour later by TRIzol, alongside protein to validate damage induction and siRNA depletion (Figure 4.19).

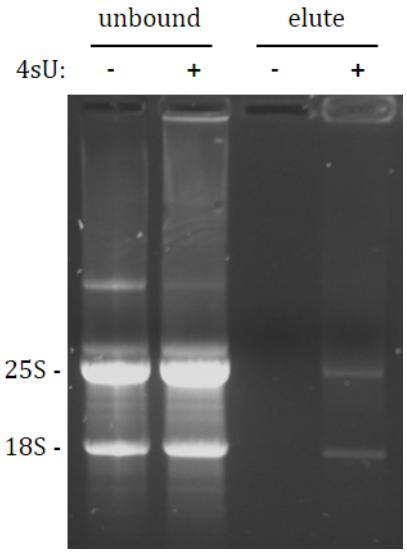


Figure 4.18. Biotinylation and streptavidin pulldown is specific for 4sU-incorporated RNA.

U2OS-AsiSI cells were treated with or without 4sU for 1 hour. RNA was harvested and subject to biotinylation reaction. This was then incubated with streptavidin-coated magnetic beads for 15 minutes. Beads were separated on a magnet and supernatant kept as unbound fraction, whereas bound fraction was washed and eluted by reversing the biotinylation reaction using DTT. These fractions were run on an agarose gel to compare 25S and 18S rRNA bands, showing that only 4sU-containing RNA is selected for with this methodology.

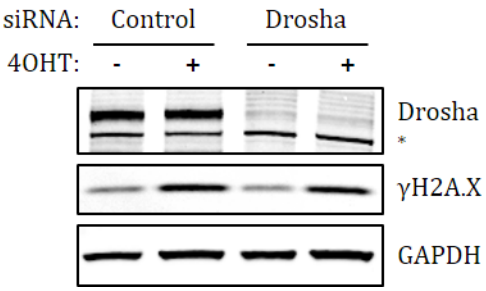


Figure 4.19. Validation of damage induction by AsiSI and Drosha depletion in samples used for sequencing.

U2OS-AsiSI cells were reverse transfected with control or Drosha siRNA for 48 hours. AsiSI was induced by 4-OHT addition for 1 hour; 4sU was added at the same time. RNA was harvested for TRIzol; in parallel, protein was collected. Representative western blot of three biological replicates. Damage induction and siRNA depletion was validated by immunoblotting for γH2AX and Drosha, respectively.

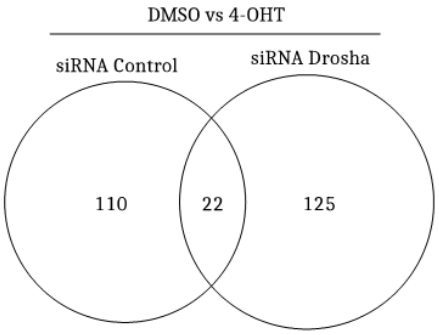
4sU-containing total RNA was biotinylated and captured by streptavidin-coated bead incubation (section 2.2.5). After several washes, RNA was eluted from beads by reducing the 4sU-biotin bond with 100 mM DTT and purified by acid phenol-chloroform extraction and ethanol precipitation. A yield of ~4% was obtained across a total of three biological replicates. These samples were then sent for stranded total RNA-seq library preparation with rRNA depletion and deep sequenced to obtain $8x10^8$ paired-end reads (section 2.2.5).

Paired FASTQ files were aligned to the human genome (GRCh38) using HISAT2, which also allows mapping of reads that span exon-exon junctions (D. Kim, Langmead, and Salzberg 2015). In addition, reads were also aligned to the human transcript reference GRCh38.85 using the pseudo-alignment programme kallisto (Bray et al. 2016). HISAT2 genome alignments were used to explore for novel RNA species around DSBs, whilst kallisto transcriptomic alignments were used to investigate how DNA damage and Drosha influences active transcription.

4.9.1 DNA damage causes expected changes to the transcriptome as detected by 4sU-seq

Changes to the transcriptome as a result of DNA damage is a well-established phenomenon, with a classic example being p53-controlled up-regulation of genes involved in cell cycle arrest. However, like most transcriptome studies, these have typically looked at steady-state mRNA levels hours after damage induction. By looking at changes to the annotated transcriptome in the 4sU-enriched RNA dataset, an idea of how DNA damage influences the transcriptional landscape in the very short term can be obtained. Kallisto was used to align 4sU-seq reads to Ensembl reference release GRCh38.85. Pseudo-aligned reads were then tested for differential expression between undamaged and damaged conditions by DESeq2 (section 2.5.6). This revealed 135 mature differentially expressed transcripts with an FDR of less than 0.1 following DNA damage in the control siRNA situation, 147 in the Drosha siRNA condition, with 22 overlaps between control and Drosha siRNA (Figure 4.20A). Most transcripts typically increased in levels following DNA damage, and this is particularly apparent in the 22 overlaps (Figure 4.20B). Interestingly, 5 transcripts within this overlap changed direction of differential expression in a Drosha-depleted background, suggesting there may be a role for Drosha in transcriptional regulation of some DNA damage

(A)



(B)

Gene	Direction	
	Control si	Drosha si
VCL	up	up
ARHGEF2	up	up
ECE1	up	up
SMG7	up	down
MCRS1	up	up
CSMD3	up	down
DMAP1	up	up
CLN6	up	up
COL22A1	up	up
GORASP1	up	up
AHCYL2	up	down
BRCA1	up	down
ARMC6	up	up
RBSN	up	up
RAB3IP	up	down
CUL5	up	up
NECAB3	up	up
GPR161	up	up
POLM	down	down
POLD1	down	down
ATP9A	down	down
CHEK2	down	down

(C)

Control siRNA
Log2(Fold Change) > 0

Gene Set Name	FDR q-value
Cell Cycle Process	1.14E-04
Mitotic Cell Cycle	1.14E-04
Cell Cycle	6.21E-04
Regulation Of Transcription From RNA Polymerase II Promoter	3.24E-03
Microtubule Cytoskeleton Organization	5.89E-03

Drosha siRNA
Log2(Fold Change) > 0

Gene Set Name	FDR q-value
Negative Regulation Of Cellular Component Organization	1.49E-03
Cellular Response To Stress	1.49E-03
Cell Junction Organization	1.49E-03
Regulation Of Cell Cycle	1.49E-03
Regulation Of Organelle Organization	1.49E-03

Log2(Fold Change) < 0

Gene Set Name	FDR q-value
DNA Metabolic Process	9.17E-04
Peptidyl Amino Acid Modification	1.01E-03
Regulation Of Intracellular Signal Transduction	1.30E-03
DNA Repair	3.40E-03
Protein Localization	1.41E-02

Log2(Fold Change) < 0

Gene Set Name	FDR q-value
Positive Regulation Of Biosynthetic Process	8.41E-05
Chromosome Organization	6.52E-03
Cellular Modified Amino Acid Metabolic Process	9.07E-03
Positive Regulation Of Sequence Specific DNA Binding Transcription Factor Activity	9.07E-03
One Carbon Metabolic Process	1.08E-02

Figure 4.20. DESeq2 analysis of transcriptome-aligned reads shows expected changes to transcription following DNA damage.

(A) Venn diagram showing overlap between control and Drosha depleted differentially expressed (DE) transcripts on DNA damage. **(B)** List of the 22 DE genes on DNA damage. Indicated is the direction (up- or down-regulated) of DE. Highlighted in yellow are the 5 genes which have contrasting direction of DE in Drosha depletion, indicative of transcriptional control by Drosha. **(C)** Gene ontology analysis of the up- or down-regulated genes on DNA damage in each siRNA condition. The top 5 in each group are shown.

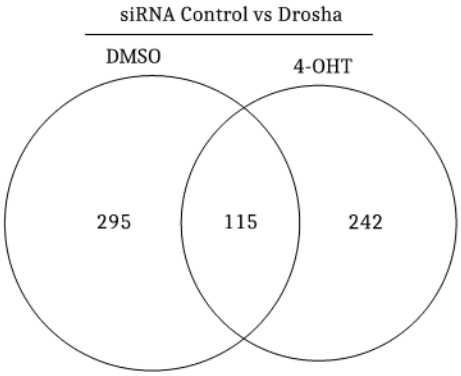
controlled genes. However, none of these differentially expressed genes were those cut by AsiSI.

Of the transcripts up-regulated following DNA damage in the control, gene ontology analysis reveals a strong enrichment for genes involved in the cell cycle (Figure 4.20C). This is not unexpected; however, this experiment shows that the change to this transcriptional network occurs immediately upon damage induction. Within the top ten enriched GO terms, two are related to cytoskeletal organisation, suggesting an immediate need to change the migration potential of the cell, or to cause nuclear rearrangements. In contrast, there are fewer related GO terms enriched in the down-regulated transcripts. Interestingly however, the top 5 terms of these include DNA repair and DNA metabolic process. This might suggest that DNA damage causes a decrease in the transcription of DNA repair factors. This is highly unintuitive, and further work to understand this will be required. In the Drosha depleted situation, DNA damage did not cause the same changes as in control (Figure 4.20AC). Some minor overlap occurs, such as cell cycle in the up-regulated set, however in the down-regulated this was not the case. This points towards Drosha being important for transcriptional regulation.

4.9.2 Drosha may be a regulator of transcription.

The comparison between control and siRNA Drosha in undamaged and damaged settings shows many more differentially expressed transcripts (449 and 381 respectively), 115 of which overlap (Figure 4.21A). Transcripts with increased transcription on Drosha depletion show enrichment for similar GO terms in the undamaged and damaged conditions, in many different cellular pathways (Figure 4.21B). The same is apparent with those that are decreased on loss of Drosha, with a major enrichment for cell cycle related transcripts. This may suggest that Drosha plays an important role in the transcription of cell cycle factors, in both normal and in stressed conditions. More gene set analysis was performed to investigate this, looking for transcription factor target and motif gene sets (Figure 4.22). In the damaged background, five out of the top ten enriched gene sets were those containing motifs for the transcription factor family E2F near their TSSs. The E2F family control transcription of many cell cycle genes. As loss of Drosha leads to a decrease in cell cycle factor transcription (Figure 4.21B), this suggests that Drosha may interact in

(A)



(B)

DMSO
Log2(Fold Change) > 0

Gene Set Name	FDR q-value
Negative Regulation Of Gene Expression	8.04E-14
Negative Regulation Of Nitrogen Compound Metabolic Process	7.55E-11
Regulation Of Intracellular Signal Transduction	6.78E-08
Regulation Of Transcription From RNA Polymerase II Promoter	8.17E-08
Chromosome Organization	1.36E-07

4-OHT
Log2(Fold Change) > 0

Gene Set Name	FDR q-value
Negative Regulation Of Gene Expression	4.51E-10
Negative Regulation Of Nitrogen Compound Metabolic Process	6.03E-07
Cellular Response To Organic Substance	6.50E-07
RNA Splicing	1.62E-06
Establishment Of Localization In Cell	1.62E-06

Log2(Fold Change) < 0

Gene Set Name	FDR q-value
Regulation Of Cell Cycle	2.50E-11
Cell Cycle	3.28E-11
Regulation Of Cell Cycle Process	3.82E-11
Mitotic Cell Cycle	7.71E-11
Cell Cycle Process	8.37E-11

Log2(Fold Change) < 0

Gene Set Name	FDR q-value
Cell Cycle	1.51E-11
Cell Cycle Process	7.08E-10
Positive Regulation Of Gene Expression	1.57E-09
Mitotic Cell Cycle	3.28E-08
Positive Regulation Of Molecular Function	5.19E-08

Figure 4.21. Differential expression of 4sU-enriched mRNA between control and Drosha knockdown conditions.

(A) Venn diagram showing the number of overlaps of DE genes on Drosha depletion between the damaged (4-OHT) and undamaged (DMSO) conditions. **(B)** GO analysis of the differentially expressed genes in the distinct groups in (A). Up- and down-regulated genes were analysed separately using the GSEA MSigDB. Top 5 enriched gene sets by significance are shown.

DMSO

Gene Set Name	FDR q-value
YRTCANNRCGC_UNKNOWN	4.35 e-19
TTGCWCAAY_CEBPB_02	1.07 e-15
TGGAAA_NFAT_Q4_01	2.25 e-14
TTTNNANAGCYR_UNKNOWN	9.48 e-14
WWTAAGGC_UNKNOWN	1.53 e-12
YAATNRNNNYNATT_UNKNOWN	2.05 e-11
AAAYWAACM_HFH4_01	5.18 e-11
TCCCRNNRTGC_UNKNOWN	2.17 e-9
TGCTGAY_UNKNOWN	3.41 e-9
WGTTNNNNNAAA_UNKNOWN	3.55 e-9

4-OHT

Gene Set Name	FDR q-value
YRTCANNRCGC_UNKNOWN	1.98 e-25
YAATNRNNNYNATT_UNKNOWN	1.59 e-10
E2F1_Q6	4.47 e-10
TTTNNANAGCYR_UNKNOWN	8.06 e-10
TGGAAA_NFAT_Q4_01	8.8 e-10
YATGNWAAT_OCT_C	1.4 e-9
E2F1DP1RB_01	1.4 e-9
E2F_Q6	1.4 e-9
E2F_Q4	1.4 e-9
E2F1DP1_01	1.4 e-9

Figure 4.22. Transcription factor and motif gene set analysis for DE transcripts following Drosha depletion shows enrichment on certain factors.

Gene set analysis was performed using the GSEA MSigDB to investigate the lists of DE transcripts on Drosha siRNA treatment, in both undamaged (DMSO) and damaged (4-OHT) conditions. Top 10 hits by significance are shown.

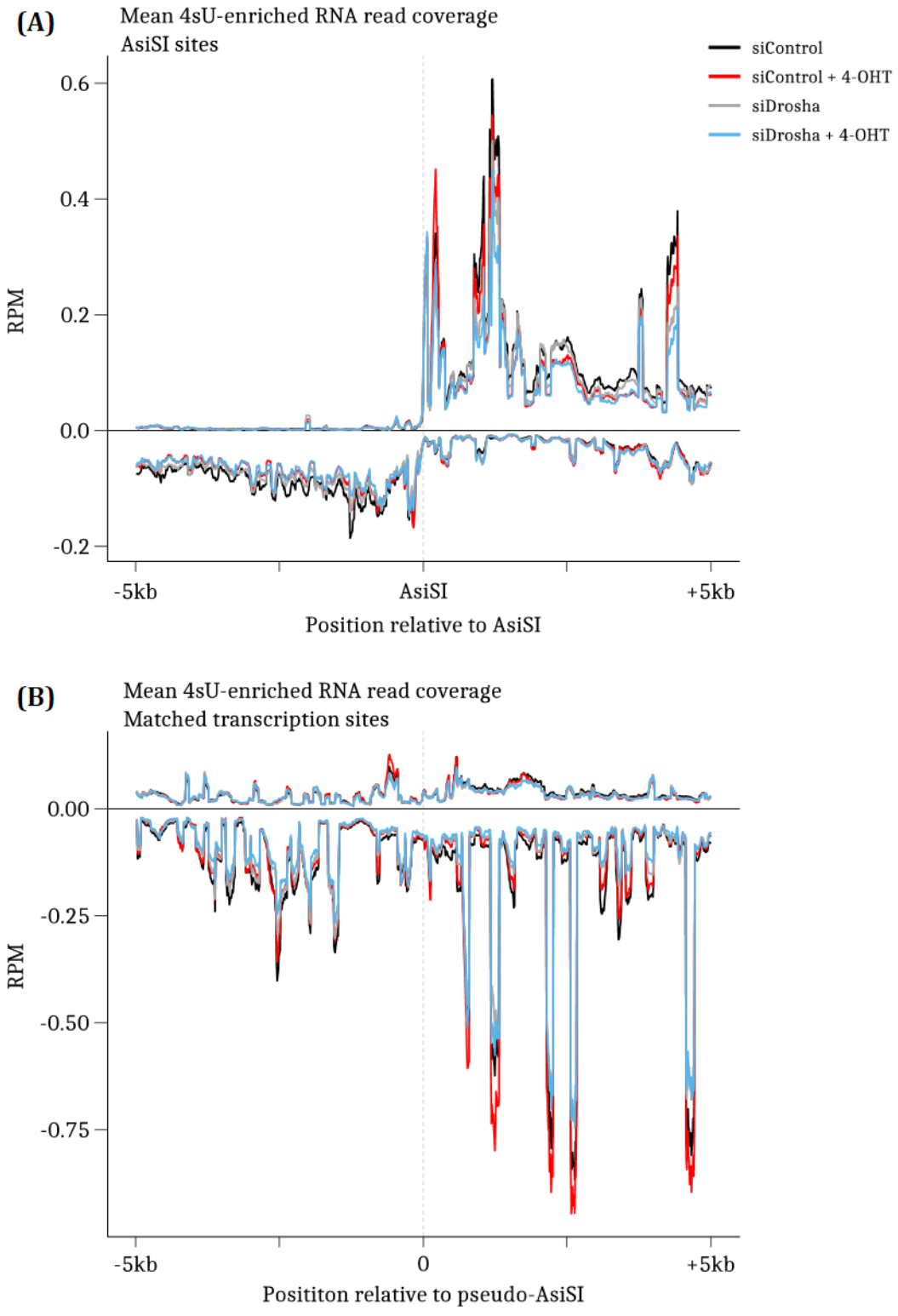
some way with E2F. However, a cursory look at the NCBI interactions section for the human Drosha gene did not show a previously documented interaction between E2F and Drosha, nor for other enriched transcription factors CEBPB, NFAT, OCTC, or HFH4 (Figure 4.22). Drosha has been demonstrated to interact with the transcription factor SP1 (Gunther, Laithier, and Brison 2000), which is known to interact with both E2F family members and CEBPB (Foti et al. 2003). Whilst SP1 transcription factor was not in the enriched lists, the top enriched motif, YRTCANNRCGC, was previously annotated as SP1 binding site. Taken together, this may implicate Drosha in transcriptional regulation.

4.10 Analysis of new transcripts around DSBs does not provide evidence that DNA damage induces local transcription.

Expected transcriptional changes following DNA damage were detected with the 4sU-seq dataset, showing that transcription can be rapidly turned on and off upon damage induction (Figure 4.20C). Therefore, if DNA damage induces local transcription at the site of the DSB immediately following break induction it should be revealed in this dataset.

Coverage profiles were generated for visualisation of transcription around AsiSI sites (Figure 4.23). While there does appear to be peaks around the AsiSI cut site, these are also present in the undamaged controls (Figure 4.23A). In the Drosha-depleted conditions there was slightly lower read density throughout, apart from directly adjacent to the cut site. The matched controls followed the same trend (Figure 4.23B); comparing just the control siRNA conditions side by side was not very informative either (Figure 4.23C). The presence of sharp peaks in these suggest that individual genes may skew the aggregate plots due to high expression. Additionally, exons are not likely to stack across multiple genes and this may produce a lot of the noise seen. Nevertheless, a strong induction of transcription at even a few sites would potentially stand out, which was not the case.

To look on an individual site basis, a statistical approach was undertaken as with the small RNA-seq experiments. DESeq2 was again used to determine whether any small changes could be detected at AsiSI cut sites. No significant differential expression was observed at any loci after damage with or without Drosha (Figure 4.24).


(A)
Mean 4sU-enriched RNA read coverage
AsiSI sites
siControl
siControl + 4-OHT
siDrosha
siDrosha + 4-OHT
0.6
0.4
0.2
0.0
-0.2
RPM
-5kb
AsiSI
+5kb
Position relative to AsiSI
(B)
Mean 4sU-enriched RNA read coverage
Matched transcription sites
0.00
-0.25
-0.50
-0.75
RPM
-5kb
0
+5kb
Posititon relative to pseudo-AsiSI

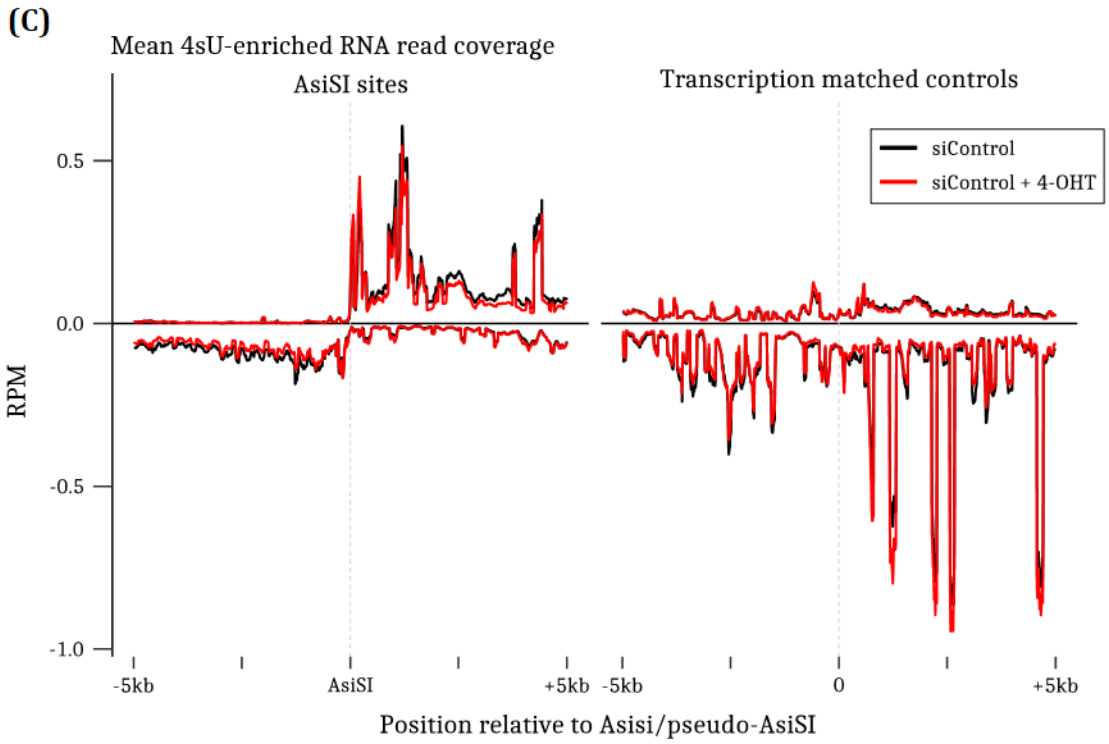


Figure 4.23. 4sU-incorporated read density plots do not show any striking changes to the transcriptional landscape around AsiSI sites after damage.

(A) Mean read coverage was calculated in 10 nt bins 5 kb either side of each AsiSI cut site. Only the mate 2 read of the sequenced pair was used, as this has the same sequence as the original RNA fragment. Reads mapping to the plus strand have positive coverage, those mapping to the minus strand have negative coverage. N=3. **(B)** As in (A), however mapping around control uncut matched transcriptional activity loci. Pseudo-AsiSI refers to the midpoint of each locus; this was calculated to be a similar distance from the TSS as the paired AsiSI site. **(C)** As in (A) and (B), however only the control siRNA conditions are plotted to better compare the AsiSI loci and the control loci.

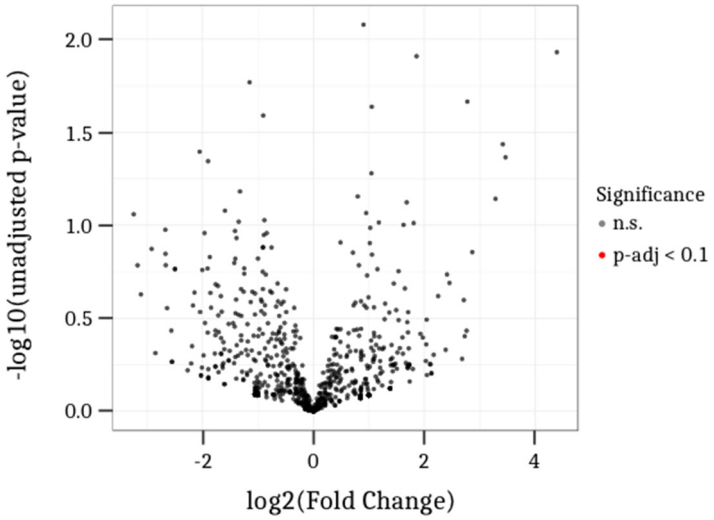


Figure 4.24. No significant change of 4sU-incorporated RNA reads occurs in the regions around any AsiSI motif.

The mate 2 read of each pair was counted within AsiSI loci and tested for differential expression using DESeq2. The GLM was built using all replicates and all conditions; every contrast tested had FDR>0.1.

Taken together, these data do not provide any strong evidence for new transcription from the sequences around DSBs immediately following induction of DNA damage.

4.11 Chapter discussion

Drosha and Dicer were shown in the previous chapter to participate in DNA damage repair. As these are RNA binding and metabolising enzymes, it would be expected that their role in DNA repair would involve an RNA. Various proteomic studies have demonstrated a significant role for RNA processing enzymes in DNA repair, indicating that RNA may indeed take part in DSB resolution (section 1.5) (Matsuoka et al. 2007; Paulsen et al. 2009; Hurov, Cotta-Ramusino, and Elledge 2010; Slabicki et al. 2010; Cotta-Ramusino et al. 2011; Adamson et al. 2012). As Drosha and Dicer are the miRNA biogenesis enzymes, the simplest hypothesis is that they could process an RNA into a small RNA which could then have some role in repair. This is shown in fungi, plants and humans, albeit mostly through reporter based systems (Lee et al. 2009; W. Wei et al. 2012; Francia et al. 2012). In human cells, this has not been demonstrated at all outside of a highly-expressed reporter system or highly repetitive systems (W. Wei et al. 2012; Rossiello et al. 2017). It was important to validate these findings at non-repetitive, endogenous loci, to investigate whether small RNA generation is part of the DDR. In this thesis, the AsiSI endonuclease system was chosen as it cuts in many different genomic locations within the genome, and has been extensively characterised (Aymard et al. 2014). For example, the dominant repair pathway undertaken by each site has been documented, meaning models can be built based on differential responses at each site. Furthermore, DSB induced by the endonuclease can be quickly introduced by the addition of 4-OHT; in Figure 3.7, damage induction could be seen in as little as 30 minutes. This allows the use of more biologically relevant timepoints (with respect to the DDR), a major advantage over other reported systems which utilised timepoints over 12 hours (W. Wei et al. 2012; Francia et al. 2012; Michalik, Böttcher, and Förstemann 2012).

In this system, a time course of 0 (uninduced), 1, 6, 12, and 24 hours following 4-OHT addition was utilised. A background control where no AsiSI is present at all was also used, as any small RNA products produced due to AsiSI cleavage should not be detected. Small RNA was deep sequenced. The use of two biological replicates and multiple timepoints allowed a robust general linearised model of the data to be fit;

however, this did not reveal any damage-induced small RNA from the regions around any of the 99 break sites, at any timepoint (Figure 4.9). All annotated small RNAs were removed prior to analyses, so that they would not affect the analysis. Across the 99 cut sites, the average number of reads was 5.3 reads per million in a 10 kb region. As such, even minute increases in the small RNA mapping to these regions would be detected. Coupled with the observation that this experiment was at least as sensitive for detection of low abundance RNA as a previously reported dataset used for small RNA discovery (Figure 4.7), and that expected changes to the miRNA pathway were validated over time (Figure 4.5), this strongly suggests that this dataset is highly reliable and from this can conclude that damage-induced small RNA is not a mechanism of the DDR at endogenous and non-repetitive loci.

A second approach was also undertaken to explore the possibility of a small RNA amplification pathway, as was shown to be required for damage-induced small RNA in plants (W. Wei et al. 2012). This process requires an RNA-dependent RNA polymerase. Secondary amplified products possess 5'-triphosphate moieties instead of the miRNA-like 5'-monophosphates, and would not be captured in standard library preparations. A method to sequence these was developed (Figure 4.11), as it could be that previously reported small RNAs were captured for sequencing due to chance dephosphorylation during cell lysis or another mechanism. This analysis did not reveal any evidence for small RNA amplification in humans at DNA break sites.

As previously reported small RNA in human cells were derived from highly expressed reporters (W. Wei et al. 2012) or repetitive reporters/loci (Francia et al. 2012; Rossiello et al. 2017), it could be that they are degradation products from those highly abundant RNAs, or arise due to repeat copy number variation due to HR repair (section 6.1). Nevertheless, the role for RNA processing enzymes in DNA repair (Chapter 3) still implicates RNA in the process. The published models for damage-induced small RNA require *de novo* RNA synthesis at the DSB (section 1.5.2). An alternative hypothesis that it would be this longer, primary transcription event that could participate in the DDR was developed, and that detected small RNA was a degradation product of this. As the majority of RNA in a cell is steady state rRNA, tRNA, and mRNA (Palazzo and Lee 2015), this would potentially drown out any newly synthesised RNA, especially if they had a short half-life. To test this hypothesis with this issue in mind, 4sU was utilised to

enrich for all newly synthesised RNA following its addition. This was added simultaneously to the addition of 4-OHT to activate the endonuclease and incubated for 1 hour, as in this model a new RNA would be transcribed before repair is completed. Analysis of annotated transcripts demonstrated that expected changes to the transcriptome due to DNA damage could be observed (section 4.9.1 and Figure 4.20), showing that this methodology enriched for newly synthesised RNA upon DNA damage. Interestingly, this also revealed that Drosha was important for the transcription of numerous genes, suggesting it may act as a transcriptional activator (section 4.9.2 and Figure 4.21). This activity of Drosha has been previously described (Gromak et al. 2013). Whether this was the mechanism by which Drosha participates in the DDR was tested by analysing the 4sU-enriched RNA read coverage around AsiSI sites as performed for the small RNA-seq experiments. This analysis did not reveal any damage-induced *de novo* transcription around AsiSI loci, affected by Drosha depletion or not (Figure 4.24). This observation coincides with the consensus among the literature that DNA damage inhibits transcription locally (section 1.5.3) (Kruhlak et al. 2007; Shanbhag et al. 2010; Pankotai et al. 2012; Manfrini et al. 2015; Solovjeva et al. 2007; Iacovoni et al. 2010; Gong et al. 2015; Iannelli et al. 2017).

Potential caveats with this methodology are that AsiSI takes a while to be activated, so 4sU needed to be added for 1 hour. This length of time captures many of the very highly transcribed RNA in the cell and many processed, mature RNAs meaning the enrichment is not as great as it could be. Further optimisation with the timing of 4-OHT so that shorter 4sU pulses can be used may also improve this method. Furthermore, long and slowly transcribing RNAs will incorporate 4sU into their 3' ends because they were still transcribing when 4sU was added, meaning some sequenced RNA is not reflective of what was newly transcribed during the 4sU pulse. Alternative methodologies could address this, such as fragmenting 4sU-incorporated RNA before enrichment to sequence only newly transcribed fragments instead of whole RNAs, as in the recently described TT-seq (Schwalb et al. 2016). Native-elongating transcript sequencing (NET-seq) is another approach, which is an RNA polymerase II RNA-immunoprecipitation which captures all RNAs actively being transcribed (Nojima et al. 2015). Another interesting possibility is that a *de novo* transcript would be modified in some way. Recently, a role for N6-methylated (m6A) RNA at UV-induced

damage sites was shown to be important for focal recruitment of downstream NER factors (see section 1.2.8.3) (Xiang et al. 2017). The presence of m6A could roadblock the reverse transcription (RT) of library preparation, meaning the likelihood of sequencing such an RNA would be reduced. Nevertheless, it is hard to image that this would prevent sequencing completely: tRNAs are heavily methylated and were still able to be sequenced in this dataset (data not shown) without special treatment to remove these.

From the experiments carried out here it does not appear that DNA damage induces new transcription or processing of RNA to small RNA from the sequences around that DSB. What Drosha or Dicer could be doing in DNA repair is still likely to be RNA related as all non-canonical functions of these enzymes still utilise their RNA binding and metabolising activities (Gromak et al. 2013; Dhir et al. 2015; Han et al. 2009; E. White et al. 2014). One of these functions of Drosha is a role in transcriptional activation (Gromak et al. 2013), which is supported from data in this chapter (Figure 4.21); however, new transcription from DSBs was not detected.

5 Chapter 5: DNA:RNA hybrids form at DNA break sites facilitated by Drosha and are important for DNA repair.

5.1 Chapter Introduction

The miRNA biogenesis apparatus was shown in Chapter 3 to participate in the DNA double-strand break repair pathway. The focus was primarily on the role of Drosha, due to the stronger effect observed upon its depletion compared to other components such as Dicer. This was shown to be a distinct activity to that of the biogenesis of miRNA; numerous publications have demonstrated an involvement of Drosha in several other RNA processing events such as transcriptional termination (Dhir et al. 2015). Previous reports of small RNA generated from sequences around DSBs led to the question of whether Drosha may be involved in the DDR by processing a nascent transcript into small RNAs that would function in DNA repair. Chapter 4 set out to investigate this, applying an array of deep sequencing approaches aiming to identify whether damage-induced small RNAs are a general feature of DSBs within the genome, and whether the nascent transcripts they would be generated from could be detected. This appeared not to be the case, suggesting previous reports may have detected artefacts due to the nature of the reporter systems or repetitive sequences they had been mapped to.

With *de novo* synthesised RNA products ruled out, the issue of the role of Drosha in DNA repair remains unresolved. As this activity of Drosha is still likely to be RNA-related, the investigations were turned to pre-existing RNA species at sites of DNA damage. Most of the genome is thought to be actively transcribed (Wong, Passey, and Yu 2001; Clark et al. 2011; Djebali et al. 2012; Hangauer, Vaughn, and McManus 2013), meaning that any randomly introduced DSB is likely to occur within a locus that has an RNA complement either in close proximity or elsewhere in the cell. Potential mechanisms for a role for RNA in the repair of its gene was discussed in section 1.5. The most exciting possibility, with several pieces of literature supporting it, is that the RNA could be used as a template for error-free repair (Storici et al. 2007; Keskin et al. 2014; Mazina et al. 2017). A critical step for such a mechanism would involve the

hybridisation of the RNA to the DNA. This structure is known as an R-loop (for RNA displacement loop), which form during transcription (section 1.3.3). These are generally considered detrimental to genomic stability, however recent data has suggested an important role for these in the torpedo model of transcription (see section 1.3.4) (Skourti-Stathaki and Proudfoot 2014).

This final chapter aims to determine whether RNA can interact with the DNA at sites of DSBs, and whether Drosha is involved in this.

5.2 R-loops are formed rapidly following DNA damage induction.

DNA:RNA hybridisation can be probed for with two major tools: the S9.6 antibody which can recognise the DNA:RNA duplex, and RNase H1, a ribonuclease that binds to DNA:RNA hybrids and in its active form degrades the RNA component. To investigate whether RNA hybridises to DNA at sites of damage an R-loop-recognising but catalytically dead (D10R E48R) *E. coli* RNase H1 (referred to henceforth as $RNH1^{mut}$), fused to mCherry and a nuclear localisation signal, was used in laser microirradiation experiments. U2OS cells were seeded in chambered coverglass wells transfected with $RNH1^{mut}$ or an mCherry control (Table 8-3). 24 hours later, media was replaced with phenol red-free media and the chambered coverglass mounted within a stage-top environment control system for the spinning disk inverted microscope and maintained at standard cell culture conditions (37 °C, 5% CO_2). Cells were pre-sensitised for laser microirradiation by the addition of Hoescht 33342 (20 μM final) for 10 minutes. Pre-sensitisation increases the susceptibility of DNA to be damaged by a laser, meaning a lower laser power can be used to reduce the off-target effects of high-powered light. Using the red (561 nm) channel, transfected cells were located and laser stripes drawn across nuclei using the microscope software. A 532 nm pulsed laser was activated across the stripes and an image taken every 500 ms in the 561 nm channel.

Laser microirradiation caused a rapid relocation of the $RNH1^{mut}$ to the laser stripes (Figure 5.1A). The average intensity along the laser stripe was normalised to a second non-irradiated stripe within the same nucleus to account for fluctuations in fluorescence. A total of 168 cells were quantified over 3 biological replicates, showing

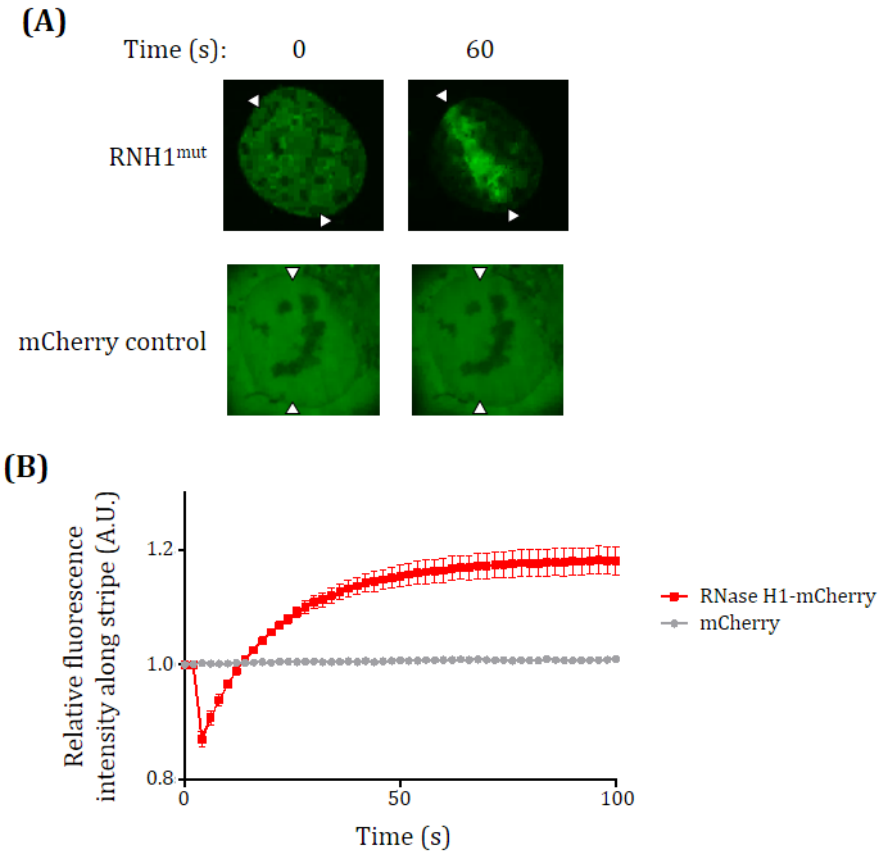


Figure 5.1. Relocalisation of RNH1^{mut} suggests DNA:RNA hybrids structures form at sites of DNA damage.

(A) Representative image of a cell used in microirradiation experiments. RNH1^{mut} or mCherry transfected U2OS cells were pre-sensitised for 532 nm laser damage by adding 20 µM Hoescht 33342 for 10 minutes. Laser stripes were drawn across nuclei using 3i Slidebook software, as shown by arrowheads. Microirradiation was induced along the stripe and images taken every 500 ms in the 561 nm channel. Images are shown in green for better visibility. **(B)** Quantification of cells as shown in (A). The fluorescence intensity along each stripe was normalised to a background undamaged stripe in the same nucleus. At least 40 cells measured per experiment, N=3, error=SEM.

a clear increase of signal over background along the laser stripe (Figure 5.1B). An mCherry transfected control did not show a similar effect, indicating the relocation of RNH1^{mut} is specific (Figure 5.1AB). This suggests that DNA:RNA hybrids form at sites of DNA damage.

5.3 DNA:RNA immunoprecipitation (DRIP) shows increased hybridisation around DSBs.

5.3.1 DRIP

To further investigate DNA damage-dependent R-loop formation, a more direct approach was undertaken. The S9.6 antibody specifically recognises the DNA:RNA helical structure, and can be utilised to perform DNA:RNA immunoprecipitation (DRIP) on fragmented DNA. Any R-loop-containing DNA fragments will be bound by the antibody and enriched. Here, like in most uses of DRIP in the literature, the RNA component of the DNA:RNA-containing fragments was then degraded by RNase 1. The R-loop-prone genomic regions can then be identified via qPCR or deep sequenced. The DRIP experiments and qPCR in this sub-section were performed in collaboration with Wei-Ting Lu (Appendix 8.1).

DNA damage was induced for 2 hours by the addition of 300 nM 4-OHT, or an equal volume of DMSO as control. Genomic DNA was isolated and fragmented by sonication (section 2.3.3). Each sample was split into two tubes and one of these was incubated in the presence RNase H1. As RNase H1 degrades the RNA component of R-loops, this acts as a negative control for the antibody. The samples were purified by phenol:chloroform:isoamylalcohol (pH 8.0) extraction and ethanol precipitation and subject to DRIP using the S9.6 antibody, with an aliquot set aside as input (this is described in detail in section 2.3.3).

Primers designed against genomic regions around AsiSI cut sites were used for qPCR of the DRIP fragments, as well as primers for the beta-actin locus as a control site (Table 8-4). DRIP C_T values were normalised to input C_T values as a measure of DNA:RNA hybridisation within those loci. This showed that DNA:RNA hybridisation increases following DNA damage around AsiSI cut sites, but not at other undamaged loci within the genome (Figure 5.2A). The RNase H1 treated samples had reduced signal for all loci tested, but was still observable (Figure 5.2A, shaded grey). This suggests that RNase H1

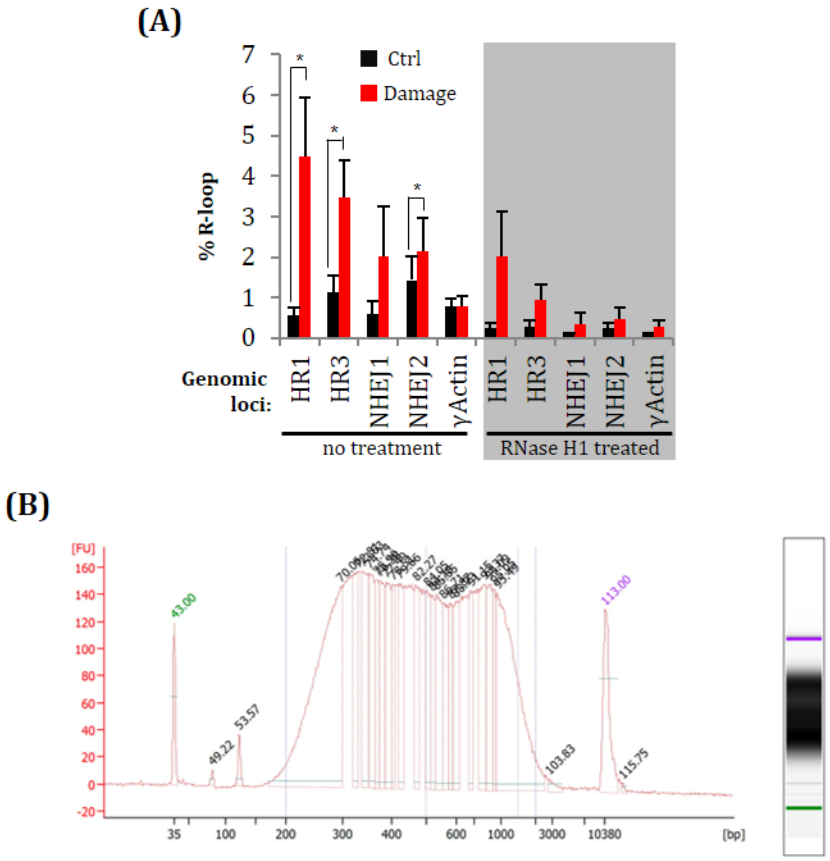


Figure 5.2. DNA:RNA hybrid immunoprecipitation (DRIP) shows an increase in R-loop formation at DSBs.

(A) DRIP followed by qPCR around AsiSI sites that are HR- or NHEJ-prone (primers used are in Table 8-4 and Table 8-5 is a list of all cut AsiSI sites by repair). As a control, primers against γ-actin exon 5 (which does not contain AsiSI motifs) were also tested. Damage by AsiSI was induced by the addition of 300 nM 4-OHT for 2 hours. Shaded in grey are the paired samples treated with RNase H1 prior to DRIP as a negative control. N=4, *p<0.05, Student's t-test. Error=SD. **(B)** Example Bioanalyzer high sensitivity DNA chip trace of a DRIP-seq library made from the DRIP-ed fragments in (A) using the NEBNext Ultra II library preparation kit. Concentrations of these libraries were excessive meaning PCR cycles can be reduced in the future.

treatment may be incomplete, and, as this method is amplification based, a small amount of DNA:RNA hybrids can still be detected by qPCR. The DRIP fragments were then used to generate NGS libraries (DRIP-seq), with the hope that a more global approach would provide mechanistic insight into DNA:RNA hybridisation following DNA damage.

5.3.2 DRIP-seq

DRIP-seq libraries were generated from DRIP-ed DNA fragments using the NEB Ultra II kit (see section 2.3.5). Briefly, 750 pg of fragments underwent an end preparation step followed by adaptor ligation. The end preparation step cleaves 5' overhangs and fills in 3' overhangs to create ends with a 5'-monophosphate and 3'-dA, which can then be ligated to double-stranded hairpin adaptors. Adaptor-ligated DNA was amplified using 11 PCR cycles, as recommended by the manufacturer's protocol. Following a final clean-up step, libraries were run on a Bioanalyzer high sensitivity DNA chip (Figure 5.2B). The final concentrations of libraries were much higher than required, suggesting a lower number of PCR cycles could be used in the future. Samples were then pooled at equimolar concentrations and sequenced on an Illumina NextSeq500 (University of Cambridge Biochemistry Department), generating a total of 400 million reads.
FASTQ files were aligned to the human genome (GRCh38) using bowtie2 (Langmead and Salzberg 2012). Samtools was then used to filter BAM files to remove low quality alignments (those with MAPQ<30; see section 2.5.8), as recommended for ChIP sequencing analysis. Coverage of mapped reads around AsiSI cut sites was calculated using bedtools (Quinlan and Hall 2010) and normalised to the total number of mapping reads genome-wide. The mean read coverage across the 99 cut sites (Table 8-5 (Aymard et al. 2014)) was plotted (Figure 5.3A). This revealed an almost 2-fold increase of read density directly at the AsiSI cut sites following 4-OHT addition. For comparison, the representative transcriptional activity control sites as used in Chapter 4 do not show this increase (Figure 5.3B). This shows that RNA does indeed base pair with DNA following DSB induction. Additionally, the trough in the coverage peak is positioned directly at the AsiSI cut site, indicating that these R-loops have indeed formed around broken DNA (Figure 5.3A). As DNA:RNA hybridisation occurs before repair has taken place, this raises the possibility that this may be part of the DDR pathway.

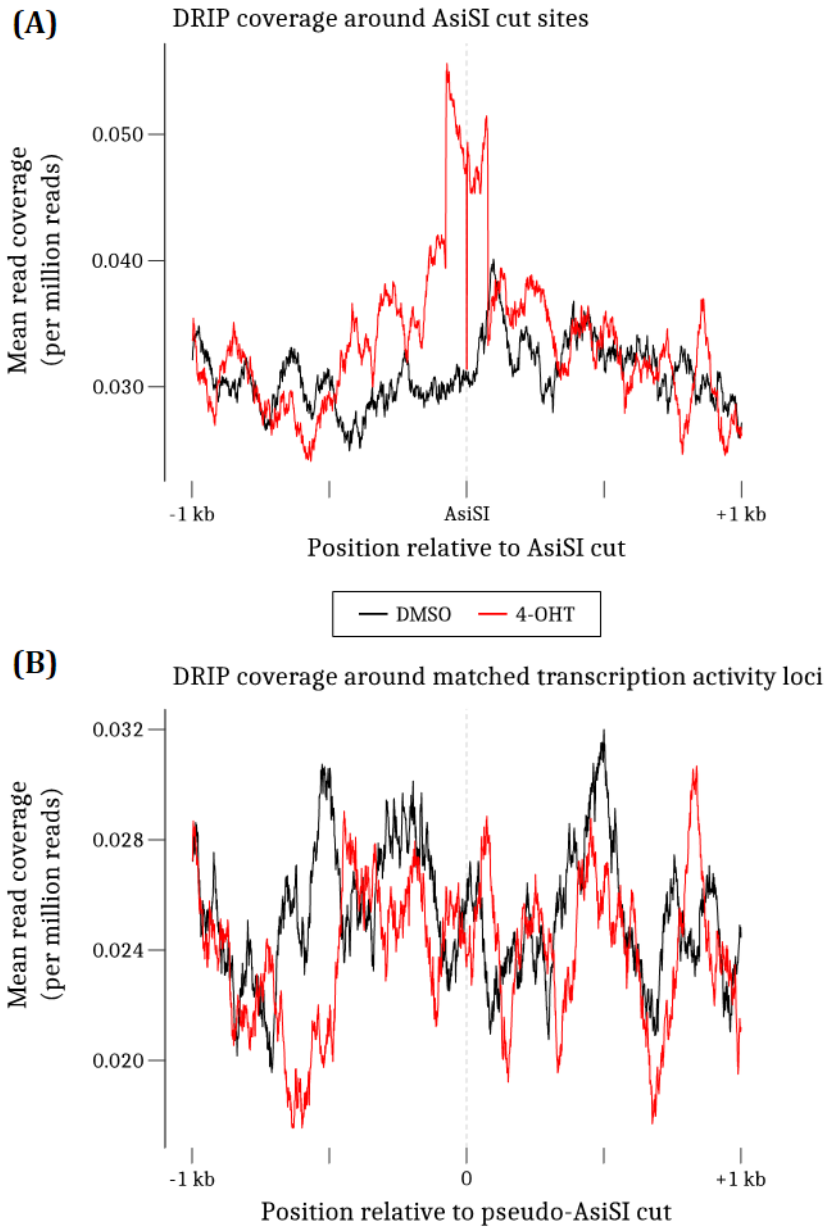


Figure 5.3. Aggregate plots of DRIP-seq reads around AsiSI cut sites show a clear increase in DNA:RNA hybridisation after DNA damage.

(A) Mean per-base coverage of DRIP-seq reads with MAPQ>30 calculated using bedtools and custom scripts and was plotted around all 99 AsiSI cut sites (Table 8-5). **(B)** As in (A), however per-base coverage was plotted around the matched transcriptional activity loci Table 8-6, as used in Chapter 4.

5.4 A follow-up DRIP-seq experiment confirms DNA:RNA hybridisation at sites of DNA damage.

The first DRIP-seq experiment had two caveats: first, single-end sequencing was performed meaning no more information about where the extent of the R-loop structure can be determined as this method identifies only one of the extreme ends of a fragment. A second potential issue is that a specific break was introduced in the 4-OHT condition by AsiSI and not in the DMSO condition. As sequenced fragments are read from one end, in the 4-OHT condition sequencing would always occur from that specific end, meaning reads could be skewed in favour of the cut site. To address these issues, another DRIP-seq was performed with two changes. Prior to the random fragmentation by sonication, all samples were treated with AsiSI endonuclease *in vitro* to cleave all AsiSI motifs. The aim of this was to introduce the same ends to the DNA in the undamaged condition as in the damaged condition. The second change to the protocol was to perform paired-end sequencing, so both ends of a DRIP-ed fragment would be sequenced. This not only greatly increases the mappability of reads, but as it is genomic DNA that is sequenced, the entire fragment can be deduced meaning a much better idea of where R-loops are forming can be determined.

As before (section 5.3.1), the DRIP experiment was performed in collaboration with Wei-Ting Lu (Appendix 8.1).

Sequencing libraries were prepared as before (section 2.3.5, 5.3.2), starting with 1 ng input DNA. The number of PCR cycles was reduced to 8 as the previous library was much more concentrated than required (Figure 5.2B); this reduces amplification bias as well as the number of PCR duplicates. A Bioanalyzer analysis confirmed successful generation of libraries (Figure 5.4) which were then pooled at equal concentrations, given a final size selection clean-up, and sequenced on an Illumina NextSeq500 (University of Cambridge Biochemistry Department) generating 800 million paired-end reads. Paired FASTQ files were aligned to the human genome (GRCh38) using bowtie2 and again filtered by mapping quality. Concordant mate reads (those mapping on the same chromosome, in close proximity and opposing one another) were then converted from BAM to paired BED format (BEDPE) using bedtools. PCR duplicates were removed by filtering for only unique read pairs (i.e. two reads mapping identically were not considered PCR duplicates if the mate reads were not also identical). The lowest and

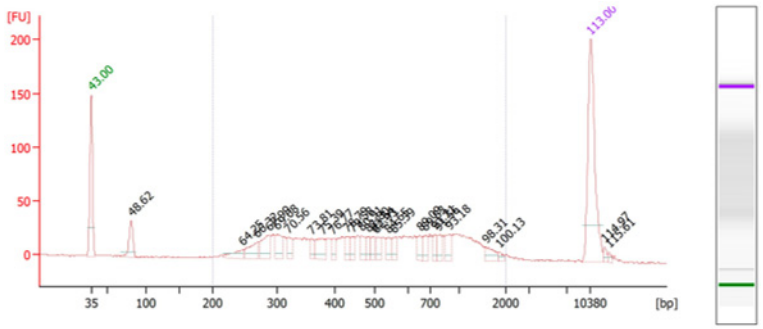


Figure 5.4. DRIP-seq libraries for paired-end sequencing were more appropriately amplified.

For the follow-up DRIP-seq experiment, 25% more input DRIP-ed fragments were used (750 ng to 1 µg), and the number of cycles was reduced to 8 from 11. This is beneficial to reduce amplification bias, and to reduce the number of PCR duplicates that arise from the amplification step.

highest mapped base of the read pair were then used to create a standard BED file containing the original sequenced fragment information. Coverage of DRIP fragments was calculated using bedtools and mean coverage as fragments per million fragments (FPM) around AsiSI sites was plotted as before. An increase in DRIP signal was observed following 4-OHT induction (Figure 5.5A). As both undamaged and damaged conditions were cleaved *in vitro* by AsiSI, this confirms that the increase in coverage is due to *in vivo* DNA:RNA hybridisation as a result of DNA damage. This also confirmed the observation that the DNA:RNA hybrids form in the most proximal regions to DSBs. In the Drosha depleted setting, no DNA damage-dependent increase can be observed compared to undamaged cells (Figure 5.5A, right). This suggests Drosha may be important for the formation of these hybrids, or for their retention.

Plotting DRIP coverage around the same random but transcriptionally matched control loci as in Figure 4.9 does not show any changes to R-loops following 4-OHT (Figure 5.5B). This confirms that the increase in signal around AsiSI loci is specific to the induction of DSBs at those loci. As an additional control, DRIP coverage was calculated around the remaining 1121 AsiSI recognition motifs in the genome (those that were determined to not be cut *in vivo* (Aymard et al. 2014)) (Figure 5.6A). This, like the transcriptionally matched control, does not show an increase in signal following DNA damage. This does show that DRIP peaks in all cases, which is to be expected as all AsiSI motifs are found primarily at gene TSSs (Figure 3.6), and the matched controls were designed to be in similar locations. For a fairer comparison to the 99 cut AsiSI sites, the 1121 sites were randomly split into four groups of 99 and similarly plotted (Figure 5.6B). These plots show a trough at the AsiSI site due to the *in vitro* AsiSI cleavage performed with this experiment; this highlights how certain recognition motifs may not be available to be cut *in vivo* due to compaction in heterochromatin whereas others will not be available in either case due to methylation. There is also the possibility that some motifs can be cleaved *in vivo*, however this may be at very low efficiency. Nevertheless, no significant changes to DRIP coverage following induction of damage can be observed at any of these subsets, further providing evidence that DNA:RNA hybridisation occurs around actively cut DSBs.

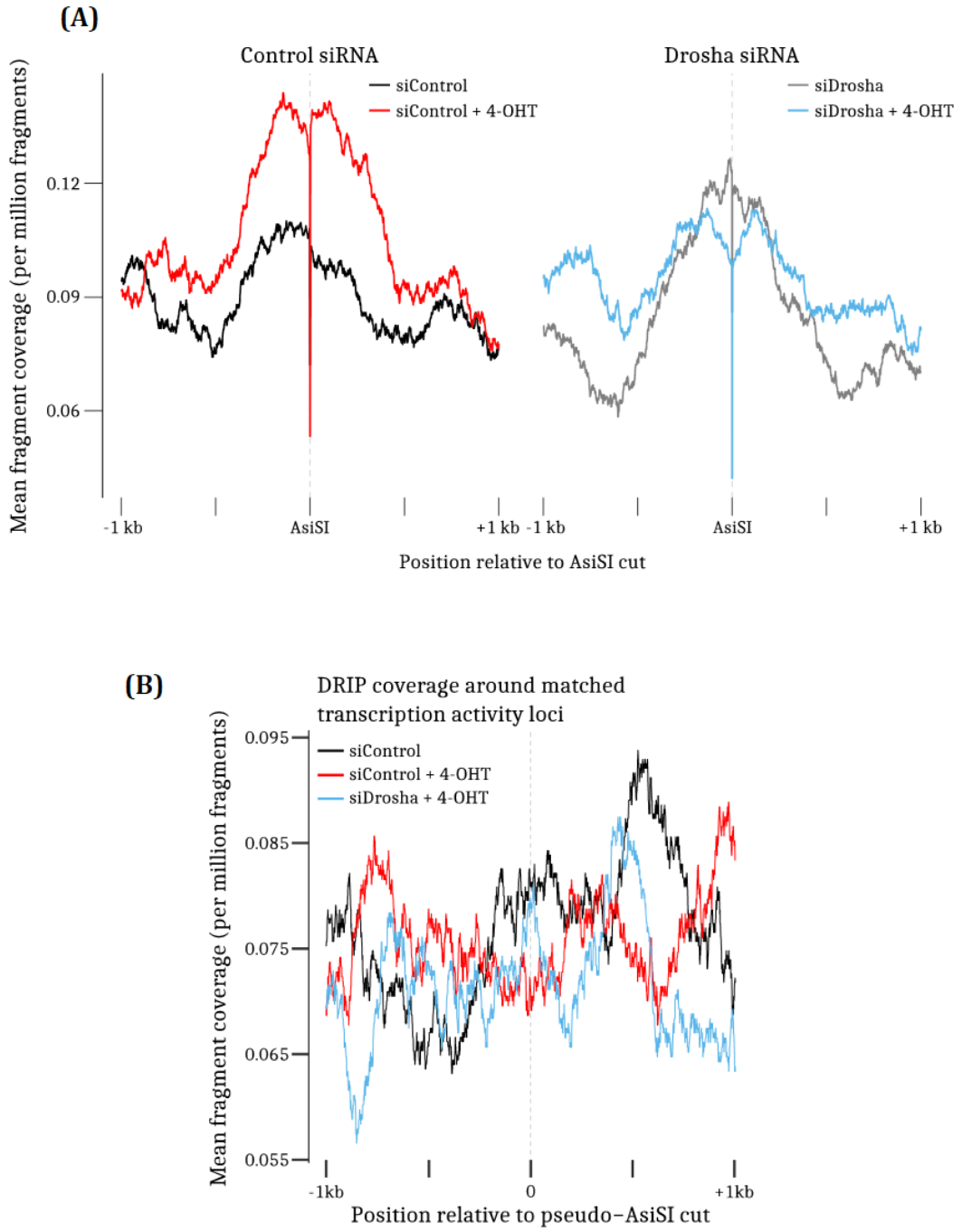


Figure 5.5. Paired-end sequencing confirms DNA:RNA hybrids form highly proximal to AsiSI-induced DSBs and show Drosha is required for their formation.

(A) Mean per-base coverage of unique DRIP-seq fragments was calculated using bedtools and custom scripts and was plotted around all 99 AsiSI cut sites (Table 8-5). This was facetted to compare the effect of DNA damage in control and Drosha siRNA setting. **(B)** As in (A), plotted around matched transcriptional activity control sites Table 8-6.

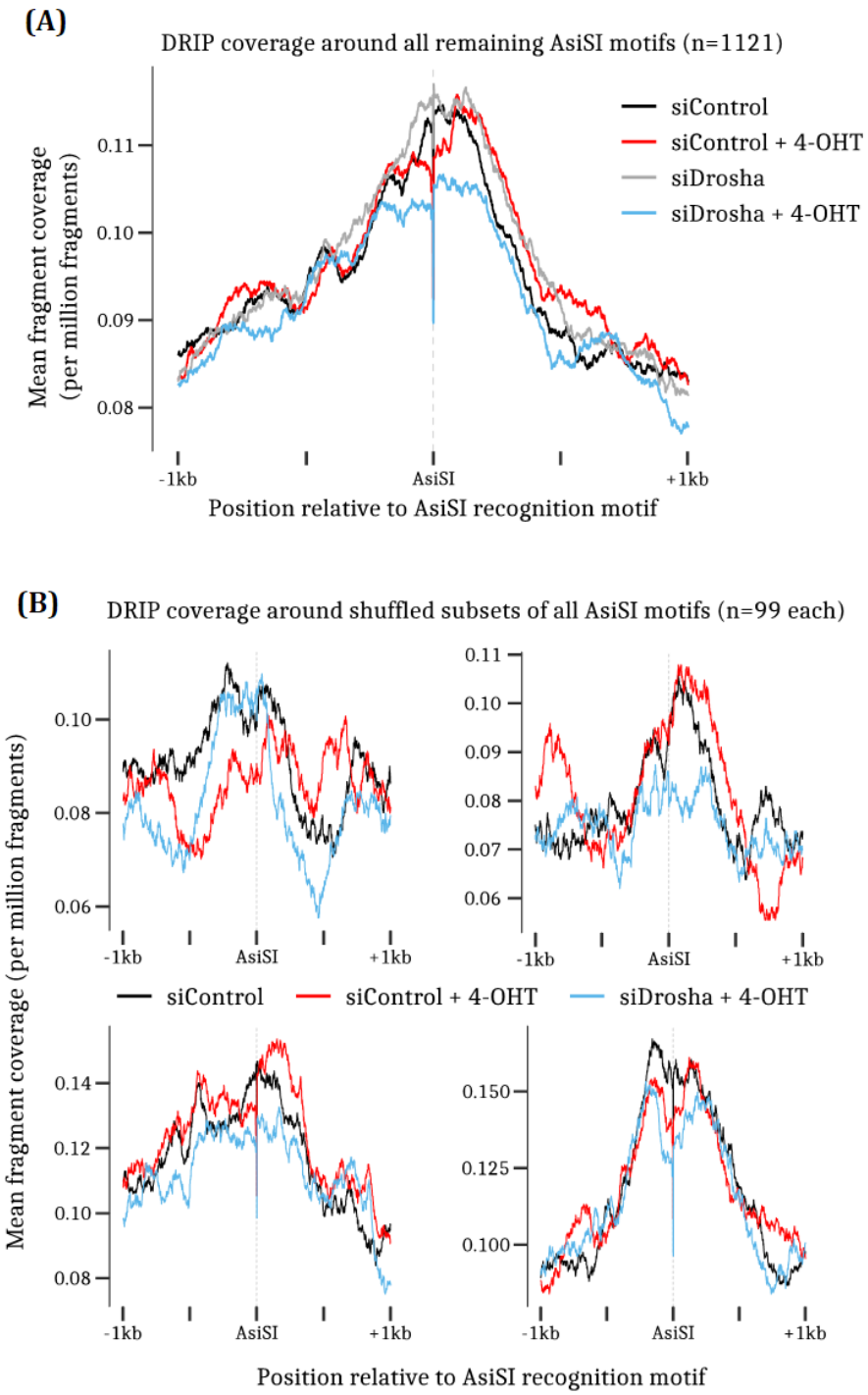


Figure 5.6. DRIP-seq coverage around the uncut AsiSI motifs are another matched control showing *in vivo* DSBs are required for R-loop formation.

(A) Mean per-base coverage of unique DRIP-seq fragments was calculated using bedtools and custom scripts and was plotted around the 1121 uncut AsiSI motifs. **(B)** Four random subsets of 99 uncut loci were plotted as in (A).

5.5 Damage-dependent R-loops form around both HR- and NHEJ-prone sites and is dependent on transcription.

The repair pathway selection at each cut AsiSI site was previously reported (Aymard et al. 2014): of the 99 cut, 25 were determined to be HR-prone, 25 NHEJ-prone, with the remaining 49 assigned ambiguous (no strong enrichment for repair factors for either pathway). To investigate whether the formation of R-loops is determined by repair pathway choice, DRIP coverage was calculated and plotted around each site as divided by repair group (Figure 5.7). DNA damage-dependent R-loop formation was observed around both HR- and NHEJ-prone sites, and Drosha depletion again reduced R-loop formation. In contrast, the ambiguously repaired sites had a much more modest increase in R-loop formation. As these sites are loci with more competition between factors from the two pathways, it could be that R-loops are cleared quickly or not permitted to form at all. Alternatively, these may be sites of fast repair kinetics, meaning damage-dependent R-loops are less likely to be isolated in the timeframe of the experiment.

The publication which defined repair pathway selection for the AsiSI cut sites also found that the transcriptional activity of a damaged locus affects its repair choice, where more transcriptionally active genes undergo error-free HR (Aymard et al. 2014). Additionally, the hypothesis made at the beginning of this chapter was that pre-existing transcripts may have a role in DNA repair, meaning more active genes should be more likely to produce R-loops. To test this, each AsiSI site was placed in one of three equal sized groups based on transcriptional activity (low, medium, high): the pseudo-alignment and transcript quantification programme kallisto (Bray et al. 2016) was used to determine transcripts per million (TPM) for all genes in the control condition of the 4sU-seq dataset in Chapter 4. DRIP coverage was then calculated and plotted around the loci in each of these groups (Figure 5.8). The resulting data shows that increased transcriptional activity is required for DNA:RNA hybridisation, as the low TPM group had negligible changes following damage induction. Drosha depletion reduced damage-dependent hybrids at both medium and high TPM loci. This suggests pre-existing transcripts may form damage-induced R-loops, rather than *de novo* transcripts.

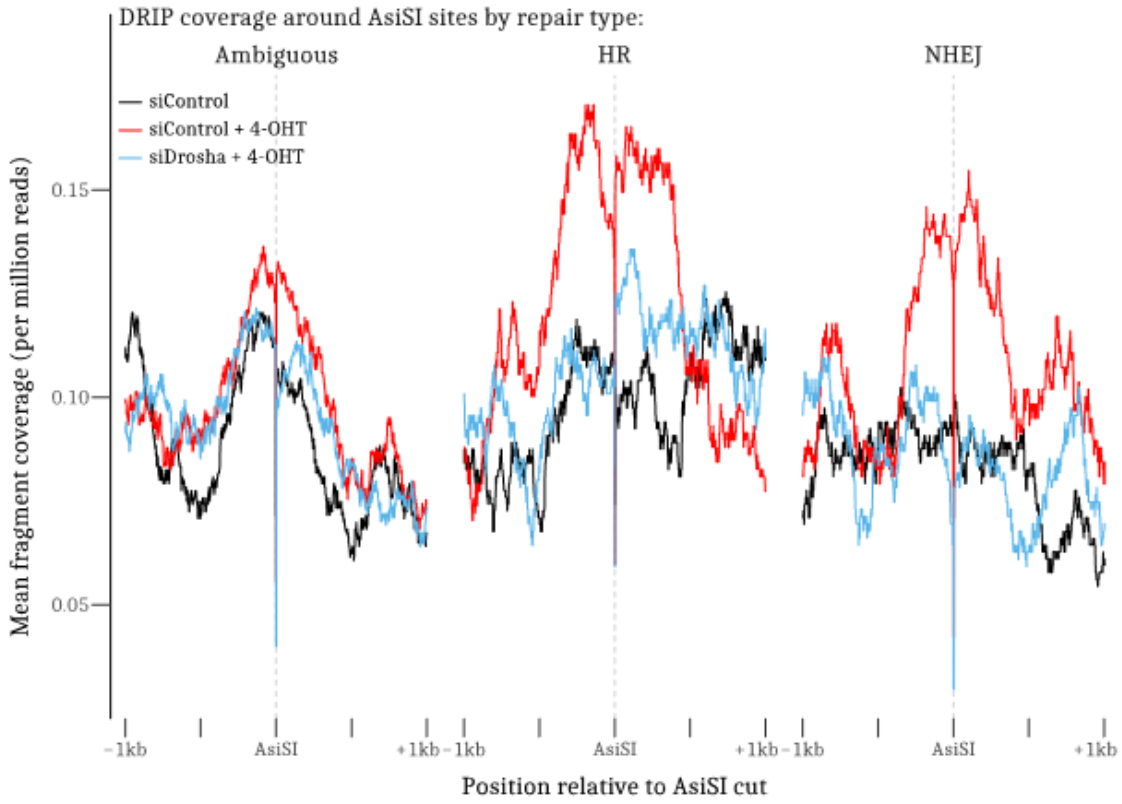


Figure 5.7. DNA:RNA hybrids form around both HR and NHEJ repaired cut sites.

Mean per-base coverage of DRIP fragments was calculated around ambiguous (i.e. not enriched or depleted for either HR or NHEJ) sites (n=49), HR sites (n=25), or NHEJ sites (n=25) using bedtools and plotted together for comparison. These sites are described in Table 8-5.

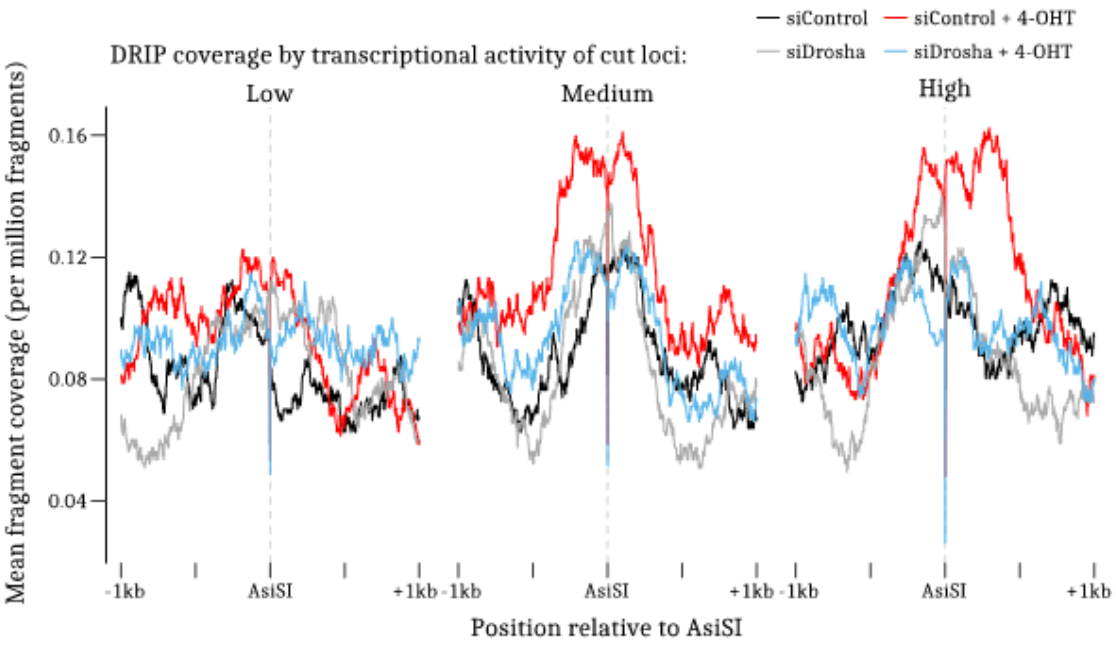


Figure 5.8. Increased basal transcriptional activity is required for damage-induced R-loop formation.

Mean per-base DRIP fragment coverage was calculated around subsets of AsiSI sites based on transcriptional activity. This was determined in section 2.5.6; briefly, the control 4sU-seq samples were used to sort all genes by transcripts per million (TPM). As this dataset enriched for all new transcripts in one hour, this is a measure of transcriptional activity. The 99 AsiSI sites were subset into groups of 33 by ascending TPM.

5.6 Damage-dependent DNA:RNA hybridisation may be an active process.

To further investigate whether transcription activity alone contributes to DNA:RNA hybrid formation, individual AsiSI cleaved genes of each repair type were chosen: *ASXL1* (HR), *CCBL2* (NHEJ), and *TRIM37* (ambiguous), and *AGA* (NHEJ). For each locus, a 40 kb region centred around the AsiSI motif was binned into 100 nt windows and DRIP fragment coverage (as FPM) was calculated within each. Additionally, coverage of 4sU-seq data (section 4.9) (as RPM) was also calculated in the same windows, where positive values are RNA from the plus strand and negative from the minus strand. These datasets were plotted together, with a schematic of the gene (Figure 5.9 and Figure 5.10).

ASXL1 and *CCBL2* show an increase in DRIP signal at the break site following DNA damage (Figure 5.9), reflecting the aggregate plots previous shown (Figure 5.5, Figure 5.7, Figure 5.8). Interestingly, the DRIP signal does not appear to be correlated with increased RNA abundance from a particular region; both the upstream of *ASXL1* and the downstream of *CCBL2* have very low RNA reads originating from them yet DNA:RNA hybrid reads are maintained on either side of the cut site.

In contrast, cuts around *AGA* and *TRIM37* do not show damage-dependent DNA:RNA hybridisation around the AsiSI motif (Figure 5.10). *AGA* has very little DRIP coverage at all, even though there is transcription within that region. *TRIM37* is highly transcriptionally active, however the DRIP signal does not increase; there are actually less sequenced fragments compared to undamaged. In summary, the observation that active transcription alone does not necessarily mean R-loops will form after damage suggests this may be a selective process.

5.7 Increased DNA:RNA hybrid clearance is detrimental to DNA repair.

As Drosha clearly plays a role in DNA repair (Chapter 3), a hypothesis that Drosha promotes R-loop formation at sites of DNA damage and that these structures then aid the repair process was made. To validate the DRIP-seq observations that Drosha was required for R-loop formation, laser microirradiation was again utilised as in section 5.2. U2OS cells were seeded into coverglass wells and reverse transfected with control

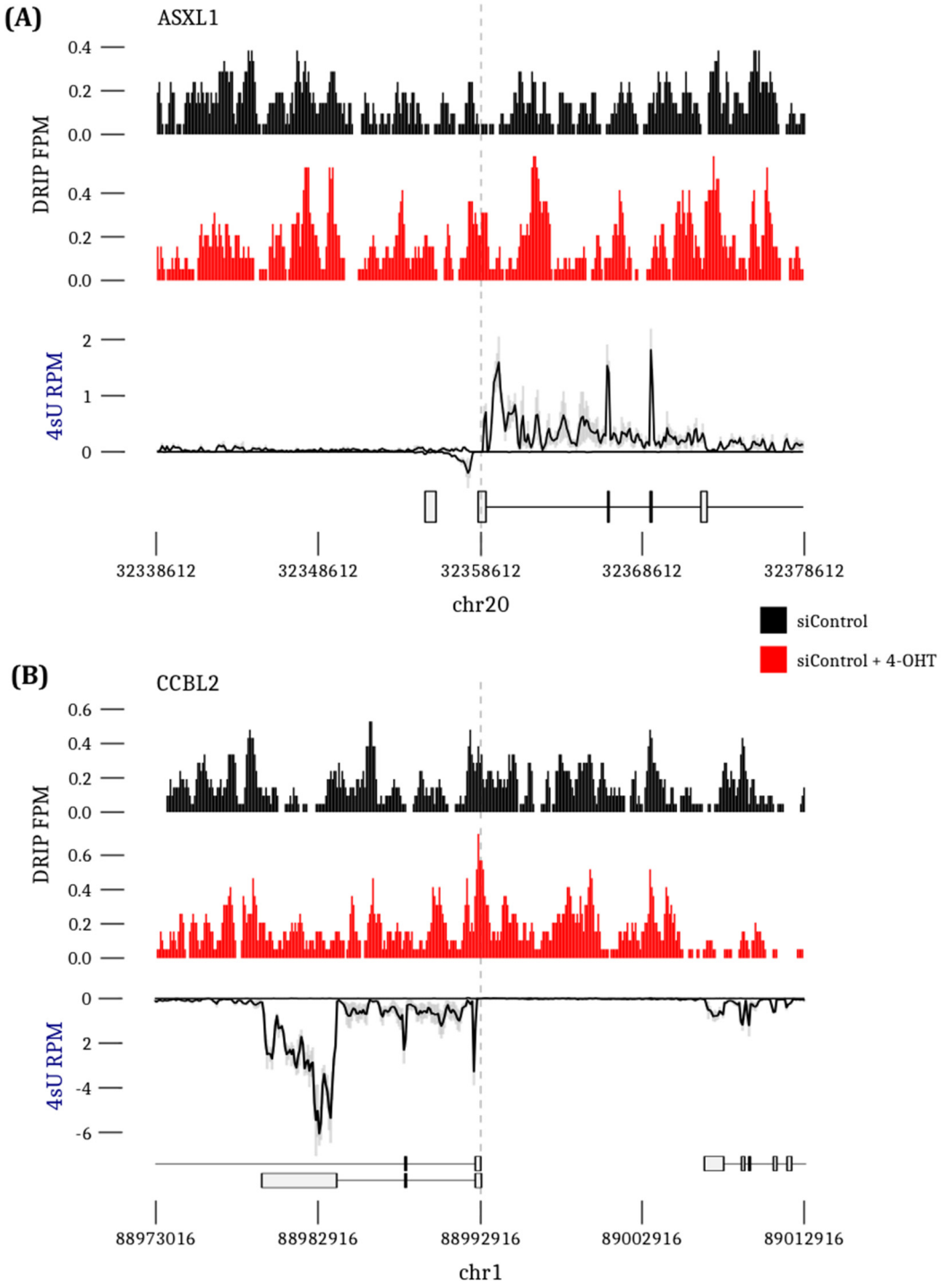


Figure 5.9. Two genes show increased DNA:RNA hybridisation at AsiSI cut sites.

(A) DRIP fragments per million (FPM) were plotted in a 40 kb region (binned into 100 nt windows) around the AsiSI cut in the *ASXL1* gene. 4sU-seq read coverage was also calculated (as reads per million; RPM). **(B)** As in (A), plotted around the AsiSI cut in *CCBL2*.

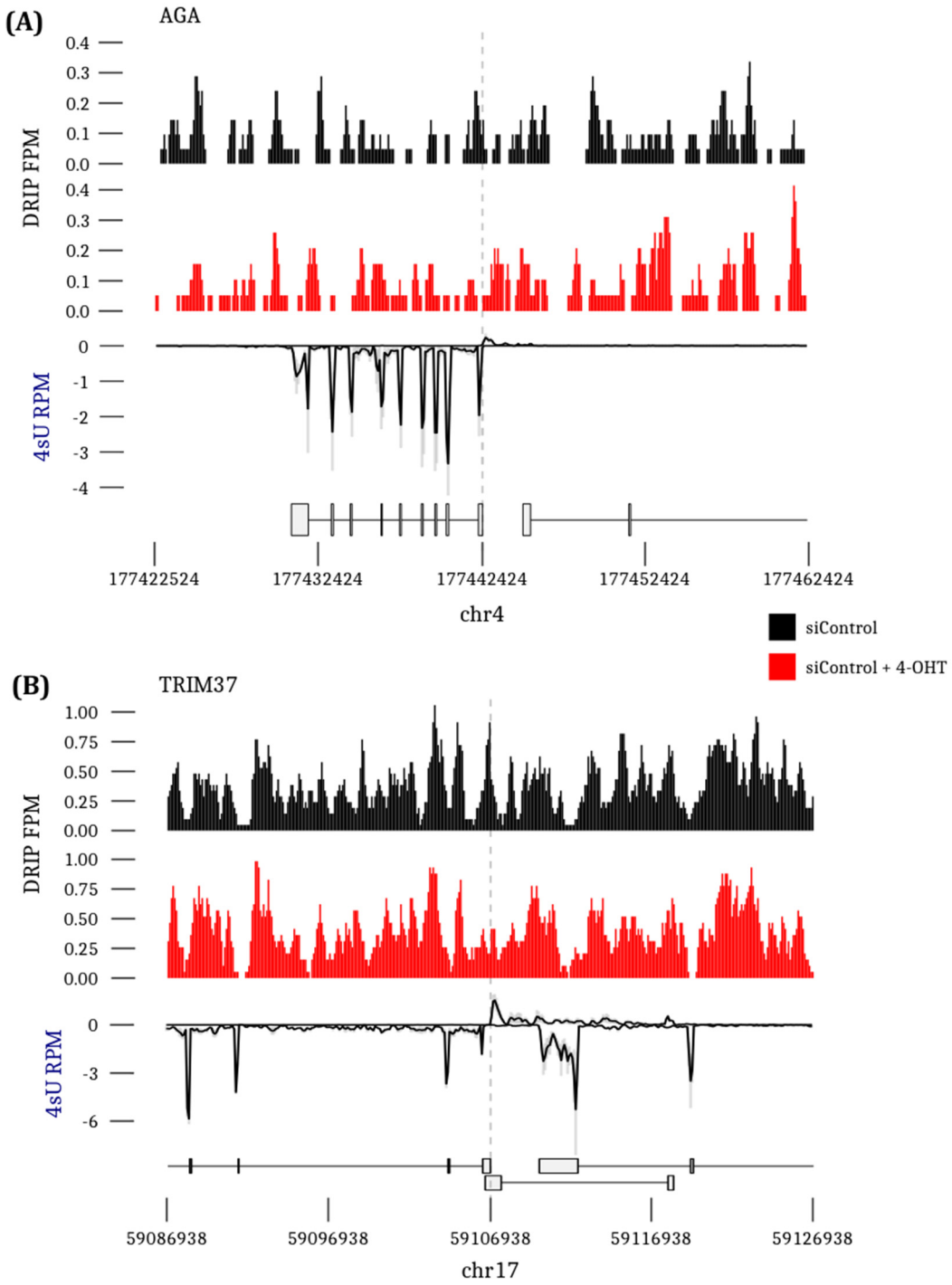


Figure 5.10. Two genes do not show increased DNA:RNA hybrids despite having active and ongoing transcription.

(A) As in Figure 5.9, plotted around the AsiSI cut in the *AGA* locus. Little DRIP signal can be seen throughout the locus, yet the 4sU track shows RNA reads. **(B)** DRIP fragments were plotted around the AsiSI cut in the *TRIM37* locus. This locus is highly transcriptionally active, but no DNA damage-induced R-loops form.

or Drosha and Dicer siRNA. The next day, RNH1^{mut} was forward transfected and expressed for a further 24 hours. Laser microirradiation induced rapid relocation of RNH1^{mut} in control siRNA cells (Figure 5.11A), as previously shown in Figure 5.1. Depletion of Drosha and Dicer however reduced the rate of this relocation, suggesting they are indeed required for R-loop formation after DNA damage.

To test whether DNA:RNA hybridisation is important for DNA repair, active RNase H1 was utilised to degrade R-loops *in situ*. RNase H1 was transfected into HR and NHEJ repair reporter cell lines (described in section 2.1.5) 6 hours prior to I-SceI transfection to initiate cleavage at the reporter locus. After 48 hours, repair efficiency was determined by measuring GFP positive cells by flow cytometry, where GFP is only expressed if repair is completed by that particular pathway. Over-expression of RNase H1 reduced repair by both HR (Figure 5.11B) and NHEJ (Figure 5.11C). This suggests that the formation of R-loops may be important for effective DNA repair.

5.8 Chapter discussion.

RNase H1 is the primary protein that remove R-loops in mammalian cells. It specifically recognises the DNA:RNA hybrid duplex and cleaves the RNA strand (Stein and Hausen 1969). Fluorescently tagged, catalytically dead mutant (D10R E48R) bacterial RNase H1 (RNH1^{mut}) retains the enzyme's affinity for R-loops. Using microirradiation, RNH1^{mut} recruitment to the laser stripes was observed, indicating DNA:RNA hybrids are forming in these regions following DNA damage (Figure 5.1). The S9.6 antibody – which specifically recognises the DNA:RNA duplex structure – was used for immunoprecipitation (DRIP) following activation of the AsiSI endonuclease, which confirmed the result from RNH1^{mut}.

By coupling this with NGS (DRIP-seq), DNA:RNA hybrids could be seen forming in the region closest to the DSB (Figure 5.3 and Figure 5.5). Transcription is inhibited following DNA damage in loci adjacent to the DSB (section 1.5.3) (Kruhlak et al. 2007; Shanbhag et al. 2010; Pankotai et al. 2012; Manfrini et al. 2015; Solovjeva et al. 2007; Iacovoni et al. 2010; Gong et al. 2015; Iannelli et al. 2017). It could be argued that this increase in R-loop signal could be due to the stalling of RNA pol II as a result of DNA damage rather than being an active process. Analysis revealed that while transcriptional activity is required for R-loop formation after damage, it alone is

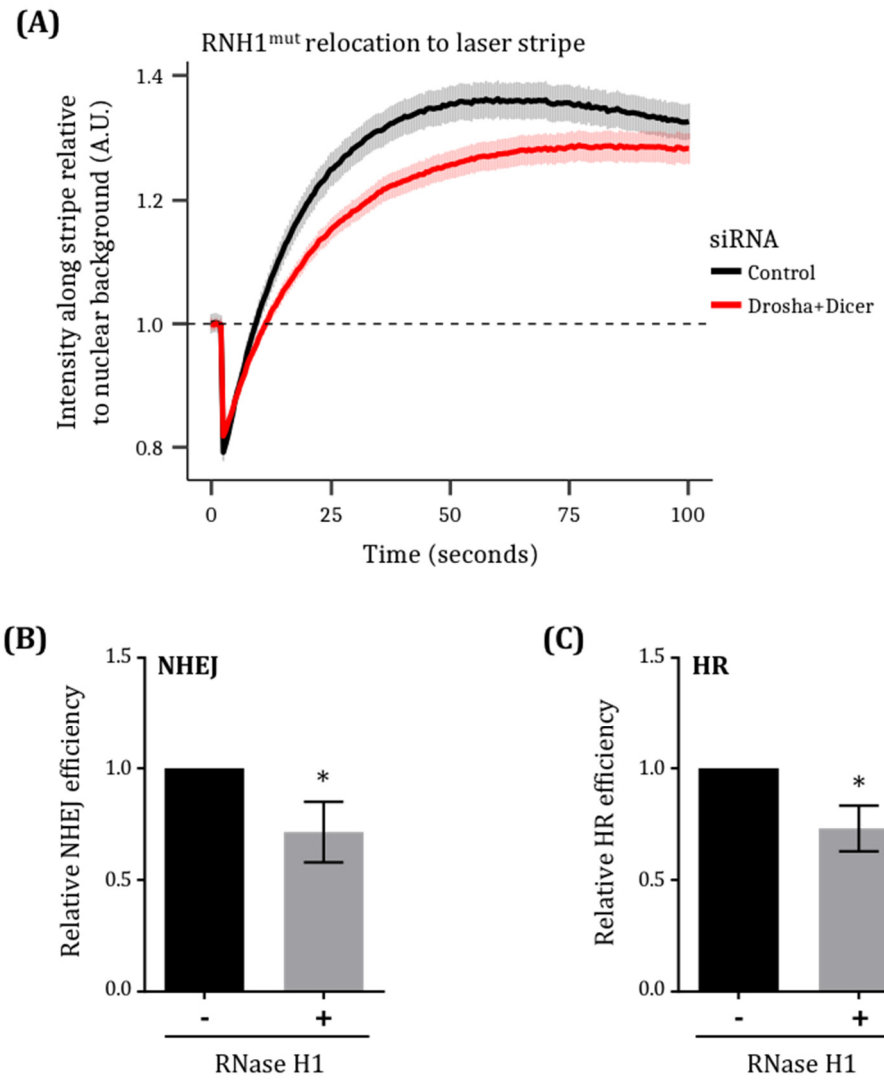


Figure 5.11. Damage- and Drosha-dependent R-loops are required for efficient DNA repair.

(A) U2OS cells were treated with control or Drosha and Dicer siRNAs for 24 hours. RNH1^{mut} was transfected for an additional 24 hours before laser microirradiation. Intensity along the laser stripe was normalised to a background, undamaged stripe in the same nucleus. 106 cells in the control and 104 in the Drosha and Dicer condition were quantified across 3 biological replicates. Mean±SEM. **(B, C)** Active RNase H1 was over-expressed for 6 hours in the U2OS-NHEJ or U2OS-HR reporter cell lines before induction of reporter damage by transfection with I-SceI. 48 hours later, cells were quantified for NHEJ/HR repair by measuring GFP-positive cells by flow cytometry. N=3, mean±SD, *p<0.05, One-sample t-test comparing to the normalised control value of 1. Error=SD.

not the only determinant of this process. The data presented in this chapter suggests that there may be a requirement for the hybridisation of RNA to DNA at break sites for efficiency repair, and this may be driven by an active process.

Depletion of Drosha abrogates the increase in R-loops around DSBs following DNA damage. This would suggest that Drosha is important in facilitating this interaction. As Drosha has been demonstrated to be involved in transcriptional termination (Dhir et al. 2015), it is possible that it cleaves the RNA to initiate its invasion into the chromatin. Alternatively, Drosha may aid in retention of the DNA:RNA hybrid, perhaps by binding to the hybrid structure although this activity has not been observed in the literature. Arguing against this possibility was the observation that depletion of Drosha and Dicer reduced the rate at which $RNH1^{mut}$ relocated to a laser stripe (Figure 5.11A), suggesting it they are indeed required for formation of the structures. Repeating these experiments with much longer timepoints will be necessary to see whether catabolism of the R-loop is affected in any way on depletion of these factors. Importantly, as Drosha was demonstrated to be involved in DNA repair (Chapter 3), this suggested that damage-induced DNA:RNA hybridisation may be important for DNA repair. These results are somewhat paradoxical as R-loop formation is typically associated with genomic instability (Skourti-Stathaki and Proudfoot 2014); indeed, loss of R-loop processing factors is associated with progression of disease such as Fragile X Syndrome (Groh et al. 2014). However, over-expression of active RNase H1 reduced both HR and NHEJ repair efficiency. As R-loops formed around both HR and NHEJ sites, this provides strong evidence for an involvement of RNA directly in the repair process of DSBs.

From this data, a model that active DNA:RNA hybridisation around DNA break sites is driven, in some part by Drosha, to aid in the efficient repair of those breaks. Whether this is through a scaffolding, signalling, or as the first step for RNA templated repair is yet to be determined.

6 Chapter 6: Discussion

The advent of next generation sequencing approaches has led to the identification of a diverse range of RNA species, many of which are still be to be fully characterised. Through this, it is now clear that RNA plays a central role in many if not all biological processes. Surprisingly, its participation in DNA repair has been barely explored despite the numerous lines of evidence pointing to transcription and RNA processing enzymes in DNA repair (Matsuoka et al. 2007; Paulsen et al. 2009; Hurov, Cotta-Ramusino, and Elledge 2010; Slabicki et al. 2010; Cotta-Ramusino et al. 2011; Adamson et al. 2012; Aymard et al. 2014; Aymard et al. 2017) (section 1.5). When considering the similar biochemical nature of RNA to DNA, and that RNA is a complimentary copy of its template DNA, one can envision many ways in which an RNA could be involved in DNA repair.

The aims of this thesis (section 1.6) were initially to validate and consolidate the previously reported roles for microRNA pathway components in DNA repair and to investigate whether the proposed *de novo* transcription at sites of damage was a general feature of DSB repair at endogenous loci. Here, these were explored using an array of experimental approaches focusing heavily on the use of NGS technologies. These findings led to the development of a new hypothesis which highlights the potential involvement of pre-existing transcripts in the repair of the damaged gene. This chapter will bring together all the data and discussions in Chapters 3-5 to propose future work and a model for the participation of RNA in DNA repair.

6.1 MicroRNA biogenesis enzymes are required for DNA repair without the generation of a small RNA product.

Reports of damage-induced small RNA have been remarkably controversial in both the DNA repair and RNA fields, with several key outstanding questions which this project aimed to answer.

Firstly, what would the mechanism of these small RNAs be? Some publications suggested a direct role in the DDR/repair pathway choice (Francia et al. 2012; W. Wei et al. 2012; Rossiello et al. 2017), whereas others proposed that these would facilitate the degradation of aberrant transcripts (Lee et al. 2009; Michalik, Böttcher, and Förstemann 2012). For the latter model, experiments showed that the presence of a

linearised plasmid led to repression of another reporter plasmid, with the conclusion that small RNA produced from damaged DNA could act as siRNA *in trans* (Michalik, Böttcher, and Förstemann 2012). However, how such a mechanism could assist in the repair of a DNA break was unclear. In the former model, damage-induced small RNAs were suggested to directly target early DDR factors (MDC1, ATM; section 1.2.2) to DSBs (Francia et al. 2012); in conflict with this, a different group proposed this occurred much later in the DDR pathway to specifically direct RAD51 to DSBs for HR repair (Gao et al. 2014). Regardless of the downstream mode of action, all these models necessitate that DSBs would induce new transcription from around the break, which would then be processed into small RNA by Drosha and Dicer. In Chapter 3, the participation of Drosha and Dicer in DNA repair was confirmed. Immunofluorescence studies in two cell lines with different damaging agents showed that the recruitment of 53BP1 was perturbed upon loss of Drosha and Dicer (Figure 3.2). The lack of a role for the TNRC6 proteins (which is absolutely crucial for miRNA-mediated repression, as verified in Figure 3.3)) demonstrated that this effect was not a consequence of the canonical miRNA pathway. The lack of an observed effect on γH2AX foci formation in all cases argued against a participation in the initial DDR cascade (Figure 3.2; section 1.2.2); this was then further confirmed by IF of pATM and MDC1 foci (Figure 3.4). These findings were discordant with both the early (Francia et al. 2012) and late (W. Wei et al. 2012) factor disputes, and instead argue that Drosha and Dicer may participate at a more central part of the DDR pathway, potentially at the chromatin modification/remodelling stage. This is further backed up by the observations that factors that compete with 53BP1 also have reduced focal recruitment (e.g. BRCA1; Figure 3.5) and that both major competing repair pathways – NHEJ and HR – are also perturbed (Figure 3.13).

The second major outstanding question was whether damage-induced small RNA generation is a general feature of all DSBs. All of these studies described experiments performed many hours after damage induction (from 12 hours to several days), begging the question of whether the detected RNAs were a primary event (i.e. made early in the DDR to participate in the DDR) or a secondary event (produced after damage repair, from degradation of aberrant transcripts, or as a result of the initiation of apoptosis due to prolonged DNA damage). Furthermore, these studies used

primarily integrated reporter constructs with high expression levels or highly repetitive sequences (W. Wei et al. 2012; Francia et al. 2012), transient transfection of many copies of plasmids (Michalik, Böttcher, and Förstemann 2012), or repetitive ribosomal/telomeric loci (Lee et al. 2009; Rossiello et al. 2017). With GFP-based reporters under control of a CMV promoter, the huge abundance of the transcribed mRNA makes inference of transcriptional changes difficult. Similarly, the transient transfection of plasmids, even if these are low expressing as in this study, would produce large quantities of RNA due to the vast number of plasmid copies per cell (Michalik, Böttcher, and Förstemann 2012)). The repair of plasmids is also unrepresentative of normal DNA damage, which occurs within chromatin. Repetitive sequences face a similar problem due to copy number; additionally, both DNA repair and the chromatin context is different at these loci compared to the bulk of the genome (Watts 2016). A final general concern with all of these studies was that the model for small RNA generation would require new transcription following DNA damage. How this can be reconciled with the observations that transcription is locally inhibited upon DNA damage is yet to be determined (section 1.5.3). It was therefore critical for the furthering of this field that damage-induced *de novo* transcription and small RNA products be studied in the context of endogenous, non-repetitive loci. This thesis utilised the AsiSI endonuclease system in U2OS cells, which cuts robustly at around 100 loci in the genome and has been extensively characterised by the Legube lab (Figure 3.6) (Iacovoni et al. 2010; Aymard et al. 2014; Aymard et al. 2017). As discussed, a role for Drosha and Dicer in the DDR was confirmed in this system (section 3.6 and above). However, deep sequencing of small RNA did not reveal any evidence for new small RNA products following induction of damage. Comparison of the annotated miRNA between these datasets and those in the previously published studies showed comparable, if not greater, coverage of small RNA showing these experiments had at least equal sensitivity (Figure 4.7). Much earlier time points (1 and 6 hours) were also used alongside the later timepoints (12 and 24 hours) to address the question of whether the generation of small RNA was a primary or secondary event. Differential expression analysis of miRNAs confirmed expected effects following DNA damage, showing that the DSBs induced by AsiSI is representative of typical sources of DNA damage (Figure 4.5). Additionally, two biological replicates were used

so that robust statistical testing could be performed; no statistical differences could be seen following damage at the 99 cut loci (Figure 4.9). This testing is described in more detail in Methods section 2.5.4. These results strongly suggested that small RNA generation was not the mechanism of Drosha and Dicer in DNA damage repair. However, because a number of groups have described these small RNAs in many organisms, a yet more comprehensive search was undertaken. In *Neurospora* and *Arabidopsis*, RNA-dependent RNA polymerases (RdRPs) were reported to be required for the small RNA generation. No dedicated RdRP exists in humans (Lee et al. 2009; W. Wei et al. 2012). In the literature, there is some evidence that RNA polymerase II may possess some minor RdRP activity (Lehmann, Brueckner, and Cramer 2007; Wagner et al. 2013), raising the possibility that RNA amplification mechanisms may exist. To test this, attention was turned to the chemistry of RdRP-amplified small RNA products, such as those from the siRNA pathway in *Neurospora* and *Arabidopsis*. Secondary small RNAs typically possess a 5'-triphosphate, in contrast to miRNA-like small RNAs which possess 5'-monophosphates. As standard small RNA-seq library preparations utilise a 5'-monophosphate, a method was developed to sequence these (section 4.6). This did not substantiate such a mechanism nor reveal the existence of any damage-induced small RNA with 5'-triphosphates (Figure 4.16).

This study did not find any evidence for damage-induced small RNA in endogenous genomic loci, suggesting that the role of Drosha and Dicer in DNA repair is not to process a *de novo* transcript into small RNA. Instead, the possibility that a precursor may be involved in the DDR was considered. There had previously been no direct evidence of such a longer nascent transcript. The approach undertaken in this thesis was to enrich for all new transcripts following DNA damage, so that abundant, steady-state RNA would not dominate the sequenced reads in an NGS experiment. 4-thiouridine (4sU) is a uridine analogue that can be used in very short pulse treatments to effectively tag all nascent transcripts (Dolken et al. 2008; Windhager et al. 2012; Rädle et al. 2013). The thiol moiety can be easily biotinylated in an oxidation reaction, which in turn can be effectively enriched from the total RNA using streptavidin-coated beads. Induction of damage via activation of the endonuclease AsiSI by the addition of 4-OHT at the same time as 4sU so that any *de novo* transcription – whether it occurs immediately or at a later point in the DDR – could be

tagged as damage occurs. The rationale behind this was that Drosha affected 53BP1 recruitment (Figure 3.2), which laser microirradiation experiments have demonstrated is recruited within minutes of damage induction (Bekker-Jensen et al. 2005; Hansen et al. 2016). If the role of Drosha in the DDR was through its processing of an RNA molecule, it would need to be produced prior to this. Sequencing of 4sU-enriched RNA did not provide strong evidence for *de novo* transcription following DNA damage (Figure 4.24). Analysis of differentially expressed transcripts demonstrated the expected transcriptional changes in response to DNA damage (i.e. increased levels of mRNA involved in the cell cycle (Figure 4.20)), showing that this methodology indeed enriched for new transcripts following DNA damage. The results from this study suggests that the previously published models of damage-induced transcription may not be a general feature of DNA damage, but instead an artefact of the reporter systems used or specific to damage repetitive sequences. For example, repetitive loci are typically condensed in heterochromatin (Trojer and Reinberg 2007). Damage induced here will cause decondensation (Watts 2016), which can result in increased transcription due to the increased availability of the sequences (Vaňková Hausnerová and Lanctôt 2017). Alternatively, HR repair in repetitive loci can result in a decrease or increase in repeat copy number (Hastings et al. 2009). As these studies looked many hours after damage, a copy number duplication after repair has been conducted in the damaged condition would ultimately result in more transcripts with the same sequence. A similar situation could arise from the integrated reporter loci (W. Wei et al. 2012); additionally, the highly active CMV promoters here produce vast quantities of RNA which, after induction of apoptosis due to prolonged DNA damage, would be degraded (Del Prete et al. 2002; Thomas et al. 2015). Importantly, his has been demonstrated in a recent publication co-authored by the corresponding author from the previous publications: small RNA production in a reporter correlated with transcriptional activity of the gene (but was not required for repair) and crucially, that this could not be detected in endogenous loci (Miki et al. 2017).

It is possible that even the 4sU-enriched RNA technique was not sensitive enough to detect *de novo* transcripts, especially if a situation arises where only one transcriptional initiation event occurs per DSB per cell or the transcripts are very short lived. Potentially, using a shorter 4sU/4-OHT pulse could address this; alternatively, a

technique such as NET-seq carried out early after damage induction may capture those initiating RNA polymerases with the nascent transcript (Nojima et al. 2015). The ultimate experiment, for both identifying primary transcripts or any processed products, would be to utilise single cell sequencing. This technique is now out of its infancy, and has been used to quantify exact numbers of specific RNA molecules per cell (Faridani et al. 2016; Gawad, Koh, and Quake 2016). As it relies on multiple technical replicates (i.e. individual cells) or conditions, many timepoints after damage induction – and using different endonucleases or CRISPR – could be tested to definitively answer whether new transcription is a true biological phenomenon or an artefact of the non-endogenous systems used in previous studies.

Nevertheless, the 4sU-seq experiment and the several small RNA sequencing experiments – alongside the previously mentioned recent publication (Miki et al. 2017) – do not support the existence of *de novo* transcription or processed products (Chapter 4). The role for Drosha and Dicer in DNA repair (Chapter 3) thus stems from either a direct interaction with DDR factors, or more likely, involves a transcript prior to DSB induction.

6.2 An associated nascent transcript may be involved in the repair of a DSB occurring in its gene.

In Chapter 5, DNA damage induction was demonstrated, using two main methods, to induce invasion of RNA into DNA, forming a DNA:RNA hybrid. A fluorescent but catalytically dead RNase H1 (RNH^{mut}) that can still bind DNA:RNA hybrids was used to visualise R-loop via microscopy (Britton et al. 2014). Laser microirradiation caused a rapid relocation of this to the damaged stripe, indicating the formation of R-loops at the damaged DNA (Figure 5.1, Figure 5.11A). The S9.6 antibody can also recognise DNA:RNA hybrids and was used to immunoprecipitate these structures (DRIP). This confirmed that R-loops form in the vicinity of DSBs (Figure 5.2A). This was recently demonstrated directly in yeast (Ohle et al. 2016) and inferred in human cells by RNase H1 digestion followed by either 53BP1 RNA IP (RIP) (Chakraborty et al. 2016) or immunofluorescence (L. Wei et al. 2015), making this the first direct observation in human cells. All three of these publications conclude that these structures are important for the repair of DNA breaks, which will be discussed later.

Further characterisation of these R-loop-containing DNA fragments was carried out by deep sequencing (DRIP-seq) and bioinformatic analyses. The data revealed that damage-induced R-loop formation occurs proximally to the break (Figure 5.5A), suggesting a central role for RNA invasion in DSB repair. This did not appear to be dependent on repair pathway choice, as both HR- and NHEJ-prone DSBs had increased DRIP fragment coverage (Figure 5.7). Curiously, those that were assigned ambiguous repair appeared to have much less hybridisation (Figure 5.7). These sites were depleted or enriched for both HR and NHEJ factors (Aymard et al. 2014). A likely explanation for these is that they are sites of very fast repair kinetics, potentially due to the chromatin landscape that makes them particularly amenable to factor recruitment. Such sites may not require R-loop formation for their repair, or due to the rapid repair may catabolise the structures reducing the likelihood that they can be immunoprecipitated.

Analysis of DRIP-seq also revealed that DNA:RNA hybrids are dependent on transcriptional activity, as would be expected (section 5.5). The lowest third of sites as scored by transcriptional activity showed only a very modest increase in these structures, whereas the remaining two-thirds showed large induction (Figure 5.8). These data indicate that more transcribed loci produce more R-loops, potentially reflecting the observations that transcriptional activity affects DNA repair (section 1.5.1) (Tang et al. 2013; Aymard et al. 2014; Aymard et al. 2017). On the other hand, it may simply suggest that R-loops form after DNA damage by the blocking or stalling of RNA polymerases (section 1.5.3). To address this, a selection of individual sites was examined. This revealed that transcription alone was not the sole determinant of RNA invasion in the DNA, suggesting that this is an active process (Figure 5.9 & Figure 5.10).

Importantly for this thesis, this work also demonstrated that Drosha is required in some part for the formation or retention of R-loops (Figure 5.5A & Figure 5.11A). Drosha has not been described to possess DNA:RNA hybrid binding activity in the literature; however, it has – as has its yeast relative Rnt1 – been shown to participate in transcriptional termination by cleaving hairpin structures (Gromak et al. 2013; Kawauchi et al. 2008). Furthermore, analysis of the 4sU-seq dataset in section 4.9 showed that Drosha may participate in transcriptional regulation (Figure 4.21 and Figure 4.22). As Drosha was required for R-loop formation or retention and for

efficient DNA repair, the hypothesis was made that the role for Drosha in DSB repair is through this R-loop-related activity, and thus DNA:RNA hybrids are involved in the DDR. To test this, active RNase H1 was over-expressed in the HR and NHEJ repair reporter cell lines. As RNase H1 degrades the RNA component of hybrids, this should degrade them as they form. Both HR and NHEJ repair efficiency was reduced, confirming the role of DNA:RNA hybrids in the DDR (Figure 5.11BC). This reflects the observations in yeast that demonstrated a DDR defect on over-expression of RNase H1 (Ohle et al. 2016). R-loops are typically considered detrimental to genomic stability (Skourti-Stathaki and Proudfoot 2014). Here, a beneficial role for DNA:RNA hybridisation at DNA break sites has been established.

6.3 A model for damage-dependent R-loop formation facilitated by Drosha.

In this thesis Drosha was demonstrated to participate in the DNA damage response outside of its well-established role in the processing of miRNA (Chapter 3). A comprehensive RNA sequencing strategy did not show evidence of small or larger damage-induced transcription products produced from the regions around breaks in endogenous loci (Chapter 4). In Chapter 5 however, RNA was demonstrated to invade the DNA duplex at double-strand break sites, and this appeared to be a part of the DNA damage response and was facilitated by Drosha.

How would DNA:RNA hybrids around DSBs participate in the DDR? One possibility is that these structures may act to signal that the DSB is an actively transcribed locus, thus affecting the repair pathway to be engaged (Tang et al. 2013; Aymard et al. 2014). Instead, a tethered RNA may have a more dynamic role by potentially participating in the homology search during HR (see section 1.2.6.3): an increased number of copies of the sequence could speed up what is the rate-limiting step of HR repair sequencing (Renkawitz et al. 2013). Alternatively, it could be more passive, simply acting as a scaffold for factors such as 53BP1 which require an RNA component (Pryde 2005). The most exciting possibility (reviewed in section 1.5.4) would be the this is the first step for RNA-templated DNA synthesis in a novel repair pathway. Preliminary work in yeast has indicated that this mechanism may be possible, and RNA-dependent DNA synthesis absolutely requires hybrid duplex formation (Storici et al. 2007; Keskin et al.

2014). In humans, two reports have produced rudimentary data that could be evidence of RNA-templated repair; many more controls and experiments will be required to demonstrate that this is possible (L. Wei et al. 2015; Chakraborty et al. 2016). However, at the time of writing this thesis, a new publication presented evidence of new NHEJ/HR hybrid repair pathway that could be a significant piece of the puzzle. This was a G1-phase, resection-dependent repair pathway that required both NHEJ and HR components (Biehs et al. 2017). Both 53BP1 and BRCA1 were required and acted in concert to tightly regulate the extent of resection: both of these factors were demonstrated in this thesis to be perturbed on loss of Drosha (Figure 3.2 & Figure 3.5). Resection-mediated repair pathways require a template to synthesise the lost DNA (sections 1.2.6.4 and 1.2.7). As there is no sister chromatid is G1, this template must either be a non-homologous region or an RNA molecule; an opinion article by these same authors speculates that this would be RNA (Löbrich and Jeggo 2017). The paper also demonstrated that this mechanism works on sites with slow repair kinetics (Biehs et al. 2017). The DRIP-seq data presented in this thesis showed that R-loops form around HR- and NHEJ-prone sites, but not ambiguous repair sites (Figure 5.7). As discussed in Chapter 5, those ambiguous sites are likely sites of fast repair kinetics. This would suggest that R-loops form around slow repair loci, which could potentially be repaired by an RNA-templated repair mechanism.

A model for how Drosha and DNA:RNA hybridisation may participate in the DDR based on this is presented in Figure 6.1. As Drosha has not been demonstrated to possess any DNA:RNA binding or processing activity, it may be that it utilises its transcriptional termination activity (Dhir et al. 2015) to cleave a hairpin structure in the nascent RNA from the elongating or stalling Pol II. A recombinase such as RAD52 could then initiate strand exchange to form the R-loop, an activity that has been described (Mazina et al. 2017). A Drosha-RAD52 interaction has not been reported; however, mass spectrometry screens have not been performed following DNA damage. Drosha has been demonstrated to interact with BRCA1 to process a subset of miRNA (Kawai and Amano 2012). Through the BRCA1-PALB2-BRCA2 axis (see section 1.2.6.2) Drosha could recruit many recombinases such as RAD52; inversely, these could also recruit Drosha to DSBs. Upon DNA:RNA hybridisation, an RNA-dependent DNA polymerase would be required to synthesise new DNA. Several polymerases have already been

described to carry out this activity, such as TERT *in vivo* (Yamaguchi et al. 2005) and Pol η *in vitro* (Franklin et al. 2004).

Much of this is speculative and requires many additional experiments. Many of the experiments performed in this thesis should be repeated in distinctly G1 vs. G2 populations of cells by the use of thymidine block or cell sorting. Drosha immunoprecipitations before and after damage will be necessary to determine whether it can interact with factors such as RAD52. The RNA component of the damage-induced R-loops is presumably unspliced, as this could lead to genomic deletions; sequencing of the RNA following DRIP would address this question. It could also identify whether the hybridised RNA has been modified in some way that may aid in its downstream function. Finally, direct evidence of RNA-templated repair could be obtained by sequencing of the genomic DNA following breaks induced by AsiSI and their repair, in a system where an RNA with or without point mutations could be tethered to the genome.

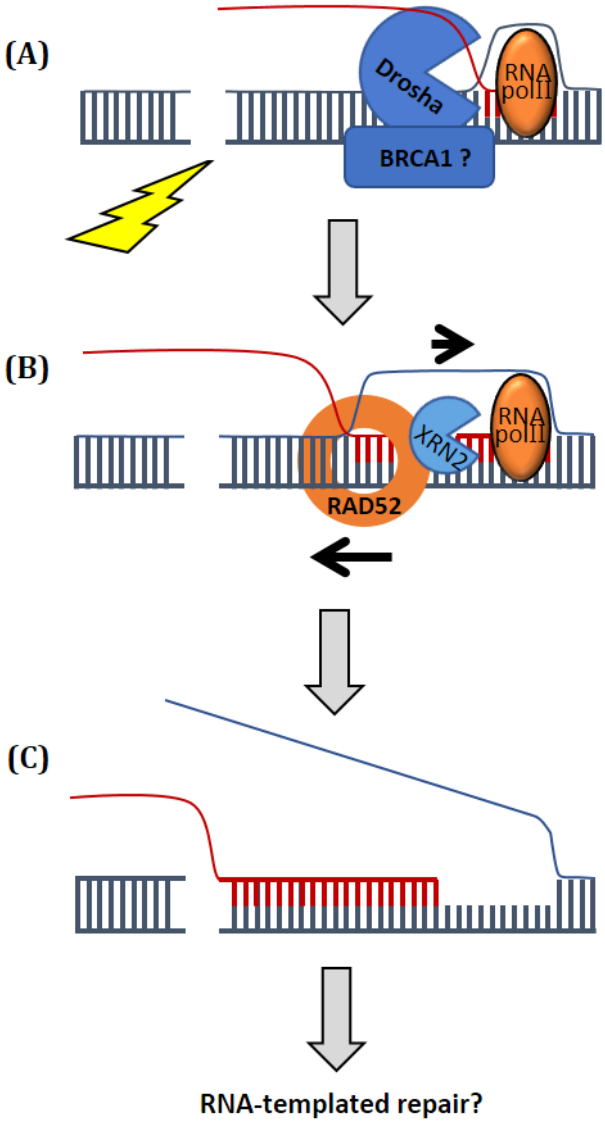


Figure 6.1. A model for damage-induced R-loop formation facilitated by Drosha in DNA repair.

(A) Upon DNA damage induction in an actively transcribing gene locus, Drosha cleaves hairpin structures in the nascent RNA. This may be facilitated by BRCA1. **(B)** This cleavage triggers XRN2 torpedo termination of Pol II, clearing the damaged locus. RAD52, either recruited through Drosha or BRCA1, carries out strand exchange with the RNA, forming R-loops. **(C)** This may extend all the way to the break site. This may then act as a template for DNA synthesis.

7 Bibliography

Adamson, Britt, Agata Smogorzewska, Frederic D. Sigoillot, Randall W. King, and Stephen J. Elledge. 2012. "A Genome-Wide Homologous Recombination Screen Identifies the RNA-Binding Protein RBMX as a Component of the DNA-Damage Response." *Nature Cell Biology* 14 (3): 318–28. doi:10.1038/ncb2426.

Ali, Ammar A E, Gyula Timinszky, Raquel Arribas-Bosacoma, Marek Kozlowski, Paul O Hassa, Markus Hassler, Andreas G Ladurner, Laurence H Pearl, and Antony W Oliver. 2012. "The Zinc-Finger Domains of PARP1 Cooperate to Recognize DNA Strand Breaks." *Nature Structural & Molecular Biology* 19 (7): 685–92. doi:10.1038/nsmb.2335.

Anders, Simon, Paul Theodor Pyl, and Wolfgang Huber. 2015. "HTSeq—a Python Framework to Work with High-Throughput Sequencing Data." *Bioinformatics* 31 (2): 166–69. http://dx.doi.org/10.1093/bioinformatics/btu638.

Arimbasseri, Aneeshkumar G, George A Kassavetis, and Richard J Maraia. 2014. "Comment on 'Mechanism of Eukaryotic RNA Polymerase III Transcription Termination.'" *Science* 345 (6196): 524 LP-524. http://science.sciencemag.org/content/345/6196/524.3.abstract.

Audebert, Marc, Bernard Salles, and Patrick Calsou. 2004. "Involvement of poly(ADP-Ribose) Polymerase-1 and XRCC1/DNA Ligase III in an Alternative Route for DNA Double-Strand Breaks Rejoining." *Journal of Biological Chemistry* 279 (53): 55117–26. doi:10.1074/jbc.M404524200.

Auyeung, Vincent C., Igor Ulitsky, Sean E. McGeary, and David P. Bartel. 2013. "Beyond Secondary Structure: Primary-Sequence Determinants License Pri-miRNA Hairpins for Processing." *Cell* 152 (4): 844–58. doi:10.1016/j.cell.2013.01.031.

Aymard, François, Marion Aguirrebengoa, Emmanuelle Guillou, Biola M Javierre, Beatrix Bugler, Coline Arnould, Vincent Rocher, et al. 2017. "Genome-Wide Mapping of Long-Range Contacts Unveils Clustering of DNA Double-Strand Breaks at Damaged Active Genes." *Nature Structural & Molecular Biology* 24 (4): 353–61. doi:10.1038/nsmb.3387.

Aymard, François, Beatrix Bugler, Christine K Schmidt, Emmanuelle Guillou, Pierre Caron, Sébastien Briois, Jason S Iacovoni, et al. 2014. "Transcriptionally Active

Chromatin Recruits Homologous Recombination at DNA Double-Strand Breaks." *Nature Structural & Molecular Biology* 21 (4): 366–74. doi:10.1038/nsmb.2796.

Baillat, David, and Ramin Shiekhattar. 2009. "Functional Dissection of the Human TNRC6 (GW182-Related) Family of Proteins." *Molecular and Cellular Biology* 29 (15): 4144–55. doi:10.1128/MCB.00380-09.

Bakkenist, Christopher J, and Michael B Kastan. 2003. "DNA Damage Activates ATM through Intermolecular Autophosphorylation and Dimer Dissociation." *Nature* 421 (6922): 499–506. doi:10.1038/nature01368.

Barber, Louise J, Jillian L Youds, Jordan D Ward, Michael J McIlwraith, Nigel J O'Neil, Mark I R Petalcorin, Julie S Martin, et al. 2008. "SPAR1/RTEL1 Maintains Genomic Stability by Suppressing Homologous Recombination." *Cell* 135 (2): 261–71. doi:10.1016/j.cell.2008.08.016.

Bartsch, I, C Schoneberg, and I Grummt. 1988. "Purification and Characterization of TTFI, a Factor That Mediates Termination of Mouse Ribosomal DNA Transcription." *Molecular and Cellular Biology* 8 (9): 3891–97. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC365448/.

Basehoar, Andrew D., Sara J. Zanton, and B. Franklin Pugh. 2004. "Identification and Distinct Regulation of Yeast TATA Box-Containing Genes." *Cell* 116 (5): 699–709. doi:10.1016/S0092-8674(04)00205-3.

Behm-Ansmant, Isabelle, Jan Rehwinkel, Tobias Doerks, Alexander Stark, Peer Bork, and Elisa Izaurralde. 2006. "mRNA Degradation by miRNAs and GW182 Requires Both CCR4:NOT Deadenylase and DCP1:DCP2 Decapping Complexes." *Genes and Development* 20 (14): 1885–98. doi:10.1101/gad.1424106.

Bekker-Jensen, Simon, Claudia Lukas, Fredrik Melander, Jiri Bartek, and Jiri Lukas. 2005. "Dynamic Assembly and Sustained Retention of 53BP1 at the Sites of DNA Damage Are Controlled by Mdc1/NFBD1." *Journal of Cell Biology* 170 (2): 201–11. doi:10.1083/jcb.200503043.

Belton, Jon Matthew, Rachel Patton McCord, Johan Harmen Gibcus, Natalia Naumova, Ye Zhan, and Job Dekker. 2012. "Hi-C: A Comprehensive Technique to Capture the Conformation of Genomes." *Methods* 58 (3): 268–76. doi:10.1016/j.ymeth.2012.05.001.

Bennardo, Nicole, Anita Cheng, Nick Huang, and Jeremy M. Stark. 2008. "Alternative-

NHEJ Is a Mechanistically Distinct Pathway of Mammalian Chromosome Break Repair." *PLoS Genetics* 4 (6). doi:10.1371/journal.pgen.1000110.

Bhargava, Ragini, David O. Onyango, and Jeremy M. Stark. 2016. "Regulation of Single-Strand Annealing and Its Role in Genome Maintenance." *Trends in Genetics* 32 (9): 566–75. doi:10.1016/j.tig.2016.06.007.

Biasiolo, Marta, Gabriele Sales, Marta Lionetti, Luca Agnelli, Katia Todoerti, Andrea Bisognin, Alessandro Coppe, Chiara Romualdi, Antonino Neri, and Stefania Bortoluzzi. 2011. "Impact of Host Genes and Strand Selection on miRNA and miRNA* Expression." *PLoS ONE* 6 (8): e23854. doi:10.1371/journal.pone.0023854.

Biehs, Ronja, Monika Steinlage, Olivia Barton, Szilvia Juhász, Julia Künzel, Julian Spies, Atsushi Shibata, Penny A. Jeggo, and Markus Löbrich. 2017. "DNA Double-Strand Break Resection Occurs during Non-Homologous End Joining in G1 but Is Distinct from Resection during Homologous Recombination." *Molecular Cell* 65 (4): 671–684.e5. doi:10.1016/j.molcel.2016.12.016.

Bieniossek, Christoph, Gabor Papai, Christiane Schaffitzel, Frederic Garzoni, Maxime Chaillet, Elisabeth Scheer, Petros Papadopoulos, Laszlo Tora, Patrick Schultz, and Imre Berger. 2013. "The Architecture of Human General Transcription Factor TFIID Core Complex." *Nature* 493 (7434): 699–702. http://dx.doi.org/10.1038/nature11791.

Bogenhagen, Daniel F., and Donald D. Brown. 1981. "Nucleotide Sequences in Xenopus 5S DNA Required for Transcription Termination." *Cell* 24 (1): 261–70. doi:10.1016/0092-8674(81)90522-5.

Bohgaki, M., T. Bohgaki, S. El Ghamrasni, T. Srikumar, G. Maire, S. Panier, A. Fradet-Turcotte, et al. 2013. "RNF168 Ubiquitylates 53BP1 and Controls Its Response to DNA Double-Strand Breaks." *Proceedings of the National Academy of Sciences* 110 (52): 20982–87. doi:10.1073/pnas.1320302111.

Bohnsack, Markus T, Kevin Czaplinski, and Dirk Gorlich. 2004. "Exportin 5 Is a RanGTP-Dependent dsRNA-Binding Protein That Mediates Nuclear Export of Pre-miRNAs." *RNA* 10 (2): 185–91. doi:10.1261/rna.5167604.

Boland, A., E. Huntzinger, S. Schmidt, E. Izaurralde, and O. Weichenrieder. 2011. "Crystal Structure of the MID-PIWI Lobe of a Eukaryotic Argonaute Protein." *Proceedings of the National Academy of Sciences* 108 (26): 10466–71.

doi:10.1073/pnas.1103946108.

Bray, N L, H Pimentel, P Melsted, and L Pachter. 2016. “Near-Optimal Probabilistic RNA-Seq Quantification.” *Nature Biotechnology* 34 (5): 525–27. doi:10.1038/nbt.3519.

Britton, Sébastien, Julia Coates, and Stephen P. Jackson. 2013. “A New Method for High-Resolution Imaging of Ku Foci to Decipher Mechanisms of DNA Double-Strand Break Repair.” *Journal of Cell Biology* 202 (3): 579–95. doi:10.1083/jcb.201303073.

Britton, Sébastien, Emma Dernoncourt, Christine Delteil, Carine Froment, Odile Schiltz, Bernard Salles, Philippe Frit, and Patrick Calsou. 2014. “DNA Damage Triggers SAF-A and RNA Biogenesis Factors Exclusion from Chromatin Coupled to R-Loops Removal.” *Nucleic Acids Research* 42 (14): 9047–62. doi:10.1093/nar/gku601.

Bunting, Samuel F, Elsa Callén, Nancy Wong, Hua-Tang Chen, Federica Polato, Amanda Gunn, Anne Bothmer, et al. 2010. “53BP1 Inhibits Homologous Recombination in Brca1-Deficient Cells by Blocking Resection of DNA Breaks.” *Cell* 141 (2): 243–54. doi:10.1016/j.cell.2010.03.012.

Bushnell, David A, Kenneth D Westover, Ralph E Davis, and Roger D Kornberg. 2004. “Structural Basis of Transcription: An RNA Polymerase II-TFIIB Cocrystal at 4.5 Angstroms.” *Science* 303 (5660): 983 LP-988. http://science.sciencemag.org/content/303/5660/983.abstract.

Cabart, Pavel, Andrea Ujvári, Mahadeb Pal, and Donal S. Luse. 2011. “Transcription Factor TFIIF Is Not Required for Initiation by RNA Polymerase II, but It Is Essential to Stabilize Transcription Factor TFIIB in Early Elongation Complexes.” *Proceedings of the National Academy of Sciences* 108 (38): 15786–91. doi:10.1073/pnas.1104591108.

Carleton, Michael, Michele A. Cleary, and Peter S. Linsley. 2007. “MicroRNAs and Cell Cycle Regulation.” *Cell Cycle* 6 (17): 2127–32. doi:10.4161/cc.6.17.4641.

Carninci, Piero, Albin Sandelin, Boris Lenhard, Shintaro Katayama, Kazuro Shimokawa, Jasmina Ponjavic, Colin A M Semple, et al. 2006. “Genome-Wide Analysis of Mammalian Promoter Architecture and Evolution.” *Nature Genetics* 38 (6): 626–35. doi:10.1038/ng1789.

Caron, Pierre, Jonathan Choudjaye, Thomas Clouaire, Béatrix Bugler, Virginie Daburon,

Marion Aguirrebengoa, Thomas Mangeat, et al. 2015. “Non-Redundant Functions of ATM and DNA-PKcs in Response to DNA Double-Strand Breaks.” *Cell Reports* 13 (8): 1598–1609. doi:10.1016/j.celrep.2015.10.024.

Carreira, A., and S. C. Kowalczykowski. 2011. “Two Classes of BRC Repeats in BRCA2 Promote RAD51 Nucleoprotein Filament Function by Distinct Mechanisms.” *Proceedings of the National Academy of Sciences* 108 (26): 10448–53. doi:10.1073/pnas.1106971108.

Chakraborty, Anirban, Nisha Tapryal, Tatiana Venkova, Nobuo Horikoshi, Raj K. Pandita, Altaf H. Sarker, Partha S. Sarkar, Tej K. Pandita, and Tapas K. Hazra. 2016. “Classical Non-Homologous End-Joining Pathway Utilizes Nascent RNA for Error-Free Double-Strand Break Repair of Transcribed Genes.” *Nature Communications* 7 (May): 13049. doi:10.1038/ncomms13049.

Chalkley, Gillian E, and C Peter Verrijzer. 1999. “DNA Binding Site Selection by RNA Polymerase II TAFs.” *EMBO Journal* 18 (17): 4835–45.

Chandradoss, Stanley D., Nicole T. Schirle, Malwina Szczepaniak, Ian J. Macrae, and Chirlmin Joo. 2015. “A Dynamic Search Process Underlies MicroRNA Targeting.” *Cell* 162 (1): 96–107. doi:10.1016/j.cell.2015.06.032.

Chen, Baohui, Luke A. Gilbert, Beth A. Cimini, Joerg Schnitzbauer, Wei Zhang, Gene Wei Li, Jason Park, et al. 2013. “Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System.” *Cell* 155 (7): 1479–91. doi:10.1016/j.cell.2013.12.001.

Chen, Mo, and James L. Manley. 2009. “Mechanisms of Alternative Splicing Regulation: Insights from Molecular and Genomics Approaches.” *Nature Reviews Molecular Cell Biology* 10 (11): 741–54. doi:10.1038/nrm2777.

Chen, Ran, and Marc S. Wold. 2014. “Replication Protein A: Single-Stranded DNA’s First Responder: Dynamic DNA-Interactions Allow Replication Protein A to Direct Single-Strand DNA Intermediates into Different Pathways for Synthesis or Repair Prospects & Overviews R. Chen and M. S. Wold.” *BioEssays* 36 (12): 1156–61. doi:10.1002/bies.201400107.

Chen, Zhucheng, Haijuan Yang, and Nikola P Pavletich. 2008. “Mechanism of Homologous Recombination from the RecA-ssDNA/dsDNA Structures.” *Nature* 453 (7194): 489–4. doi:10.1038/nature06971.

Chivers, Claire E., Apurba L. Koner, Edward D. Lowe, and Mark Howarth. 2011. “How the Biotin–streptavidin Interaction Was Made Even Stronger: Investigation via Crystallography and a Chimaeric Tetramer.” *Biochemical Journal* 435 (1): 55–63. doi:10.1042/BJ20101593.

Chou, D. M., B. Adamson, N. E. Dephoure, X. Tan, A. C. Nottke, K. E. Hurov, S. P. Gygi, M. P. Colaiacovo, and S. J. Elledge. 2010. “A Chromatin Localization Screen Reveals Poly (ADP Ribose)-Regulated Recruitment of the Repressive Polycomb and NuRD Complexes to Sites of DNA Damage.” *Proceedings of the National Academy of Sciences* 107 (43): 18475–80. doi:10.1073/pnas.1012946107.

Ciccia, Alberto, and Stephen J. Elledge. 2010. “The DNA Damage Response: Making It Safe to Play with Knives.” *Molecular Cell* 40 (2): 179–204. doi:10.1016/j.molcel.2010.09.019.

Clark, Michael B, Paulo P Amaral, Felix J Schlesinger, Marcel E Dinger, Ryan J Taft, John L Rinn, Chris P Ponting, et al. 2011. “The Reality of Pervasive Transcription.” *PLoS Biology* 9 (7): e1000625; discussion e1001102. doi:10.1371/journal.pbio.1000625.

Corden, Jeffry L. 1990. “Tails of RNA Polymerase II.” *Trends in Biochemical Sciences* 15 (10): 383–87. doi:http://dx.doi.org/10.1016/0968-0004(90)90236-5.

Cortez, D. 1999. “Requirement of ATM-Dependent Phosphorylation of Brca1 in the DNA Damage Response to Double-Strand Breaks.” *Science* 286 (5442): 1162–66. doi:10.1126/science.286.5442.1162.

Cotta-Ramusino, C., E. R. McDonald, K. Hurov, M. E. Sowa, J. W. Harper, and S. J. Elledge. 2011. “A DNA Damage Response Screen Identifies RHINO, a 9-1-1 and TopBP1 Interacting Protein Required for ATR Signaling.” *Science* 332 (6035): 1313–17. doi:10.1126/science.1203430.

Dalmay, T, A Hamilton, S Rudd, S Angell, and D C Baulcombe. 2000. “An RNA-Dependent RNA Polymerase Gene in Arabidopsis Is Required for Posttranscriptional Gene Silencing Mediated by a Transgene but Not by a Virus.” *Cell* 101 (5): 543–53. doi:S0092-8674(00)80864-8 [pii].

Davidson, Lee, Lisa Muniz, and Steven West. 2014. “3’ End Formation of Pre-mRNA and Phosphorylation of Ser2 on the RNA Polymerase II CTD Are Reciprocally Coupled in Human Cells.” *Genes and Development* 28 (4): 342–56. doi:10.1101/gad.231274.113.

De Laat, Wouter L., Esther Appeldoorn, Kaoru Sugasawa, Eric Weterings, Nicolaas G J Jaspers, and Jan H J Hoeijmakers. 1998. “DNA-Binding Polarity of Human Replication Protein A Positions Nucleases in Nucleotide Excision Repair.” *Genes and Development* 12 (16): 2598–2609. doi:10.1101/gad.12.16.2598.

DeFazio, Lisa G., Rachel M. Stansel, Jack D. Griffith, and Gilbert Chu. 2002. “Synapsis of DNA Ends by DNA-Dependent Protein Kinase.” *EMBO Journal* 21 (12). Oxford, UK: 3192–3200. doi:10.1093/emboj/cdf299.

Del Prete, M. Julieta, Maria S. Robles, Ana Guáo, Carlos Martínez-A, Manuel Izquierdo, and Jose A. Garcia-Sanz. 2002. “Degradation of Cellular mRNA Is a General Early Apoptosis-Induced Event.” *The FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology* 16 (14): 2003–5. doi:10.1096/fj.02-0392fje.

Dhir, Ashish, Somdutta Dhir, Nick J Proudfoot, and Catherine L Jopling. 2015. “Microprocessor Mediates Transcriptional Termination of Long Noncoding RNA Transcripts Hosting microRNAs.” *Nature Structural & Molecular Biology* 22 (4): 319–27. doi:10.1038/nsmb.2982.

Djebali, Sarah, Carrie A. Davis, Angelika Merkel, Alex Dobin, Timo Lassmann, Ali Mortazavi, Andrea Tanzer, et al. 2012. “Landscape of Transcription in Human Cells.” *Nature* 489 (7414): 101–8. doi:10.1038/nature11233.

Dolken, L., Z. Ruzsics, B. Radle, C. C. Friedel, R. Zimmer, J. Mages, R. Hoffmann, et al. 2008. “High-Resolution Gene Expression Profiling for Simultaneous Kinetic Parameter Analysis of RNA Synthesis and Decay.” *RNA* 14 (9): 1959–72. doi:10.1261/rna.1136108.

Downs, Jessica A., Stéphane Allard, Olivier Jobin-Robitaille, Ali Javaheri, Andréanne Auger, Nathalie Bouchard, Stephen J. Kron, Stephen P. Jackson, and Jacques Côté. 2004. “Binding of Chromatin-Modifying Activities to Phosphorylated Histone H2A at DNA Damage Sites.” *Molecular Cell* 16 (6): 979–90. doi:10.1016/j.molcel.2004.12.003.

Engreitz, Jesse M, Jenna E Haines, Elizabeth M Perez, Glen Munson, Jenny Chen, Michael Kane, Patrick E McDonel, Mitchell Guttman, and Eric S Lander. 2016. “Local Regulation of Gene Expression by lncRNA Promoters, Transcription and Splicing.” *Nature* 539 (7629): 452–55. http://dx.doi.org/10.1038/nature20149.

Fan, Pei, Zixuan Chen, Peng Tian, Wen Liu, Yan Jiao, Yi Xue, Anindya Bhattacharya, et al. 2013. “miRNA Biogenesis Enzyme Drosha Is Required for Vascular Smooth Muscle Cell Survival.” Edited by Maurizio Pesce. *PLoS ONE* 8 (4). San Francisco, USA: e60888. doi:10.1371/journal.pone.0060888.

Faridani, Omid R, Ilgar Abdullayev, Michael Hagemann-Jensen, John P Schell, Fredrik Lanner, and Rickard Sandberg. 2016. “Single-Cell Sequencing of the Small-RNA Transcriptome.” *Nature Biotechnology* 34 (12): 18–22. doi:10.1038/nbt.3701.

Fejes-Toth, Katalin, Vihra Sotirova, Ravi Sachidanandam, Gordon Assaf, Gregory J. Hannon, Philipp Kapranov, Sylvain Foissac, et al. 2009. “Post-Transcriptional Processing Generates a Diversity of 5′-Modified Long and Short RNAs.” *Nature* 457 (7232): 1028–32. doi:10.1038/nature07759.

Fishburn, J., and S. Hahn. 2012. “Architecture of the Yeast RNA Polymerase II Open Complex and Regulation of Activity by TFIIF.” *Molecular and Cellular Biology* 32 (1): 12–25. doi:10.1128/MCB.06242-11.

Fong, Nova, Kristopher Brannan, Benjamin Erickson, Hyunmin Kim, Michael A. Cortazar, Ryan M. Sheridan, Tram Nguyen, Shai Karp, and David L. Bentley. 2015. “Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition.” *Molecular Cell* 60 (2): 256–67. doi:10.1016/j.molcel.2015.09.026.

Foti, D., R. Iuliano, E. Chiefari, and A. Brunetti. 2003. “A Nucleoprotein Complex Containing Sp1, C/EBP , and HMGI-Y Controls Human Insulin Receptor Gene Transcription.” *Molecular and Cellular Biology* 23 (8): 2720–32. doi:10.1128/MCB.23.8.2720-2732.2003.

Fradet-Turcotte, Amélie, Marella D Canny, Cristina Escribano-Díaz, Alexandre Orthwein, Charles C Y Leung, Hao Huang, Marie-Claude Landry, et al. 2013. “53BP1 Is a Reader of the DNA-Damage-Induced H2A Lys 15 Ubiquitin Mark.” *Nature* 499 (7456): 50–54. doi:10.1038/nature12318.

Francia, Sofia, Flavia Michelini, Alka Saxena, Dave Tang, Michiel de Hoon, Viviana Anelli, Marina Mione, Piero Carninci, and Fabrizio d’Adda di Fagagna. 2012. “Site-Specific DICER and DROSHA RNA Products Control the DNA-Damage Response.” *Nature* 488 (7410): 231–35. doi:10.1038/nature11179.

Franklin, Andrew, Peter J Milburn, Robert V Blanden, and Edward J Steele. 2004.

"Human DNA Polymerase Eta, an A-T Mutator in Somatic Hypermutation of Rearranged Immunoglobulin Genes, Is a Reverse Transcriptase." *Immunol Cell Biol* 82 (2): 219–25. http://dx.doi.org/10.1046/j.0818-9641.2004.01221.x.

Froberg, John E., Lin Yang, and Jeannie T. Lee. 2013. "Guided by RNAs: X-Inactivation as a Model for lncRNA Function." *Journal of Molecular Biology* 425 (19): 3698–3706. doi:10.1016/j.jmb.2013.06.031.

Fu, Qinqin, and Y. Adam Yuan. 2013. "Structural Insights into RISC Assembly Facilitated by dsRNA-Binding Domains of Human RNA Helicase A (DHX9)." *Nucleic Acids Research* 41 (5): 3457–70. doi:10.1093/nar/gkt042.

Fuss, Jill O., and John A. Tainer. 2011. "XPB and XPD Helicases in TFIIH Orchestrate DNA Duplex Opening and Damage Verification to Coordinate Repair with Transcription and Cell Cycle via CAK Kinase." *DNA Repair* 10 (7): 697–713. doi:10.1016/j.dnarep.2011.04.028.

Gao, Min, Wei Wei, Ming-Ming Li, Yong-Sheng Wu, Zhaoqing Ba, Kang-Xuan Jin, Miao-Miao Li, et al. 2014. "Ago2 Facilitates Rad51 Recruitment and DNA Double-Strand Break Repair by Homologous Recombination." *Cell Research* 24 (5): 532–41. doi:10.1038/cr.2014.36.

Gawad, Charles, Winston Koh, and Stephen R. Quake. 2016. "Single-Cell Genome Sequencing: Current State of the Science." *Nature Reviews Genetics* 17 (3): 175–88. doi:10.1038/nrg.2015.16.

Giannattasio, Michele, Cindy Follonier, Hélène Tourrière, Fabio Puddu, Federico Lazzaro, Philippe Pasero, Massimo Lopes, Paolo Plevani, and Marco Muzi-Falconi. 2010. "Exo1 Competes with Repair Synthesis, Converts NER Intermediates to Long ssDNA Gaps, and Promotes Checkpoint Activation." *Molecular Cell* 40 (1): 50–62. doi:10.1016/j.molcel.2010.09.004.

Gong, Fade, Li Ya Chiu, Ben Cox, François Aymard, Thomas Clouaire, Justin W. Leung, Michael Cammarata, et al. 2015. "Screen Identifies Bromodomain Protein ZMYND8 in Chromatin Recognition of Transcription-Associated DNA Damage That Promotes Homologous Recombination." *Genes and Development* 29 (2): 197–211. doi:10.1101/gad.252189.114.

Grawunder, U, M Wilm, X Wu, P Kulesza, T E Wilson, M Mann, and M R Lieber. 1997. "Activity of DNA Ligase IV Stimulated by Complex Formation with XRCC4 Protein

in Mammalian Cells." *Nature* 388 (6641): 492–95. doi:10.1038/41358.

Groh, Matthias, Michele M P Lufino, Richard Wade-Martins, and Natalia Gromak. 2014. "R-Loops Associated with Triplet Repeat Expansions Promote Gene Silencing in Friedreich Ataxia and Fragile X Syndrome." Edited by Andrés Aguilera. *PLoS Genetics* 10 (5). San Francisco, USA: e1004318. doi:10.1371/journal.pgen.1004318.

Gromak, Natalia, Martin Dienstbier, Sara Macias, Mireya Plass, Eduardo Eyras, Javier F. Cáceres, and Nicholas J. Proudfoot. 2013. "Drosha Regulates Gene Expression Independently of RNA Cleavage Function." *Cell Reports* 5 (6): 1499–1510. doi:10.1016/j.celrep.2013.11.032.

Gulyaeva, Lyudmila F., and Nicolay E. Kushlinskiy. 2016. "Regulatory Mechanisms of microRNA Expression." *Journal of Translational Medicine* 14 (1). London: 143. doi:10.1186/s12967-016-0893-x.

Gunther, M, M Laithier, and O Brison. 2000. "A Set of Proteins Interacting with Transcription Factor Sp1 Identified in a Two-Hybrid Screening." *Molecular and Cellular Biochemistry* 210 (1–2): 131–42. doi:10.1023/A:1007177623283.

Gupta, Rajnish A., Nilay Shah, Kevin C. Wang, Jeewon Kim, Hugo M. Horlings, David J. Wong, Miao-Chih Tsai, et al. 2010. "Long Non-Coding RNA HOTAIR Reprograms Chromatin State to Promote Cancer Metastasis." *Nature* 464 (7291): 1071–76. doi:10.1038/nature08975.

Guzmán, E, and J T Lis. 1999. "Transcription Factor TFIIH Is Required for Promoter Melting in Vivo." *Molecular and Cellular Biology* 19 (8): 5652–58. doi:10.1128/MCB.19.8.5652.

Haince, Jean François, Darin McDonald, Amélie Rodrigue, Ugo Déry, Jean Yves Masson, Michael J. Hendzel, and Guy G. Poirier. 2008. "PARP1-Dependent Kinetics of Recruitment of MRE11 and NBS1 Proteins to Multiple DNA Damage Sites." *Journal of Biological Chemistry* 283 (2): 1197–1208. doi:10.1074/jbc.M706734200.

Han, Jinju, Yoontae Lee, Kyu-hyun Yeom, Young-kook Kim, Hua Jin, and V Narry Kim. 2004. "The Drosha – DGCR8 Complex in Primary microRNA Processing." *Genes & Development* 18 (24): 3016–27. doi:10.1101/gad.1262504.mic.

Han, Jinju, Jakob S. Pedersen, S. Chul Kwon, Cassandra D. Belair, Young Kook Kim, Kyu Hyeon Yeom, Woo Young Yang, David Haussler, Robert Blelloch, and V. Narry Kim.

2009. “Posttranscriptional Crossregulation between Drosha and DGCR8.” *Cell* 136 (1): 75–84. doi:10.1016/j.cell.2008.10.053.

Hangauer, Matthew J., Ian W. Vaughn, and Michael T. McManus. 2013. “Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs.” *PLoS Genetics* 9 (6). doi:10.1371/journal.pgen.1003569.

Hansen, Rebecca Kring, Andreas Mund, Sara Lund Poulsen, Maria Sandoval, Karolin Klement, Katerina Tsouroula, Maxim A. X. Tollenaere, et al. 2016. “SCAI Promotes DNA Double-Strand Break Repair in Distinct Chromosomal Contexts.” *Nature Cell Biology* 18 (12): 1357–66. doi:10.1038/ncb3436.

Hastings, P. J., James R. Lupski, Susan M. Rosenberg, and Grzegorz Ira. 2009. “Mechanisms of Change in Gene Copy Number.” *Nature Reviews Genetics* 10 (8): 551–64. doi:10.1038/nrg2593.

Hawley, Ben R, Wei-Ting Lu, Ania Wilczynska, and Martin Bushell. 2017. “The Emerging Role of RNAs in DNA Damage Repair.” *Cell Death and Differentiation*. doi:10.1038/cdd.2017.16.

He, Miao, Yu Liu, Xiaowo Wang, Michael Q. Zhang, Gregory J. Hannon, and Z. Josh Huang. 2012. “Cell-Type-Based Analysis of MicroRNA Profiles in the Mouse Brain.” *Neuron* 73 (1): 35–48. doi:10.1016/j.neuron.2011.11.010.

Henderson, Kate L., Lindsey C. Felth, Cristen M. Molzahn, Irina Shkel, Si Wang, Munish Chhabra, Emily F. Ruff, Lauren Bieter, Joseph E. Kraft, and M. Thomas Record. 2017. “Mechanism of Transcription Initiation and Promoter Escape by *E* . *Coli* RNA Polymerase.” *Proceedings of the National Academy of Sciences* 114 (15): E3032–40. doi:10.1073/pnas.1618675114.

Henriques, Telmo, Daniel Gilchrist, Sergei Nechaev, Michael Bern, Ginger Muse, Adam Burkholder, David Fargo, and Karen Adelman. 2013. “Stable Pausing by RNA Polymerase II Provides an Opportunity to Target and Integrate Regulatory Signals.” *Molecular Cell* 52 (4): 517–28. doi:10.1016/j.molcel.2013.10.001.

Hoeijmakers, Jan H J. 2009. “DNA Damage, Aging, and Cancer” 361 (15): 1475–85. doi:10.1056/NEJMra0804615.

Holstege, F C, P C van der Vliet, and H T Timmers. 1996. “Opening of an RNA Polymerase II Promoter Occurs in Two Distinct Steps and Requires the Basal

Transcription Factors IIE and IIH." *The EMBO Journal* 15 (7): 1666–77. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC450078/.

Hoogstraten, Deborah, Steven Bergink, Jessica M Y Ng, Vincent H M Verbiest, Martijn S Luijsterburg, Bart Geverts, Anja Raams, et al. 2008. "Versatile DNA Damage Detection by the Global Genome Nucleotide Excision Repair Protein XPC." *Journal of Cell Science* 121 (Pt 17): 2850–59. doi:10.1242/jcs.03503.

Hurov, Kristen E., Cecilia Cotta-Ramusino, and Stephen J. Elledge. 2010. "A Genetic Screen Identifies the Triple T Complex Required for DNA Damage Signaling and ATM and ATR Stability." *Genes and Development* 24 (17): 1939–50. doi:10.1101/gad.1934210.

Iacovoni, Jason S, Pierre Caron, Imen Lassadi, Estelle Nicolas, Laurent Massip, Didier Trouche, and Gaëlle Legube. 2010. "High-Resolution Profiling of gammaH2AX around DNA Double Strand Breaks in the Mammalian Genome." *The EMBO Journal* 29 (8): 1446–57. doi:10.1038/emboj.2010.38.

Iannelli, Fabio, Alessandro Galbiati, Ilaria Capozzo, Quan Nguyen, Brian Magnuson, Flavia Michelini, Giuseppina D'Alessandro, et al. 2017. "A Damaged Genome's Transcriptional Landscape through Multilayered Expression Profiling around in Situ-Mapped DNA Double-Strand Breaks." *Nature Communications* 8 (May): 15656. http://dx.doi.org/10.1038/ncomms15656.

Ira, Grzegorz, Dominik Satory, and James E Haber. 2006. "Conservative Inheritance of Newly Synthesized DNA in Double-Strand Break-Induced Gene Conversion ." *Molecular and Cellular Biology* 26 (24): 9424–29. doi:10.1128/MCB.01654-06.

Jasiak, Anna J., Karim Jean Armache, Birgit Martens, Ralf Peter Jansen, and Patrick Cramer. 2006. "Structural Biology of RNA Polymerase III: Subcomplex C17/25 X-Ray Structure and 11 Subunit Enzyme Model." *Molecular Cell* 23 (1): 71–81. doi:10.1016/j.molcel.2006.05.013.

Jena, N. R. 2012. "DNA Damage by Reactive Species: Mechanisms, Mutation and Repair." *Journal of Biosciences* 37 (3): 503–7. doi:10.1007/s12038-012-9218-2.

Jonkers, Iris, Hojoong Kwak, and John T Lis. 2014. "Genome-Wide Dynamics of Pol II Elongation and Its Interplay with Promoter Proximal Pausing, Chromatin, and Exons." Edited by Kevin Struhl. *eLife* 3: e02407. doi:10.7554/eLife.02407.

Jonkers, Iris, and John T. Lis. 2015. "Getting up to Speed with Transcription Elongation

by RNA Polymerase II." *Nature Reviews Molecular Cell Biology* 16 (3): 167–77. doi:10.1038/nrm3953.

Kapanidis, A. N., E. Margeat, S. O. Ho, E. Kortkhonjia, S. Weiss, and R. H. Ebright. 2006. "Initial Transcription by RNA Polymerase Proceeds Through a DNA-Scrunching Mechanism." *Science* 314 (5802): 1144–47. doi:10.1126/science.1131399.

Kapranov, Philipp, Jill Cheng, Sujit Dike, David A. Nix, Radharani Duttagupta, Aarron T. Willingham, Peter F. Stadler, et al. 2007. "RNA Maps Reveal New RNA Classes and a Possible Function for Pervasive Transcription." *Science* 316 (5830): 1484–88. doi:10.1126/science.1138341.

Kapranov, Philipp, Fatih Ozsolak, Sang Woo Kim, Sylvain Foissac, Doron Lipson, Chris Hart, Steve Roels, et al. 2010. "New Class of Gene-Termini-Associated Human RNAs Suggests a Novel RNA Copying Mechanism." *Nature* 466 (7306): 642–46. doi:10.1038/nature09190.

Kato, M, T Paranjape, R Ullrich, S Nallur, E Gillespie, K Keane, A Esquela-Kerscher, J B Weidhaas, and F J Slack. 2009. "The Mir-34 microRNA Is Required for the DNA Damage Response in Vivo in C. Elegans and in Vitro in Human Breast Cancer Cells." *Oncogene* 28 (25): 2419–24. doi:10.1038/onc.2009.106.

Kawahara, Yukio, Molly Megraw, Edward Kreider, Hisashi Iizasa, Louis Valente, Artemis G. Hatzigeorgiou, and Kazuko Nishikura. 2008. "Frequency and Fate of microRNA Editing in Human Brain." *Nucleic Acids Research* 36 (16): 5270–80. doi:10.1093/nar/gkn479.

Kawahara, Yukio, Boris Zinshteyn, Thimmaiah P Chendrimada, Ramin Shiekhattar, and Kazuko Nishikura. 2007. "RNA Editing of the microRNA-151 Precursor Blocks Cleavage by the Dicer–TRBP Complex." *EMBO Reports* 8 (8): 763–69. doi:10.1038/sj.embor.7401011.

Kawai, Shinji, and Atsuo Amano. 2012. "BRCA1 Regulates microRNA Biogenesis via the DROSHA Microprocessor Complex." *Journal of Cell Biology* 197 (2): 201–8. doi:10.1083/jcb.201110008.

Kawauchi, Junya, Hannah Mischo, Priscilla Braglia, Ana Rondon, and Nick J. Proudfoot. 2008. "Budding Yeast RNA Polymerases I and II Employ Parallel Mechanisms of Transcriptional Termination." *Genes and Development* 22 (8): 1082–92. doi:10.1101/gad.463408.

Kent, Tatiana, Gurushankar Chandramouly, Shane Michael McDevitt, Ahmet Y Ozdemir, and Richard T Pomerantz. 2015. “Mechanism of Microhomology-Mediated End-Joining Promoted by Human DNA Polymerase θ.” *Nature Structural & Molecular Biology* 22 (3): 230–37. doi:10.1038/nsmb.2961.

Keskin, Havva, Ying Shen, Fei Huang, Mikir Patel, Taehwan Yang, Katie Ashley, Alexander V. Mazin, and Francesca Storici. 2014. “Transcript-RNA-Templated DNA Recombination and Repair.” *Nature* 515 (7527): 436–39. doi:10.1038/nature13682.

Kim, Daehwan, Ben Langmead, and Steven L Salzberg. 2015. “HISAT: A Fast Spliced Aligner with Low Memory Requirements.” *Nature Methods* 12 (4): 357–60. doi:10.1038/nmeth.3317.

Kim, Joseph L, Dimitar B Nikolov, and Stephen K Burley. 1993. “Co-Crystal Structure of TBP Recognizing the Minor Groove of a TATA Element.” *Nature* 365 (6446): 520–27. http://dx.doi.org/10.1038/365520a0.

Kim, Nayun, and Sue Jinks-Robertson. 2012. “Transcription as a Source of Genome Instability.” *Nature Reviews Genetics* 13 (3): 204–14. doi:10.1038/nrg3152.

Kim, Tae-Kyung, Richard H Ebright, and Danny Reinberg. 2000. “Mechanism of ATP-Dependent Promoter Melting by Transcription Factor IIH.” *Science* 288 (5470): 1418 LP-1421. http://science.sciencemag.org/content/288/5470/1418.abstract.

Kim, Young-Kook, Boseon Kim, and V. Narry Kim. 2016. “Re-Evaluation of the Roles of *DROSHA* , *Exportin 5* , and *DICER* in microRNA Biogenesis.” *Proceedings of the National Academy of Sciences* 113 (13): E1881–89. doi:10.1073/pnas.1602532113.

Kim, Youngchang, James. H Geiger, Steven Hahn, and Paul B Sigler. 1993. “Crystal Structure of a Yeast TBP/TATA-Box Complex.” *Nature* 365 (6446): 512–20. http://dx.doi.org/10.1038/365512a0.

Knuckles, Philip, Miriam A Vogt, Sebastian Lugert, Marta Milo, Mark M W Chong, Guillaume M Hautbergue, Stuart A Wilson, Dan R Littman, and Verdon Taylor. 2012. “Drosha Regulates Neurogenesis by Controlling Neurogenin 2 Expression Independent of microRNAs.” *Nature Neuroscience* 15 (7): 962–69. doi:10.1038/nn.3139.

Kozomara, Ana, and Sam Griffiths-Jones. 2014. “MiRBase: Annotating High Confidence

microRNAs Using Deep Sequencing Data." *Nucleic Acids Research* 42 (D1): D68–73. doi:10.1093/nar/gkt1181.

Krol, Jacek, Krzysztof Sobczak, Urszula Wilczynska, Maria Drath, Anna Jasinska, Danuta Kaczynska, and Wlodzimierz J. Krzyzosiak. 2004. "Structural Features of microRNA (miRNA) Precursors and Their Relevance to miRNA Biogenesis and Small Interfering RNA/short Hairpin RNA Design." *Journal of Biological Chemistry* 279 (40): 42230–39. doi:10.1074/jbc.M404931200.

Kruhlak, Michael, Elizabeth E. Crouch, Marika Orlov, Carolina Montaño, Stanislaw A. Gorski, André Nussenzweig, Tom Misteli, Robert D. Phair, and Rafael Casellas. 2007. "The ATM Repair Pathway Inhibits RNA Polymerase I Transcription in Response to Chromosome Breaks." *Nature* 447 (7145): 730–34. doi:10.1038/nature05842.

Kuhn, Claus-D., Sebastian R. Geiger, Sonja Baumli, Marco Gartmann, Jochen Gerber, Stefan Jennebach, Thorsten Mielke, Herbert Tschochner, Roland Beckmann, and Patrick Cramer. 2007. "Functional Architecture of RNA Polymerase I." *Cell* 131 (7): 1260–72. doi:10.1016/j.cell.2007.10.051.

Kunkel, Thomas A. 2004. "DNA Replication Fidelity." *Journal of Biological Chemistry* 279 (17): 16895–98. doi:10.1074/jbc.R400006200.

Kuzminov, A. 2001. "Single-Strand Interruptions in Replicating Chromosomes Cause Double-Strand Breaks." *Proceedings of the National Academy of Sciences* 98 (15): 8241–46. doi:10.1073/pnas.131009198.

Kwon, S. Chul, Tuan Anh Nguyen, Yeon Gil Choi, Myung Hyun Jo, Sungchul Hohng, V Narry Kim, and Jae Sung Woo. 2016. "Structure of Human DROSHA." *Cell* 164 (1–2): 81–90. doi:10.1016/j.cell.2015.12.019.

Lagrange, T, A N Kapanidis, H Tang, D Reinberg, and R H Ebright. 1998. "New Core Promoter Element in RNA Polymerase II-Dependent Transcription: Sequence Specific DNA Binding by Transcription Factor IIB." *Genes and Development* 12 (1): 34–44. papers3://publication/uuid/D6AAF1D4-B67A-403C-BCDA-DCD38B168159.

Langmead, Ben, and Steven L Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods* 9 (4): 357–59. doi:10.1038/nmeth.1923.

Lee, Heng-Chi, Shwu-Shin Chang, Swati Choudhary, Antti P Aalto, Mekhala Maiti, Dennis H Bamford, and Yi Liu. 2009. "qiRNA Is a New Type of Small Interfering

RNA Induced by DNA Damage." *Nature* 459 (7244): 274–77. doi:10.1038/nature08041.

Lehman, Jason A., Derek J. Hoelz, and John J. Turchi. 2008. "DNA-Dependent Conformational Changes in the Ku Heterodimer." *Biochemistry* 47 (15): 4359–68. doi:10.1021/bi702284c.

Lehmann, Elisabeth, Florian Brueckner, and Patrick Cramer. 2007. "Molecular Basis of RNA-Dependent RNA Polymerase II Activity." *Nature* 450 (7168): 445–49. doi:10.1038/nature06290.

Leung, Yuk Yee, Pavel P. Kuksa, Alexandre Amlie-Wolf, Otto Valladares, Lyle H. Ungar, Sampath Kannan, Brian D. Gregory, and Li San Wang. 2016. "DASHR: Database of Small Human Noncoding RNAs." *Nucleic Acids Research* 44 (D1): D216–22. doi:10.1093/nar/gkv1188.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25 (16): 2078–79. doi:10.1093/bioinformatics/btp352.

Liang, Li, Li Deng, Son C. Nguyen, Xin Zhao, Christopher D. Maulion, Changshun Shao, and Jay A. Tischfield. 2008. "Human DNA Ligases I and III, but Not Ligase IV, Are Required for Microhomology-Mediated End Joining of DNA Double-Strand Breaks." *Nucleic Acids Research* 36 (10): 3297–3310. doi:10.1093/nar/gkn184.

Lipardi, Concetta, and Bruce M Paterson. 2009. "Identification of an RNA-Dependent RNA Polymerase in Drosophila Involved in RNAi and Transposon Suppression." *Proceedings of the National Academy of Sciences of the United States of America* 106 (37): 15645–50. doi:10.1073/pnas.0904984106.

Liu, Jie, Tammy Doty, Bryan Gibson, and Wolf-Dietrich Heyer. 2010. "Human BRCA2 Protein Promotes RAD51 Filament Formation on RPA-Covered Single-Stranded DNA." *Nature Structural & Molecular Biology* 17 (10): 1260–62. doi:10.1038/nsmb.1904.

Liu, Ying, Xuecheng Ye, Feng Jiang, Chunyang Liang, Dongmei Chen, Junmin Peng, Lisa N Kinch, Nick V Grishin, and Qinghua Liu. 2009. "C3PO, an Endoribonuclease That Promotes RNAi by Facilitating RISC Activation." *Science* 325 (5941): 750–53. doi:10.1126/science.1176325.

Llorente, Bertrand, and Lorraine S Symington. 2004. “The Mre11 Nuclease Is Not Required for 5’ to 3’ Resection at Multiple HO-Induced Double-Strand Breaks.” *Molecular and Cellular Biology* 24 (21): 9682–94. doi:10.1128/MCB.24.21.9682-9694.2004.

Löbrich, Markus, and Penny Jeggo. 2017. “A Process of Resection-Dependent Nonhomologous End Joining Involving the Goddess Artemis.” *Trends in Biochemical Sciences* 42 (9): 690–701. doi:10.1016/j.tibs.2017.06.011.

London, Robert E. 2015. “The Structural Basis of XRCC1-Mediated DNA Repair.” *DNA Repair* 30 (June): 90–103. doi:https://doi.org/10.1016/j.dnarep.2015.02.005.

Lou, Zhenkun, Katherine Minter-Dykhouse, Sonia Franco, Monica Gostissa, Melissa A. Rivera, Arkady Celeste, John P. Manis, et al. 2006. “MDC1 Maintains Genomic Stability by Participating in the Amplification of ATM-Dependent DNA Damage Signals.” *Molecular Cell* 21 (2): 187–200. doi:10.1016/j.molcel.2005.11.025.

Louder, Robert K, Yuan He, José Ramón López-Blanco, Jie Fang, Pablo Chacón, and Eva Nogales. 2016. “Structure of Promoter-Bound TFIID and Model of Human Pre-Initiation Complex Assembly.” *Nature* 531 (7596): 604–9. http://dx.doi.org/10.1038/nature17394.

Love, Michael I, Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12): 550. doi:10.1186/s13059-014-0550-8.

Ma, Yunmei, Haihui Lu, Brigette Tippin, Myron F. Goodman, Noriko Shimazaki, Osamu Koiwai, Chih Lin Hsieh, Klaus Schwarz, and Michael R. Lieber. 2004. “A Biochemically Defined System for Mammalian Nonhomologous DNA End Joining.” *Molecular Cell* 16 (5): 701–13. doi:10.1016/j.molcel.2004.11.017.

MacRae, I. J. 2006. “Structural Basis for Double-Stranded RNA Processing by Dicer.” *Science* 311 (5758): 195–98. doi:10.1126/science.1121638.

Mahaney, Brandi L., Katheryn Meek, and Susan P. Lees-Miller. 2009. “Repair of Ionizing Radiation-Induced DNA Double-Strand Breaks by Non-Homologous End-Joining.” *Biochemical Journal* 417 (3): 639–50. doi:10.1042/BJ20080413.

Manfrini, Nicola, Michela Clerici, Maxime Wery, Chiara Vittoria Colombo, Marc Descrimes, Antonin Morillon, Fabrizio d’Adda di Fagagna, and Maria Pia Longhese. 2015. “Resection Is Responsible for Loss of Transcription around a Double-Strand

Break in Saccharomyces Cerevisiae." Edited by Jessica K Tyler. *eLife* 4: e08942. doi:10.7554/eLife.08942.

Mari, P.-O., B. I. Florea, S. P. Persengiev, N. S. Verkaik, H. T. Bruggenwirth, M. Modesti, G. Giglia-Mari, et al. 2006. "Dynamic Assembly of End-Joining Complexes Requires Interaction between Ku70/80 and XRCC4." *Proceedings of the National Academy of Sciences* 103 (49): 18597–602. doi:10.1073/pnas.0609061103.

Maroney, Patricia A, Yang Yu, Jesse Fisher, and Timothy W Nilsen. 2006. "Evidence That microRNAs Are Associated with Translating Messenger RNAs in Human Cells." *Nature Structural & Molecular Biology* 13 (12): 1102–7. doi:10.1038/nsmb1174.

Martin, Marcel. 2011. "Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads." *EMBnet.journal* 17 (1): 10. doi:10.14806/ej.17.1.200.

Marzi, Matteo J., Francesco Ghini, Benedetta Cerruti, Stefano De Pretis, Paola Bonetti, Chiara Giacomelli, Marcin M. Gorski, et al. 2016. "Degradation Dynamics of Microrna Revealed by a Novel Pulse-Chase Approach." *Genome Research* 26 (4): 554–65. doi:10.1101/gr.198788.115.

Mason, S W, E E Sander, and I Grummt. 1997. "Identification of a Transcript Release Activity Acting on Ternary Transcription Complexes Containing Murine RNA Polymerase I." *The EMBO Journal* 16 (1): 163–72. doi:10.1093/emboj/16.1.163.

Matera, A. Gregory, and Zefeng Wang. 2014. "A Day in the Life of the Spliceosome." *Nature Reviews Molecular Cell Biology* 15 (2): 108–21. doi:10.1038/nrm3742.

Mathonnet, G., M. R. Fabian, Y. V. Svitkin, A. Parsyan, L. Huck, T. Murata, S. Biffo, et al. 2007. "MicroRNA Inhibition of Translation Initiation in Vitro by Targeting the Cap-Binding Complex eIF4F." *Science* 317 (5845): 1764–67. doi:10.1126/science.1146067.

Matsuoka, S., B. A. Ballif, A. Smogorzewska, E. R. McDonald, K. E. Hurov, J. Luo, C. E. Bakalarski, et al. 2007. "ATM and ATR Substrate Analysis Reveals Extensive Protein Networks Responsive to DNA Damage." *Science* 316 (5828): 1160–66. doi:10.1126/science.1140321.

Matsuzaki, Hajime, George A Kassavetis, and E.Peter Geiduschek. 1994. "Analysis of RNA Chain Elongation and Termination by Saccharomyces Cerevisiae RNA Polymerase III." *Journal of Molecular Biology* 235 (4): 1173–92.

doi:https://doi.org/10.1006/jmbi.1994.1072.

Mattiroli, Francesca, Joseph H a Vissers, Willem J van Dijk, Pauline Ikpa, Elisabetta Citterio, Wim Vermeulen, Jurgen a Marteijn, and Titia K Sixma. 2012. “RNF168 Ubiquitinates K13-15 on H2A/H2AX to Drive DNA Damage Signaling.” *Cell* 150 (6): 1182–95. doi:10.1016/j.cell.2012.08.005.

Maxon, Mary E., James A. Goodrich, and Robert Tjian. 1994. “Transcription Factor IIE Binds Preferentially to RNA Polymerase IIa and Recruits TFIIH: A Model for Promoter Clearance.” *Genes and Development* 8 (5): 515–24. doi:10.1101/gad.8.5.515.

Mazina, Olga M., Havva Keskin, Kritika Hanamshet, Francesca Storici, and Alexander V. Mazin. 2017. “Rad52 Inverse Strand Exchange Drives RNA-Templated DNA Double-Strand Break Repair.” *Molecular Cell* 67 (1): 19–29.e3. doi:10.1016/j.molcel.2017.05.019.

McElhinny, Stephanie A.Nick, Jody M. Havener, Miguel Garcia-Diaz, Raquel Juárez, Katarzyna Bebenek, Barbara L. Kee, Luis Blanco, Thomas A. Kunkel, and Dale A. Ramsden. 2005. “A Gradient of Template Dependence Defines Distinct Biological Roles for Family X Polymerases in Nonhomologous End Joining.” *Molecular Cell* 19 (3): 357–66. doi:10.1016/j.molcel.2005.06.012.

McVey, Mitch, Varandt Y. Khodaverdian, Damon Meyer, Paula Gonçalves Cerqueira, and Wolf-Dietrich Heyer. 2016. “Eukaryotic DNA Polymerases in Homologous Recombination.” *Annual Review of Genetics* 50 (1): 393–421. doi:10.1146/annurev-genet-120215-035243.

McVey, Mitch, and Sang Eun Lee. 2008. “MMEJ Repair of Double-Strand Breaks (Director’s Cut): Deleted Sequences and Alternative Endings.” *Trends in Genetics* 24 (11): 529–38. doi:10.1016/j.tig.2008.08.007.

Meek, David W. 2004. “The p53 Response to DNA Damage.” *DNA Repair* 3 (8–9): 1049–56. doi:https://doi.org/10.1016/j.dnarep.2004.03.027.

Mehta, Anuja, and James E. Haber. 2014. “Sources of DNA Double-Strand Breaks and Models of Rec.” *Cold Spring Harbor Perspectives in Biology* 6: 1–19. doi:10.1101/cshperspect.a016428.

Meijer, H. A., Y. W. Kong, W. T. Lu, A. Wilczynska, R. V. Spriggs, S. W. Robinson, J. D. Godfrey, A. E. Willis, and M. Bushell. 2013. “Translational Repression and eIF4A2

Activity Are Critical for MicroRNA-Mediated Gene Regulation." *Science* 340 (6128): 82–85. doi:10.1126/science.1231197.

Meister, Gunter, Markus Landthaler, Agnieszka Patkaniowska, Yair Dorsett, Grace Teng, and Thomas Tuschl. 2004. "Human Argonaute2 Mediates RNA Cleavage Targeted by miRNAs and siRNAs." *Molecular Cell* 15 (2): 185–97. doi:10.1016/j.molcel.2004.07.007.

Melé, Marta, and John L. Rinn. 2016. "'Cat's Cradling' the 3D Genome by the Act of LncRNA Transcription." *Molecular Cell* 62 (5): 657–64. doi:10.1016/j.molcel.2016.05.011.

Melo, Sonia A., Catia Moutinho, Santiago Ropero, George A. Calin, Simona Rossi, Riccardo Spizzo, Agustin F. Fernandez, et al. 2010. "A Genetic Defect in Exportin-5 Traps Precursor MicroRNAs in the Nucleus of Cancer Cells." *Cancer Cell* 18 (4): 303–15. doi:10.1016/j.ccr.2010.09.007.

Michalik, Katharina M., Romy Böttcher, and Klaus Förstemann. 2012. "A Small RNA Response at DNA Ends in Drosophila." *Nucleic Acids Research* 40 (19): 9596–9603. doi:10.1093/nar/gks711.

Miki, Daisuke, Peiying Zhu, Wencan Zhang, Yanfei Mao, Zhengyan Feng, Huan Huang, Hui Zhang, et al. 2017. "Efficient Generation of diRNAs Requires Components in the Posttranscriptional Gene Silencing Pathway." *Scientific Reports* 7 (1): 301. doi:10.1038/s41598-017-00374-7.

Mimitou, Eleni P, and Lorraine S Symington. 2008. "Sae2, Exo1 and Sgs1 Collaborate in DNA Double-Strand Break Processing." *Nature* 455 (7214): 770–74. doi:10.1038/nature07312.

Min, Jung-Hyun, and Nikola P Pavletich. 2007. "Recognition of DNA Damage by the Rad4 Nucleotide Excision Repair Protein." *Nature* 449 (7162): 570–75. http://dx.doi.org/10.1038/nature06155.

Mocquet, Vincent, Jean Philippe Lainé, Thilo Riedl, Zhou Yajin, Marietta Y Lee, and Jean Marc Egly. 2008. "Sequential Recruitment of the Repair Factors during NER: The Role of XPG in Initiating the Resynthesis Step." *The EMBO Journal* 27 (1): 155–67. doi:10.1038/sj.emboj.7601948.

Mondol, Vanessa, Byoung Chan Ahn, and Amy E. Pasquinelli. 2015. "Splicing Remodels the Let-7 Primary microRNA to Facilitate Drosha Processing in Caenorhabditis

Elegans.” *RNA* 21 (8): 1396–1403. doi:10.1261/rna.052118.115.

Moser, Jill, Marcel Volker, Hanneke Kool, Sergei Alekseev, Harry Vrieling, Akira Yasui, Albert A van Zeeland, and Leon H F Mullenders. 2005. “The UV-Damaged DNA Binding Protein Mediates Efficient Targeting of the Nucleotide Excision Repair Complex to UV-Induced Photo Lesions.” *DNA Repair* 4 (5): 571–82. doi:https://doi.org/10.1016/j.dnarep.2005.01.001.

Moshous, Despina, Isabelle Callebaut, Régina De Chasseval, Barbara Corneo, Marina Cavazzana-Calvo, Françoise Le Deist, Ilhan Tezcan, et al. 2001. “Artemis, a Novel DNA Double-Strand Break repair/V(D)J Recombination Protein, Is Mutated in Human Severe Combined Immune Deficiency.” *Cell* 105 (2): 177–86. doi:10.1016/S0092-8674(01)00309-9.

Motycka, Teresa A., Tadayoshi Bessho, Sean M. Post, Patrick Sung, and Alan E. Tomkinson. 2004. “Physical and Functional Interaction between the XPF/ERCC1 Endonuclease and hRad52.” *Journal of Biological Chemistry* 279 (14): 13634–39. doi:10.1074/jbc.M313779200.

Murr, Rabih, Joanna I. Loizou, Yun-Gui Yang, Cyrille Cuenin, Hai Li, Zhao-Qi Wang, and Zdenko Herceg. 2006. “Histone Acetylation by Trrap–Tip60 Modulates Loading of Repair Proteins and Repair of DNA Double-Strand Breaks.” *Nature Cell Biology* 8 (1): 91–99. doi:10.1038/ncb1343.

Nguyen, Tuan Anh, Myung Hyun Jo, Yeon-Gil Choi, Joha Park, S Chul Kwon, Sungchul Hohng, V Narry Kim, and Jae-Sung Woo. 2015. “Functional Anatomy of the Human Microprocessor.” *Cell* 161 (6): 1374–87. doi:10.1016/j.cell.2015.05.010.

Nielsen, Soren, Yulia Yuzenkova, and Nikolay Zenkin. 2013. “Mechanism of Eukaryotic RNA Polymerase III Transcription Termination.” *Science* 340 (6140): 1577 LP-1580. http://science.sciencemag.org/content/340/6140/1577.abstract.

Nielsen, Soren, and Nikolay Zenkin. 2014. “Response to Comment on ‘Mechanism of Eukaryotic RNA Polymerase III Transcription Termination.’” *Science* 345 (6196): 524 LP-524. http://science.sciencemag.org/content/345/6196/524.4.abstract.

Nojima, Takayuki, Tomás Gomes, Ana Rita Fialho Grosso, Hiroshi Kimura, Michael J. Dye, Somdutta Dhir, Maria Carmo-Fonseca, and Nicholas J. Proudfoot. 2015. “Mammalian NET-Seq Reveals Genome-Wide Nascent Transcription Coupled to RNA Processing.” *Cell* 161 (3): 526–40. doi:10.1016/j.cell.2015.03.027.

Noland, C. L., and J. A. Doudna. 2013. “Multiple Sensors Ensure Guide Strand Selection in Human RNAi Pathways.” *RNA* 19 (5): 639–48. doi:10.1261/rna.037424.112.

Ochs, Fena, Kumar Somyajit, Matthias Altmeyer, Maj-Britt Rask, Jiri Lukas, and Claudia Lukas. 2016. “53BP1 Fosters Fidelity of Homology-Directed DNA Repair.” *Nature Structural & Molecular Biology* 23 (8): 714–21. doi:10.1038/nsmb.3251.

Ogi, Tomoo, Siripan Limsirichaikul, René M. Overmeer, Marcel Volker, Katsuya Takenaka, Ross Cloney, Yuka Nakazawa, et al. 2010. “Three DNA Polymerases, Recruited by Different Mechanisms, Carry Out NER Repair Synthesis in Human Cells.” *Molecular Cell* 37 (5): 714–27. doi:10.1016/j.molcel.2010.02.009.

Ohkuma, Yoshiaki, and Robert G Roeder. 1994. “Regulation of TFIIH ATPase and Kinase Activities by TFIIE during Active Initiation Complex Formation.” *Nature* 368 (6467): 160–63. http://dx.doi.org/10.1038/368160a0.

Ohle, Corina, Rafael Tesorero, Géza Schermann, Nikolay Dobrev, Irmgard Sinning, and Tamás Fischer. 2016. “Transient RNA-DNA Hybrids Are Required for Efficient Double-Strand Break Repair.” *Cell* 167 (4): 1001–1013.e7. doi:10.1016/j.cell.2016.10.001.

Okamoto, Tomoko, Seiji Yamamoto, Yoshinori Watanabe, Tsutomu Ohta, Fumio Hanaoka, Robert G Roeder, and Yoshiaki Ohkuma. 1998. “Analysis of the Role of TFIIE in Transcriptional Regulation through Structure-Function Studies of the TFIIEβ Subunit.” *Journal of Biological Chemistry* 273 (31): 19866–76. doi:10.1074/jbc.273.31.19866.

Pak, J., and A. Fire. 2007. “Distinct Populations of Primary and Secondary Effectors During RNAi in C. Elegans.” *Science* 315 (5809): 241–44. doi:10.1126/science.1132839.

Palazzo, Alexander F., and Eliza S. Lee. 2015. “Non-Coding RNA: What Is Functional and What Is Junk?” *Frontiers in Genetics* 5 (JAN): 1–11. doi:10.3389/fgene.2015.00002.

Pan, Yaoqian, Louisa Balazs, Gabor Tigyi, and Junming Yue. 2011. “Conditional Deletion of Dicer in Vascular Smooth Muscle Cells Leads to the Developmental Delay and Embryonic Mortality.” *Biochemical and Biophysical Research Communications* 408 (3): 369–74. doi:10.1016/j.bbrc.2011.02.119.

Panier, Stephanie, and Simon J. Boulton. 2013. “Double-Strand Break Repair: 53BP1

Comes into Focus." *Nature Reviews Molecular Cell Biology* 15 (1): 7–18. doi:10.1038/nrm3719.

Pankotai, Tibor, Céline Bonhomme, David Chen, and Evi Soutoglou. 2012. "DNAPKcs-Dependent Arrest of RNA Polymerase II Transcription in the Presence of DNA Breaks." *Nature Structural & Molecular Biology* 19 (3): 276–82. doi:10.1038/nsmb.2224.

Park, Jong-Eun, Inha Heo, Yuan Tian, Dhirendra K Simanshu, Hyeshik Chang, David Jee, Dinshaw J Patel, and V Narry Kim. 2011. "Dicer Recognizes the 5' End of RNA for Efficient and Accurate Processing." *Nature* 475 (7355): 201–5. doi:10.1038/nature10198.

Parsons, Jason L, Bradley D Preston, Timothy R O'Connor, and Grigory L Dianov. 2007. "DNA Polymerase Delta-Dependent Repair of DNA Single Strand Breaks Containing 3'-end Proximal Lesions." *Nucleic Acids Research* 35 (4): 1054–63. doi:10.1093/nar/gkl1115.

Paull, TT. 2015. "Mechanisms of ATM Activation." *Annual Review of Biochemistry*, no. December: 1–28. doi:10.1146/annurev-biochem-060614-034335.

Paulsen, Renee D., Deena V. Soni, Roy Wollman, Angela T. Hahn, Muh Ching Yee, Anna Guan, Jayne A. Hesley, et al. 2009. "A Genome-Wide siRNA Screen Reveals Diverse Cellular Processes and Pathways That Mediate Genome Stability." *Molecular Cell* 35 (2): 228–39. doi:10.1016/j.molcel.2009.06.021.

Peña-Diaz, Javier, and Josef Jiricny. 2012. "Mammalian Mismatch Repair: Error-Free or Error-Prone?" *Trends in Biochemical Sciences* 37 (5): 206–14. doi:10.1016/j.tibs.2012.03.001.

Penny, Graeme D., Graham F. Kay, Steven A. Sheardown, Sohaila Rastan, and Neil Brockdorff. 1996. "Requirement for Xist in X Chromosome Inactivation." *Nature* 379 (6561): 131–37. doi:10.1038/379131a0.

Price, Morgan N., Eric J. Alm, and Adam P. Arkin. 2005. "Interruptions in Gene Expression Drive Highly Expressed Operons to the Leading Strand of DNA Replication." *Nucleic Acids Research* 33 (10): 3224–34. doi:10.1093/nar/gki638.

Pryde, F. 2005. "53BP1 Exchanges Slowly at the Sites of DNA Damage and Appears to Require RNA for Its Association with Chromatin." *Journal of Cell Science* 118 (9): 2043–55. doi:10.1242/jcs.02336.

Quinlan, Aaron R, and Ira M Hall. 2010. “BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features.” *Bioinformatics* 26 (6): 841–42. doi:10.1093/bioinformatics/btq033.

Rädle, Bernd, Andrzej J. Rutkowski, Zsolt Ruzsics, Caroline C. Friedel, Ulrich H. Koszinowski, and Lars Dölken. 2013. “Metabolic Labeling of Newly Transcribed RNA for High Resolution Gene Expression Profiling of RNA Synthesis, Processing and Decay in Cell Culture.” *Journal of Visualized Experiments*, no. 78(August): 50195. doi:10.3791/50195.

Reisz, Julie A., Nidhi Bansal, Jiang Qian, Weiling Zhao, and Cristina M. Furdui. 2014. “Effects of Ionizing Radiation on Biological Molecules—Mechanisms of Damage and Emerging Methods of Detection.” *Antioxidants & Redox Signaling* 21 (2). 140 Huguenot Street, 3rd FloorNew Rochelle, NY 10801USA: 260–92. doi:10.1089/ars.2013.5489.

Renkawitz, Jörg, Claudio A. Lademann, Marian Kalocsay, and Stefan Jentsch. 2013. “Monitoring Homology Search during DNA Double-Strand Break Repair In Vivo.” *Molecular Cell* 50 (2): 261–72. doi:10.1016/j.molcel.2013.02.020.

Revyakin, Andrey, Chenyu Liu, Richard H Ebright, and Terence R Strick. 2006. “Abortive Initiation and Productive Initiation by RNA Polymerase Involve DNA Scrunching.” *Science* 314 (5802): 1139 LP-1143. http://science.sciencemag.org/content/314/5802/1139.abstract.

Riddles, P. W., and I. R. Lehman. 1985. “The Formation of Paranemic and Plectonemic Joints between DNA Molecules by the recA and Single-Stranded DNA-Binding Proteins of Escherichia Coli.” *Journal of Biological Chemistry* 260 (1): 165–69. http://www.jbc.org/content/260/1/165.abstract.

Rossiello, Francesca, Julio Aguado, Sara Sepe, Fabio Iannelli, Quan Nguyen, Sethuramasundaram Pitchiaya, Piero Carninci, and Fabrizio d’Adda di Fagagna. 2017. “DNA Damage Response Inhibition at Dysfunctional Telomeres by Modulation of Telomeric DNA Damage Response RNAs.” *Nature Communications* 8 (May 2016): 13980. doi:10.1038/ncomms13980.

Roth, Braden M, Daniella Ishimaru, and Mirko Hennig. 2013. “The Core Microprocessor Component DiGeorge Syndrome Critical Region 8 (DGCR8) Is a Nonspecific RNA-Binding Protein.” *Journal of Biological Chemistry* 288 (37): 26785–99.

doi:10.1074/jbc.M112.446880.

Rothenberg, E., J. M. Grimme, M. Spies, and T. Ha. 2008. “Human Rad52-Mediated Homology Search and Annealing Occurs by Continuous Interactions between Overlapping Nucleoprotein Complexes.” *Proceedings of the National Academy of Sciences* 105 (51): 20274–79. doi:10.1073/pnas.0810317106.

Roy, Sunetra, Abinadabe J de Melo, Yao Xu, Satish K Tadi, Aurélie Négrel, Eric Hendrickson, Mauro Modesti, and Katheryn Meek. 2015. “XRCC4/XLF Interaction Is Variably Required for DNA Repair and Is Not Required for Ligase IV Stimulation.” *Molecular and Cellular Biology* 35 (17): 3017–28. doi:10.1128/MCB.01503-14.

Ryan, Kevin, Olga Calvo, and James L Manley. 2004. “Evidence That Polyadenylation Factor CPSF-73 Is the mRNA 3’ Processing Endonuclease.” *RNA* 10 (4): 565–73. doi:10.1261/rna.5214404.quence.

Sainsbury, Sarah, Jürgen Niesser, and Patrick Cramer. 2012. “Structure and Function of the Initially Transcribing RNA Polymerase II–TFIIB Complex.” *Nature* 493 (7432): 437–40. doi:10.1038/nature11715.

Salzman, D W, K Nakamura, S Nallur, M T Dookwah, C Metheetrairut, F J Slack, and J B Weidhaas. 2016. “miR-34 Activity Is Modulated through 5’-end Phosphorylation in Response to DNA Damage.” *Nature Communications* 7 (May 2015): 10954. doi:10.1038/ncomms10954.

Schwalb, Björn, Margaux Michel, Benedikt Zacher, Katja Frühauf, Carina Demel, Achim Tresch, Julien Gagneur, and Patrick Cramer. 2016. “TT-Seq Maps the Human Transient Transcriptome.” *Science* 352 (6290): 1225–28. doi:10.1126/science.aad9841.

Schwarz, Dianne S., György Hutvágner, Tingting Du, Zuoshang Xu, Neil Aronin, and Phillip D. Zamore. 2003. “Asymmetry in the Assembly of the RNAi Enzyme Complex.” *Cell* 115 (2): 199–208. doi:10.1016/S0092-8674(03)00759-1.

Schwertman, Petra, Anna Lagarou, Dick H W Dekkers, Anja Raams, Adriana C van der Hoek, Charlie Laffeber, Jan H J Hoeijmakers, et al. 2012. “UV-Sensitive Syndrome Protein UVSSA Recruits USP7 to Regulate Transcription-Coupled Repair.” *Nature Genetics* 44 (5): 598–602. doi:10.1038/ng.2230.

Seila, A. C., J. M. Calabrese, S. S. Levine, G. W. Yeo, P. B. Rahl, R. A. Flynn, R. A. Young,

and P. A. Sharp. 2008. “Divergent Transcription from Active Promoters.” *Science* 322 (5909): 1849–51. doi:10.1126/science.1162253.

Seluanov, Andrei, Zhiyong Mao, and Vera Gorbunova. 2010. “Analysis of DNA Double-Strand Break (DSB) Repair in Mammalian Cells.” *Journal of Visualized Experiments : JoVE*, no. 43(January): 3–9. doi:10.3791/2002.

Shanbhag, Niraj M, Ilona U Rafalska-Metcalf, Carlo Balane-Bolivar, Susan M Janicki, and Roger a Greenberg. 2010. “ATM-Dependent Chromatin Changes Silence Transcription in Cis to DNA Double-Strand Breaks.” *Cell* 141 (6): 970–81. doi:10.1016/j.cell.2010.04.038.

Siaud, Nicolas, Maria A Barbera, Akinori Egashira, Isabel Lam, Nicole Christ, Katharina Schlacher, Bing Xia, and Maria Jasin. 2011. “Plasticity of BRCA2 Function in Homologous Recombination: Genetic Interactions of the PALB2 and DNA Binding Domains.” *PLOS Genetics* 7 (12): e1002409. https://doi.org/10.1371/journal.pgen.1002409.

Sijen, T., F. A. Steiner, K. L. Thijssen, and R. H. A. Plasterk. 2007. “Secondary siRNAs Result from Unprimed RNA Synthesis and Form a Distinct Class.” *Science* 315 (5809): 244–47. doi:10.1126/science.1136699.

Skourti-Stathaki, Konstantina, and Nicholas J. Proudfoot. 2014. “A Double-Edged Sword: R Loops as Threats to Genome Integrity and Powerful Regulators of Gene Expression.” *Genes and Development* 28 (13): 1384–96. doi:10.1101/gad.242990.114.

Skourti-Stathaki, Konstantina, Nicholas J. Proudfoot, and Natalia Gromak. 2011. “Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination.” *Molecular Cell* 42 (6): 794–805. doi:10.1016/j.molcel.2011.04.026.

Slabicki, Mikolaj, Mirko Theis, Dragomir B. Krastev, Sergey Samsonov, Emeline Mundwiller, Magno Junqueira, Maciej Paszkowski-Rogacz, et al. 2010. “A Genome-Scale DNA Repair RNAi Screen Identifies SPG48 as a Novel Gene Associated with Hereditary Spastic Paraplegia.” *PLoS Biology* 8 (6). doi:10.1371/journal.pbio.1000408.

Smid, Amke, Martin Finsterer, and Ingrid Grummt. 1992. “Limited Proteolysis Unmasks Specific DNA-Binding of the Murine RNA Polymerase I-Specific Transcription

Termination Factor TTFI." *Journal of Molecular Biology* 227 (3): 635–47. doi:http://dx.doi.org/10.1016/0022-2836(92)90213-4.

Solovjeva, Liudmila V, Maria P Svetlova, Vadim O Chagin, and Nikolai V Tomilin. 2007. "Inhibition of Transcription at Radiation-Induced Nuclear Foci of Phosphorylated Histone H2AX in Mammalian Cells." *Chromosome Research* 15 (6): 787–97. doi:10.1007/s10577-007-1162-x.

Staresincic, Lidija, Adebanke F Fagbemi, Jacqueline H Enzlin, Audrey M Gourdin, Nils Wijgers, Isabelle Dunand-Sauthier, Giuseppina Giglia-Mari, Stuart G Clarkson, Wim Vermeulen, and Orlando D Schärer. 2009. "Coordination of Dual Incision and Repair Synthesis in Human Nucleotide Excision Repair." *The EMBO Journal* 28 (8): 1111–20. doi:10.1038/emboj.2009.49.

Stein, Hans, and Peter Hausen. 1969. "Enzyme from Calf Thymus Degrading the RNA Moiety of DNA-RNA Hybrids: Effect on DNA-Dependent RNA Polymerase." *Science* 166 (3903): 393 LP-395. http://science.sciencemag.org/content/166/3903/393.abstract.

Stewart, Grant S, Stephanie Panier, Kelly Townsend, Abdallah K Al-Hakim, Nadine K Kolas, Edward S Miller, Shinichiro Nakada, et al. 2009. "The RIDDLE Syndrome Protein Mediates a Ubiquitin-Dependent Signaling Cascade at Sites of DNA Damage." *Cell* 136 (3): 420–34. doi:10.1016/j.cell.2008.12.042.

Storici, Francesca, Katarzyna Bebenek, Thomas A. Kunkel, Dmitry A. Gordenin, and Michael A. Resnick. 2007. "RNA-Templated DNA Repair." *Nature* 447 (7142): 338–41. doi:10.1038/nature05720.

Su, Hong, Melanie I. Trombly, Jian Chen, and Xiaozhong Wang. 2009. "Essential and Overlapping Functions for Mammalian Argonautes in microRNA Silencing." *Genes and Development* 23 (3): 304–17. doi:10.1101/gad.1749809.

Subramanian, Aravind, Pablo Tamayo, and Vamsi Mootha. 2014. "GSEA : Gene Set Enrichment Analysis Gene Set Enrichment Analysis : A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles." *Proceedings of the National Academy of Sciences of the United States of America* 102 (43): 15545–50. doi:10.1073/pnas.0506580102.

Sung Rhee, Ho, and B. Franklin Pugh. 2012. "Genome-Wide Structure and Organization of Eukaryotic Pre-Initiation Complexes." *Nature* 487 (7405): 128–128.

doi:10.1038/nature11266.

Suwa, a, M Hirakata, Y Takeda, S a Jesch, T Mimori, and J a Hardin. 1994. “DNA-Dependent Protein Kinase (Ku Protein-p350 Complex) Assembles on Double-Stranded DNA.” *Proceedings of the National Academy of Sciences* 91 (July): 6904–8. doi:10.1073/pnas.91.15.6904.

Sy, S. M. H., M. S. Y. Huen, and J. Chen. 2009. “PALB2 Is an Integral Component of the BRCA Complex Required for Homologous Recombination Repair.” *Proceedings of the National Academy of Sciences* 106 (17): 7155–60. doi:10.1073/pnas.0811159106.

Taft, Ryan J, Evgeny A Glazov, Nicole Cloonan, Cas Simons, Stuart Stephen, Geoffrey J Faulkner, Timo Lassmann, et al. 2009. “Tiny RNAs Associated with Transcription Start Sites in Animals.” *Nature Genetics* 41 (5): 572–78. doi:10.1038/ng.312.

Tang, Jiangbo, Nam Woo Cho, Gaofeng Cui, Erica M Manion, Niraj M Shanbhag, Maria Victoria Botuyan, Georges Mer, and Roger A Greenberg. 2013. “Acetylation Limits 53BP1 Association with Damaged Chromatin to Promote Homologous Recombination.” *Nature Structural & Molecular Biology* 20 (3): 317–25. doi:10.1038/nsmb.2499.

Thomas, Marshall P., Xing Liu, Jennifer Whangbo, Geoffrey McCrossan, Keri B. Sanborn, Emre Basar, Michael Walch, and Judy Lieberman. 2015. “Apoptosis Triggers Specific, Rapid, and Global mRNA Decay with 3’ Uridylated Intermediates Degraded by DIS3L2.” *Cell Reports* 11 (7): 1079–89. doi:10.1016/j.celrep.2015.04.026.

Thompson, Larry H. 2012. “Recognition, Signaling, and Repair of DNA Double-Strand Breaks Produced by Ionizing Radiation in Mammalian Cells: The Molecular Choreography.” *Mutation Research - Reviews in Mutation Research*. doi:10.1016/j.mrrev.2012.06.002.

Trojer, Patrick, and Danny Reinberg. 2007. “Facultative Heterochromatin: Is There a Distinctive Molecular Signature?” *Molecular Cell* 28 (1): 1–13. doi:10.1016/j.molcel.2007.09.011.

Trujillo, K M, S S Yuan, E Y Lee, and P Sung. 1998. “Nuclease Activities in a Complex of Himan Recombination and DNA Repair Factors Rad50,Mre11,andp95_JBC_1998.pdf.” *Journal of Biological Chemistry* 273 (34):

21447–50. doi:10.1074/jbc.273.34.21447.

Truong, Lan N, Yongjiang Li, Linda Z Shi, Patty Yi-Hwa Hwang, Jing He, Hailong Wang, Niema Razavian, Michael W Berns, and Xiaohua Wu. 2013. “Microhomology-Mediated End Joining and Homologous Recombination Share the Initial End Resection Step to Repair DNA Double-Strand Breaks in Mammalian Cells.” *Proceedings of the National Academy of Sciences* 110 (19): 7720–25. doi:10.1073/pnas.1213431110.

Tuzon, Creighton T., Tanya Spektor, Xiaodong Kong, Lauren M. Congdon, Shumin Wu, Gunnar Schotta, Kyoko Yokomori, and Judd C. Rice. 2014. “Concerted Activities of Distinct H4K20 Methyltransferases at DNA Double-Strand Breaks Regulate 53BP1 Nucleation and NHEJ-Directed Repair.” *Cell Reports* 8 (2): 430–38. doi:10.1016/j.celrep.2014.06.013.

Uhlmann, F, and K Nasmyth. 1998. “Cohesion between Sister Chromatids Must Be Established during DNA Replication.” *Current Biology* 8 (20): 1095–1101. doi:10.1016/S0960-9822(98)70463-4.

Vaňková Hausnerová, Viola, and Christian Lanctôt. 2017. “Chromatin Decondensation Is Accompanied by a Transient Increase in Transcriptional Output.” *Biology of the Cell* 109 (1): 65–79. doi:10.1111/boc.201600032.

Vlachos, Ioannis S., Maria D. Paraskevopoulou, Dimitra Karagkouni, Georgios Georgakilas, Thanasis Vergoulis, Ilias Kanellos, Ioannis Laertis Anastasopoulos, et al. 2015. “DIANA-TarBase v7.0: Indexing More than Half a Million Experimentally Supported miRNA:mRNA Interactions.” *Nucleic Acids Research* 43 (D1): D153–59. doi:10.1093/nar/gku1215.

Vlachos, Ioannis S., Konstantinos Zagganas, Maria D. Paraskevopoulou, Georgios Georgakilas, Dimitra Karagkouni, Thanasis Vergoulis, Theodore Dalamagas, and Artemis G. Hatzigeorgiou. 2015. “DIANA-miRPath v3.0: Deciphering microRNA Function with Experimental Support.” *Nucleic Acids Research* 43 (W1): W460–66. doi:10.1093/nar/gkv403.

Wagner, Stacey D, Petro Yakovchuk, Benjamin Gilman, Steven L Ponicsan, Linda F Drullinger, Jennifer F Kugel, and James A Goodrich. 2013. “RNA Polymerase II Acts as an RNA-Dependent RNA Polymerase to Extend and Destabilize a Non-Coding RNA.” *The EMBO Journal* 32 VN-r (6): 781–90. doi:10.1038/emboj.2013.18.

Walker, J R, R A Corpina, and J Goldberg. 2001. “Structure of the Ku Heterodimer Bound to DNA and Its Implications for Double-Strand Break Repair.” *Nature* 412 (6847): 607–14. doi:10.1038/35088000.

Wang, Hailong, Linda Z. Shi, Catherine C L Wong, Xuemei Han, Patty Yi Hwa Hwang, Lan N. Truong, Qingyuan Zhu, et al. 2013. “The Interaction of CtIP and Nbs1 Connects CDK and ATM to Regulate HR-Mediated Double-Strand Break Repair.” *PLoS Genetics* 9 (2): e1003277. doi:10.1371/journal.pgen.1003277.

Wang, Hong-Wei, Cameron Noland, Bunpote Siridechadilok, David W Taylor, Enbo Ma, Karin Felderer, Jennifer A Doudna, and Eva Nogales. 2009. “Structural Insights into RNA Processing by the Human RISC-Loading Complex.” *Nature Structural & Molecular Biology* 16 (11): 1148–53. doi:10.1038/nsmb.1673.

Wang, Minli, Weizhong Wu, Wenqi Wu, Bustanur Rosidi, Lihua Zhang, Huichen Wang, and George Iliakis. 2006. “PARP-1 and Ku Compete for Repair of DNA Double Strand Breaks by Distinct NHEJ Pathways.” *Nucleic Acids Research* 34 (21): 6170–82. doi:10.1093/nar/gkl840.

Watts, Felicity Z. 2016. “Repair of DNA Double-Strand Breaks in Heterochromatin.” Edited by Rob de Bruin. *Biomolecules* 6 (4): 47. doi:10.3390/biom6040047.

Weber, ChristopherM, Srinivas Ramachandran, and Steven Henikoff. 2014. “Nucleosomes Are Context-Specific, H2A.Z-Modulated Barriers to RNA Polymerase.” *Molecular Cell* 53 (5): 819–30. doi:10.1016/j.molcel.2014.02.014.

Wei, Leizhen, Satoshi Nakajima, Stefanie Böhm, Kara A. Bernstein, Zhiyuan Shen, Michael Tsang, Arthur S. Levine, and Li Lan. 2015. “DNA Damage during the G0/G1 Phase Triggers RNA-Templated, Cockayne Syndrome B-Dependent Homologous Recombination.” *Proceedings of the National Academy of Sciences* 112 (27): E3495–3504. doi:10.1073/pnas.1507105112.

Wei, Wei, Zhaoqing Ba, Min Gao, Yang Wu, Yanting Ma, Simon Amiard, Charles I White, Jannie Michaela Rendtlew Danielsen, Yun-Gui Yang, and Yijun Qi. 2012. “A Role for Small RNAs in DNA Double-Strand Break Repair.” *Cell* 149 (1): 101–12. doi:10.1016/j.cell.2012.03.002.

White, Eleanor, Margarita Schlackow, Kinga Kamieniarz-Gdula, Nick J Proudfoot, and Monika Gullerova. 2014. “Human Nuclear Dicer Restricts the Deleterious Accumulation of Endogenous Double-Stranded RNA.” *Nature Structural &*

Molecular Biology 21 (6): 552–59. doi:10.1038/nsmb.2827.

White, Ryan R., Patricia Sung, C. Greer Vestal, Gregory Benedetto, Noelle Cornelio, and Christine Richardson. 2013. "Double-Strand Break Repair by Interchromosomal Recombination: An in Vivo Repair Mechanism Utilized by Multiple Somatic Tissues in Mammals." *PLoS ONE* 8 (12): 1–16. doi:10.1371/journal.pone.0084379.

Wickham, Hadley. 2009. *ggplot2: Elegant Graphics for Data Analysis*. *Springer Publishing Company, Incorporated*. 2nd ed. Vol. 35. doi:10.1007/978-0-387-98141-3.

Wilczynska, Ania, and Martin Bushell. 2015. "The Complexity of miRNA-Mediated Repression." *Cell Death and Differentiation* 22 (1): 22–33. doi:10.1038/cdd.2014.112.

Windhager, Lukas, Thomas Bonfert, Kaspar Burger, Zsolt Ruzsics, Stefan Krebs, Stefanie Kaufmann, Georg Malterer, et al. 2012. "Ultrashort and Progressive 4sU-Tagging Reveals Key Characteristics of RNA Processing at Nucleotide Resolution." *Genome Research* 22 (10): 2031–42. doi:10.1101/gr.131847.111.

Wittschieben, Birgitte Ø, Shigenori Iwai, and Richard D Wood. 2005. "DDB1-DDB2 (Xeroderma Pigmentosum Group E) Protein Complex Recognizes a Cyclobutane Pyrimidine Dimer, Mismatches, Apurinic/Apyrimidinic Sites, and Compound Lesions in DNA." *Journal of Biological Chemistry* 280 (48): 39982–89. doi:10.1074/jbc.M507854200.

Wold, Marc S. 1997. "Replication Protein A: A Heterotrimeric, Single-Stranded DNA-Binding Protein Required for Eukaryotic DNA Metabolism." *Annual Review of Biochemistry* 66 (1): 61–92. doi:10.1146/annurev.biochem.66.1.61.

Wong, Gane Ka Shu, Douglas A. Passey, and Jun Yu. 2001. "Most of the Human Genome Is Transcribed." *Genome Research* 11 (12): 1975–77. doi:10.1101/gr.202401.

Xiang, Yang, Benoit Laurent, Chih-Hung Hsu, Sigrid Nachtergaele, Zhike Lu, Wanqiang Sheng, Chuanyun Xu, et al. 2017. "RNA m6A Methylation Regulates the Ultraviolet-Induced DNA Damage Response." *Nature* 543 (7646): 573–76. doi:10.1038/nature21671.

Xie, Mingyi, Mingfeng Li, Anna Vilborg, Nara Lee, Mei Di Shu, Valeria Yartseva, Nenad Šestan, and Joan A. Steitz. 2013. "Mammalian 5'-Capped microRNA Precursors

That Generate a Single microRNA." *Cell* 155 (7): 1568–80. doi:10.1016/j.cell.2013.11.027.

Xiong, Xiahui, Zhanwen Du, Ying Wang, Zhihui Feng, Pan Fan, Chunhong Yan, Henning Willers, and Junran Zhang. 2015. "53BP1 Promotes Microhomology-Mediated End-Joining in G1-Phase Cells." *Nucleic Acids Research* 43 (3): 1659–70. doi:10.1093/nar/gku1406.

Xu, Ping, Martina Billmeier, Irina Mohorianu, Darrell Green, William D Fraser, and Tamas Dalmay. 2015. "An Improved Protocol for Small RNA Library Construction Using High Definition Adapters." *Methods in Next Generation Sequencing* 2 (1): 1–10. doi:10.1515/mngs-2015-0001.

Yamaguchi, Hiroki, Rodrigo T. Calado, Hinh Ly, Sachiko Kajigaya, Gabriela M. Baerlocher, Stephen J. Chanock, Peter M. Lansdorp, and Neal S. Young. 2005. "Mutations in TERT, the Gene for Telomerase Reverse Transcriptase, in Aplastic Anemia." *The New England Journal of Medicine* 352 (14): 1413–24. doi:10.1056/NEJMoa042980.

Yang, Chuhu, Eugene Bolotin, Tao Jiang, Frances M. Sladek, and Ernest Martinez. 2007. "Prevalence of the Initiator over the TATA Box in Human and Yeast Genes and Identification of DNA Motifs Enriched in Human TATA-Less Core Promoters." *Gene* 389 (1): 52–65. doi:10.1016/j.gene.2006.09.029.

Yang, Haijuan, Philip D Jeffrey, Julie Miller, Elspeth Kinnucan, Yutong Sun, Nicolas H Thomä, Ning Zheng, Phang-Lang Chen, Wen-Hwa Lee, and Nikola P Pavletich. 2002. "BRCA2 Function in DNA Binding and Recombination from a BRCA2-DSS1-ssDNA Structure." *Science* 297 (5588): 1837 LP-1848. http://science.sciencemag.org/content/297/5588/1837.abstract.

Yang, Weidong, Thimmaiah P Chendrimada, Qingde Wang, Miyoko Higuchi, Peter H Seeburg, Ramin Shiekhattar, and Kazuko Nishikura. 2006. "Modulation of microRNA Processing and Expression through RNA Editing by ADAR Deaminases." *Nature Structural & Molecular Biology* 13 (1): 13–21. doi:10.1038/nsmb1041.

Yanling Zhao, Dorothy, Gerald Gish, Ulrich Braunschweig, Yue Li, Zuyao Ni, Frank W. Schmitges, Guoqing Zhong, et al. 2015. "SMN and Symmetric Arginine Dimethylation of RNA Polymerase II C-Terminal Domain Control Termination." *Nature* 529 (7584): 48–53. doi:10.1038/nature16469.

Yates, Andrew, Wasiu Akanni, M. Ridwan Amode, Daniel Barrell, Konstantinos Billis, Denise Carvalho-Silva, Carla Cummins, et al. 2016. “Ensembl 2016.” *Nucleic Acids Research* 44 (D1): D710–16. doi:10.1093/nar/gkv1157.

Ye, Xuecheng, Nian Huang, Ying Liu, Zain Paroo, Carlos Huerta, Peng Li, She Chen, Qinghua Liu, and Hong Zhang. 2011. “Structure of C3PO and Mechanism of Human RISC Activation.” *Nature Structural & Molecular Biology* 18 (6): 650–57. doi:10.1038/nsmb.2032.

Yeom, Kyu Hyeon, Yoontae Lee, Jinju Han, Mi Ra Suh, and V. Narry Kim. 2006. “Characterization of DGCR8/Pasha, the Essential Cofactor for Drosha in Primary miRNA Processing.” *Nucleic Acids Research* 34 (16): 4622–29. doi:10.1093/nar/gkl458.

Yi, Rui, Yi Qin, Ian G. Macara, and Bryan R. Cullen. 2003. “Exportin-5 Mediates the Nuclear Export of Pre-microRNAs and Short Hairpin RNAs.” *Genes and Development* 17 (24): 3011–16. doi:10.1101/gad.1158803.

Yoo, S, and W S Dynan. 1999. “Geometry of a Complex Formed by Double Strand Break Repair Proteins at a Single DNA End: Recruitment of DNA-PKcs Induces Inward Translocation of Ku Protein.” *Nucleic Acids Research* 27 (24): 4679–86. doi:10.1093/nar/27.24.4679.

Zaborowska, Justyna, Sylvain Egloff, and Shona Murphy. 2016. “The Pol II CTD: New Twists in the Tail.” *Nature Structural & Molecular Biology* 23 (9): 771–77. doi:10.1038/nsmb.3285.

Zhang, Feng, Jianglin Ma, Jiaxue Wu, Lin Ye, Hong Cai, Bing Xia, and Xiaochun Yu. 2009. “PALB2 Links BRCA1 and BRCA2 in the DNA-Damage Response.” *Current Biology* 19 (6): 524–29. doi:10.1016/j.cub.2009.02.018.

Zhang, Huimin, Frank Rigo, and Harold G. Martinson. 2015. “Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism That Does Not Require Cleavage at the Poly(A) Site.” *Molecular Cell* 59 (3): 437–48. doi:10.1016/j.molcel.2015.06.008.

Zhao, Weixing, Sivaraja Vaithiyalingam, Joseph San Filippo, David G. Maranon, Judit Jimenez-Sainz, Gerald V. Fontenay, Youngho Kwon, et al. 2015. “Promotion of BRCA2-Dependent Homologous Recombination by DSS1 via RPA Targeting and DNA Mimicry.” *Molecular Cell* 59 (2): 176–87. doi:10.1016/j.molcel.2015.05.032.

Zhou, Chun Hong, Xiao Peng Zhang, Feng Liu, and Wei Wang. 2014. “Involvement of miR-605 and miR-34a in the DNA Damage Response Promotes Apoptosis Induction.” *Biophysical Journal* 106 (8): 1792–1800. doi:10.1016/j.bpj.2014.02.032.

Zhu, Zhu, Woo Hyun Chung, Eun Yong Shim, Sang Eun Lee, and Grzegorz Ira. 2008. “Sgs1 Helicase and Two Nucleases Dna2 and Exo1 Resect DNA Double-Strand Break Ends.” *Cell* 134 (6): 981–94. doi:10.1016/j.cell.2008.08.037.

8 Appendix

8.1 Statement of contributions to the project

For section 5.3.1 and 5.4 Wei-Ting Lu performed the DNA:RNA immunoprecipitation (DRIP). This involved growing the cells, inducing damage, harvesting the material and performing DRIP. Wei-Ting Lu also performed the qPCR to validate the experiment. Ben Hawley received the DRIP-ed material and continued from that point (library preparation, sequencing analysis).

At the time of final printing, the work in this thesis and additional work by Wei-Ting Lu was accepted for publication in Nature Communications with the title "Drosha drives the formation of DNA:RNA hybrids around DNA break sites to facilitate DNA repair".

8.2 Tables for methods

The section contains all the tables of antibodies, primers, and siRNA used in this thesis. Also included is a list of the 99 AsiSI sites and the 99 matched control sites used in Chapters 4 and 5.

Target	Supplier	Cat. No.	Sequence
Control	Dharmacon	D-001810-03	Proprietary sequence
Drosha	Dharmacon (custom)	(Dhir et al. 2015)	CGAGUAGGCUUCGUGACUUdTdT
Dicer	Ambion Silencer Select	s23754	GAUCCUAUGUUCAAUCUAAdTdT
TNRC6A	Ambion Silencer Select	s26154	GTCGCAAAATGGAGATTGAdTdT
TNRC6B	Ambion Silencer Select	s23060	CTCCCATTGTAGATACTGAdTdT
TNRC6C	Ambion Silencer Select	s33601	CCAAGAGUUCUGUCUAAUAdTdT
BRCA1	Dharmacon	J-003461-09	CAACAUGCCCACAGAUCAAdTdT
53BP1	Ambion Silencer Select	s14314	GCACACUUGUCACUCGUGUdTdT
DGCR8	Dharmacon (custom)	(Dhir et al. 2015)	CAUCGGACAAGAGUGUGAUdTdT
DDX5	Ambion Silencer Select	s4009	GGAUUAUAUUCAUCGAAUUdTdT
DDX17	Ambion Silencer Select	s20622	GAGAGACUCUGCAAGCUAUdTdT
Ago1	Dharmacon	L-004638-00	(pool of 4 sequences)
Ago2	Dharmacon	L-004639-00	(pool of 4 sequences)

Table 8-1 – List of siRNA.
siRNAs used in this thesis are indicated with supplier and sequence information.

Antibody	Supplier	Cat. No.	Application	Dilution
Drosha	Santa Cruz	sc39591	WB	1:500
Dicer	Santa Cruz	sc30226	WB	1:500
TNRC6	Novus	NBP1-28757	WB	1:1500
HA	Roche	1867423001	WB, IF	1:500
γH2AX	Millipore	05-636	IF	1:500
γH2AX	Cell Signalling	9718s	WB	1:1000
53BP1	Abcam	ab36823	WB, IF	1:500
BRCA1	Santa Cruz	sc6954	WB, IF	1:50
Ago2	Abcam	ab32381	WB	1:2000
ATM	Santa Cruz	sc23921	WB	1:500
pATM (Ser1981)	Cell Signalling	4526	IF	1:500
pATM (Ser1981)	Abcam	ab81292	WB	1:2000
GAPDH	Santa Cruz	sc32233	WB	1:2000
MDC1	Abcam	ab11169	IF	1:600
RAD51	Santa Cruz	sc8349	IF	1:50
Vinculin	Abcam	ab18058	WB	1:3000
S9.6	Kerafast	ENH001	DRIP	1μg/2μg DNA

Table 8-2 – List of antibodies.
Antibodies used for western blotting (WB), immunofluorescence (IF), or DNA:RNA hybrid immunoprecipitation (DRIP) is indicated with dilutions used.

Plasmid	Use	Reference	Details
pBABE-HA-ER-AsiSI	-	(Iacovoni et al. 2010)	Sub-cloned into pCI
pCI-AsiSI	Expression in U2OS (Figure X)	-	1µg/1e6 cells
$RNH1^{mut}$	Laser microirradiation	Addgene #60367	250ng/well
RNase H1	Over-expression	Addgene #65784	2.5µg/6-well
HR	Repair reporter cell line	(Seluanov, Mao, and Gorbunova 2010)	Integrated into U2OS
NHEJ	Repair reporter cell line	(Seluanov, Mao, and Gorbunova 2010)	Integrated into U2OS
I-SceI	Induce repair reporter break	(Seluanov, Mao, and Gorbunova 2010)	2µg/6-well

Table 8-3 – List of plasmids.
Plasmids as used in this thesis are collated here, with details of where this was obtained and amounts used.

Primer	Sequence	Use
pBABE-AsiSI-F	TTGGTCCGCGGAATTCACCATGGCATACCC	Cloning
pBABE-AsiSI-R	AGCTGGTCGACTCACAACATC	Cloning
pCI-Neo-F	AACCACCGCGGATCGCTCGAGGCTAGCCTATAG	Cloning
pCI-Neo-R	TTGGTGTCGACGGTTCCCAATAGCTGAAGCGG	Cloning
HR1	FWD:CCGCCAGAAAGTTTCCTAGA REV:CTCACCCTTGCAGCACTTG	DRIP
HR2	FWD:GAGGAGCGCAGGACACTG REV:CCAATTAGAGACCACCCGTTT	DRIP
NHEJ1	FWD:ATCGGGCCAATCTCAGAGG REV:GCGACGCTAACGTTAAAGCA	DRIP
NHEJ2	FWD:GGTGCCCACAGCTCTCTATG REV:GAAGCCAGAGGAGTGTCCTG	DRIP
γActin exon5	FWD:GTGACACAGCATCACTAAGG REV:ACAGCACCGTGTTGGCGT	DRIP

Table 8-4 – List of primers.
Primers used in this study for cloning or qPCR.

Chr	Start	Stop	Repair
1	9589386	9589394	NHEJ
1	9651806	9651814	Ambiguous
1	13749685	13749693	Ambiguous
1	19485659	19485667	HR
1	40508970	40508978	NHEJ
1	88992912	88992920	NHEJ
1	109494076	109494084	Ambiguous
1	109984951	109984959	Ambiguous
1	230868486	230868494	Ambiguous
1	242524475	242524483	Ambiguous
10	92291256	92291264	Ambiguous
11	24496928	24496936	NHEJ
11	75814714	75814722	NHEJ
11	85664609	85664617	Ambiguous
12	121328023	121328031	Ambiguous
12	121537151	121537159	Ambiguous
14	52695637	52695645	HR
14	52778790	52778798	Ambiguous
14	54489106	54489114	Ambiguous
15	58933089	58933097	Ambiguous
15	68819780	68819788	NHEJ
15	91853707	91853715	Ambiguous
17	4366227	4366235	HR
17	5486899	5486907	HR
17	6070641	6070649	HR
17	21042986	21042994	HR
17	39868135	39868143	Ambiguous
17	40100150	40100158	Ambiguous
17	44237549	44237557	Ambiguous
17	44363565	44363573	Ambiguous
17	59106934	59106942	Ambiguous
17	63773494	63773502	NHEJ
17	73645084	73645092	Ambiguous
17	82273233	82273241	HR
17	82292963	82292971	HR
18	7566713	7566721	Ambiguous
18	7568353	7568361	HR
18	21740842	21740850	Ambiguous
19	2302877	2302885	Ambiguous
19	2455993	2456001	HR
19	2456094	2456102	HR
19	2456596	2456604	HR
19	29528579	29528587	NHEJ
19	41397836	41397844	Ambiguous
2	5697062	5697070	Ambiguous
2	16665576	16665584	Ambiguous
2	43131198	43131206	NHEJ
2	55281963	55281971	HR
2	68157615	68157623	NHEJ
2	85595469	85595477	HR
2	119366988	119366996	NHEJ
2	207166002	207166010	NHEJ
2	242088413	242088421	Ambiguous
20	32358508	32358516	HR
20	33444279	33444287	HR
20	43458476	43458484	HR
21	31873204	31873212	NHEJ
21	44801873	44801881	Ambiguous
22	20358480	20358488	Ambiguous
22	20436353	20436361	Ambiguous
22	20496019	20496027	HR
22	38468095	38468103	HR
3	39809478	39809486	NHEJ
3	52198145	52198153	HR
3	179148748	179148756	Ambiguous
3	196868109	196868117	Ambiguous

4	2058957	2058965	Ambiguous
4	83013132	83013140	Ambiguous
4	177442420	177442428	NHEJ
5	11903656	11903664	Ambiguous
5	69167022	69167030	Ambiguous
5	80488319	80488327	Ambiguous
5	137498544	137498552	Ambiguous
5	143405483	143405491	NHEJ
6	27177586	27177594	Ambiguous
6	37354034	37354042	Ambiguous
6	49949868	49949876	NHEJ
6	89638466	89638474	NHEJ
6	135498208	135498216	NHEJ
6	144286431	144286439	NHEJ
6	149566968	149566976	NHEJ
7	2727436	2727444	HR
7	4682497	4682505	Ambiguous
7	76178187	76178195	NHEJ
7	93232176	93232184	NHEJ
7	100081883	100081891	Ambiguous
7	101815567	101815575	Ambiguous
8	22565957	22565965	Ambiguous
8	115668403	115668411	Ambiguous
9	29212800	29212808	NHEJ
9	36258515	36258523	HR
9	68356723	68356731	Ambiguous
9	127930890	127930898	HR
9	128127127	128127135	HR
9	134326229	134326237	HR
X	45507147	45507155	Ambiguous
X	48957420	48957428	Ambiguous
X	53082243	53082251	Ambiguous
X	73563265	73563273	NHEJ

Table 8-5 – List of 99 cut AsiSI sites used in this thesis

List received from Gaelle Legube (University of Toulouse, France), and described in (Aymard et al. 2014; Aymard et al. 2017).

Chr	Start	Stop	Gene
1	39841130	39843138	TRIT1
1	46040396	46042404	PIK3R3
1	89005725	89007733	GBP3
1	1.03E+08	1.03E+08	COL11A1
1	1.61E+08	1.61E+08	USP21
1	2.17E+08	2.17E+08	GPATCH2
10	70204678	70206686	PPA1
10	95915734	95917742	CC2D2B
10	1.05E+08	1.05E+08	
10	1.25E+08	1.25E+08	FAM53B
11	4094472	4096480	RRM1
11	14442694	14444702	COPB1
12	2875836	2877844	TULP3
12	6683767	6685775	ZNF384
12	12926758	12928766	RP11-392P7.6
12	53267973	53269981	ESPL1
12	53507602	53509610	ATF7
12	77325212	77327220	NAV3
12	1.03E+08	1.03E+08	RP11-328J6.1
12	1.12E+08	1.12E+08	PTPN11
12	1.21E+08	1.21E+08	MLEC
13	17001097	17003105	
13	44431014	44433022	TSC22D1
14	5035905	5037913	
14	21384472	21386480	CHD8
14	49962456	49964464	LINC01588
14	69258669	69260677	GALNT16
14	87866840	87868848	GALC
14	94104519	94106527	IFI27
15	37467633	37469641	RP11-720L8.1
15	44287565	44289573	CASC4
15	54011862	54013870	UNC13C
16	16231588	16233596	NOMO3
16	24660393	24662401	LINC01567
17	4530107	4532115	SPNS2
17	15380944	15382952	RP11-849N15.4
17	35263007	35265015	SLFN5
18	5953918	5955926	L3MBTL4
18	36969367	36971375	KIAA1328
19	2731433	2733441	SLC39A3
19	17718275	17720283	MAP1S
19	19532256	19534264	YJEFN3
19	19784443	19786451	ZNF506
19	31627058	31629066	
19	42312661	42314669	TMEM145
19	42400971	42402979	LIPE
19	51684929	51686937	SPACA6P-AS
2	6960247	6962255	RNF144A
2	38750671	38752679	GEMIN6
2	42404410	42406418	COX7A2L
2	86505075	86507083	RNF103-CHMP3
2	94866650	94868658	AC097374.2
2	2.13E+08	2.13E+08	AC079610.1
2	2.39E+08	2.39E+08	
20	20388564	20390572	RALGAPA2
20	23625780	23627788	CST3
3	14135681	14137689	RP11-434D12.1
3	15069464	15071472	RBSN
3	25597158	25599166	TOP2B
3	40519931	40521939	
3	50351021	50353029	CYB561D2
3	66377862	66379870	LRIG1
3	1.34E+08	1.34E+08	CEP63
3	1.65E+08	1.65E+08	
3	1.71E+08	1.71E+08	RPL22L1
4	2070807	2072815	POLN

4	75513903	75515911	THAP6
4	1.03E+08	1.03E+08	CENPE
4	1.23E+08	1.23E+08	FGF2
5	4356969	4358977	
5	53882608	53884616	ARL15
5	58453530	58455538	PLK2
5	94703471	94705479	MCTP1
5	1.11E+08	1.11E+08	WDR36
5	1.12E+08	1.12E+08	STARD4
5	1.35E+08	1.35E+08	PCBD2
5	1.36E+08	1.36E+08	LECT2
5	1.81E+08	1.81E+08	RACK1
6	31398687	31400695	MICA
6	1.44E+08	1.44E+08	ZC2HC1B
6	1.44E+08	1.44E+08	UTRN
7	25854284	25856292	
7	28127980	28129988	JAZF1
7	55886825	55888833	ZNF713
7	99438578	99440586	CPSF4
7	1E+08	1E+08	AP4M1
7	1.5E+08	1.5E+08	ZNF767P
8	19816629	19818637	INTS10
8	79985376	79987384	MRPS28
8	80484681	80486689	ZBTB10
8	86332172	86334180	CTD-2284J15.1
8	92765003	92767011	RP11-100L22.3
9	81976992	81979000	RP11-15B24.5
X	41980835	41982843	
X	72809361	72811369	DMRTC1B
X	73611687	73613695	CHIC1

Table 8-6 – List of the transcriptionally matched control sites

This list was generating by find comparable genes to AsiSI cut genes (similar length, TPM in 4sU-seq). For each gene, a pseudo-AsiSI site was assigned to be approximately as close to the TSS as the matched AsiSI site was to its gene.