Establishing a method to monitor replicative polymerase usage genome-wide in human cells

David John Buist

Submitted for the degree of Doctor of Philosophy

University of Sussex

September 2018

Declaration

I hereby declare that this thesis has not been and will not be submitted in whole or in part to another University for the award of any other degree.

Signature:

Date:

Acknowledgements

I would like to thank Tony Carr for allowing me the opportunity to be a part of his lab. I would also like to thank Helfrid Hochegger for his role as my co-supervisor and for guiding much of the human cell work.

I am indebted to Owen Wells and Nadia Hegarat for their patience, their ability to steer me in the right direction at every juncture and for their technical knowledge. Without them, this thesis would be a lot shorter. I am also grateful to Adam Watson for his help with the yeast work in this project, but in particular for his willingness to help anyone, at any time, regardless of his own workload.

As part of this project I was lucky enough to spend 7 weeks in the Takeda laboratory in Kyoto, Japan. I am grateful to Professor Shunichi Takeda for welcoming me into his lab and to Hiroyuki Sasanuma for teaching me gene targeting techniques.

I was lucky to be able to draw on the academic experience of Philippe Dedieu, Roberto Steiner, Marc Sitbon and Naomi Taylor, who all offered words of encouragement along the way

UNIVERSITY OF SUSSEX

David Buist

A thesis submitted for the degree of Doctor of Philosophy

Establishing a method to monitor replicative polymerase usage genome-wide in human cells

The majority of DNA replication in eukaryotic cells is carried out by two replicative DNA polymerases: polymerases δ and ε (Pols δ and ε). Whilst both polymerases exhibit high specificity for deoxyribonucleotides, both frequently misincorporate ribonucleotides (rNTPs) which are rapidly removed via ribonucleotide excision repair (RER) by the heterotrimeric enzyme RNase H2.

In yeast, using mutant polymerases that incorporate elevated levels of rNTPs (and in the absence of the catalytic subunit of RNase H2 (RNase H2A)) unrepaired rNTPs serve as a molecular footprint of the replicative polymerases. This allows the determination of genome-wide polymerase usage using a deep sequencing methodology that we have termed polymerase usage sequencing (Pu-seq).

The aim of this project is to develop cell lines that can be used for Pu-seq in human cells. Ultimately, this system could then be used to monitor DNA replication dynamics in human cells to aid in the study of replication stress responses. As in yeast, the development of the Pu-seq system in human cells requires an absence of RNase H2 activity and mutant alleles of Pol δ and Polε that result in excess ribonucleotide incorporation.

This thesis describes the steps taken towards developing Pu-seq in human cell lines; firstly, by establishing that in human cells expressing normal levels of p53, RNase H2A is an essential gene, then by developing mechanisms that allow for the shutoff of RNase H2 activity through the targeted downregulation of one of its accessory subunits, RNase H2C. Attempts to develop a novel method of quantifying genomic ribonucleotides are described in addition to showing that previously published protocols for assessing overall RNase H2 activity and genomic ribonucleotide presence provide the sensitivity required to screen candidate human Pu-seq cell lines. Finally, progress made in screening mutations in Pols δ and ε that may cause increased ribonucleotide incorporation is described.

Contents

List of tables

List of figures

List of abbreviations

ACS	Autonomous consensus sequence
AGS	Aicardi-Goutières syndrome
AID	Auxin-inducible degron
ARS	Autonomous replication sequences
ASV	Asunaprevir
AtRNL	*Arabidopsis thaliana* tRNA ligase
CFS	Common fragile site
CRISPR	Clustered regularly interspaced short palindromic repeats
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide
Dox	doxycycline
DSB	Double strand break
G4	G-quadruplex
gDNA	Genomic DNA
gRNA	guide RNA
HR	Homologous recombination
IAA	Indole-3-acetic acid
mAID	Mini auxin-inducible degron
MCM	Minichromosome maintenance complex
MMR	Mismatch repair
NHEJ	Non-homologous end joining
OF	Okazaki fragment
OGRE	Origin G-rich repeated element
ORC	Origin recognition complex
OsTIR1	*Oryza sativa* transport inhibitor response 1
PAM	Protospacer-adjacent motif
PCNA	Proliferating cell nuclear antigen
PIP	PCNA-interacting protein
pre-IC	Pre-initiation complex
pre-RC	Pre-replication complex
Pu-Seq	Polymerase usage sequencing
rDNA	Ribosomal DNA
RER	Ribonucleotide excision repair
RFB	Replication fork barrier
RNA	Ribonucleic acid
rNTP	Ribonucleotide
SMASh	Small molecule-assisted shutoff
ssDNA	Single-stranded DNA
Tc	Tetracycline
TCTA	Tetracycline controlled transcriptional activation

Chapter 1

Introduction

1.1 General introduction

Before cells divide they must make a copy of their entire genome during a process known as DNA replication. The majority of DNA replication is carried out by two replicative DNA polymerases, Polymerase delta (Polδ) which replicates the lagging strand and Polymerase epsilon (Polε) which replicates the leading strand. Accurate replication of DNA is essential as mutations arising from endogenous or exogenous sources can lead to genome instability and genetic diseases.

In yeast, the genome-wide division of labour of the two major replicative polymerases can be mapped using mutant polymerases that incorporate elevated levels of ribonucleotides. In these cells, the catalytic subunit of RNase H2 (RNase H2A), the major enzyme responsible for the removal of misincorporated ribonucleotides, is knocked out allowing for unrepaired ribonucleotides to serve as a molecular footprint of the mutant replicative polymerases. The presence of unrepaired genomic ribonucleotides allows genome-wide mapping of polymerase usage using a deep sequencing methodology that has been termed polymerase usage sequencing (Pu-seq) (Daigaku et al. 2015).

The aim of this project is to develop cell lines that can be used for Pu-seq in human cells. As in yeast, the development of the Pu-seq system in human cells requires an absence of RNase H2 activity and mutant alleles of Pol δ and Polε that result in excess ribonucleotide incorporation. This thesis described progress made towards developing human Pu-seq cell lines as well as the development of assays to assess RNase H2 activity and genomic ribonucleotide incorporation.

1.2 DNA replication

1.2.1 DNA replication overview

Before undergoing mitosis, cells have to fully replicate the entire genome during S phase in a tightly regulated and highly accurate manner. The process of genomic duplication can be loosely divided into two distinct steps: unwinding of the DNA duplex to expose the sequence of each strand, followed by semiconservative replication of each template. The point at which duplex unwinding begins marks the site of DNA replication initiation at specific sites termed replication origins. The majority of prokaryotes have single, circular chromosomes that are replicated from a single, sequence-specific origin of replication. Following initiation, replication proceeds in a bi-directional manner to replicate the entire chromosome (O'Donnell et al. 2013). Eukaryotic replication is a more complicated process, with the linear, typically numerous, chromosome pairs replicated from multiple origins interspersed along the length of each chromosome, with each replication fork proceeding until it meets a converging fork at which point replication terminates (Figure 1-1). The number of origins contained within an entire genome correlates approximately with the total length of the genome with an estimated 30,000-50,000 activated origins in human cells (Mechali 2010). The importance of the accurate coordination of replication initiation is evident when one considers that at any one time an adult human body is composed of an estimated 3.7×10^{13} cells (Bianconi et al. 2013) requiring the correct propagation of in the region of 1.5×10^{18} replication forks. The total number of replication events increases significantly when one considers the renewal of cells that takes place constantly throughout the human lifespan.

1.2.2 Eukaryotic origin identity

Among the widely used eukaryotic model organisms, only the budding yeast *Saccharomyces cerevisiae* is known to use consensus sequence-specific replication origins known as autonomous replication sequences (ARS). ARS are marked by a well-defined 11-12 bp AT-rich consensus sequence termed the autonomous consensus sequence (ACS) which can also function as a replicator when inserted into plasmids (Stinchcomb et al. 1979; Marahrens & Stillman 1992; Shirahige et al. 1993; Rao et al. 1994). The presence of an ACS does not however guarantee that replication will begin

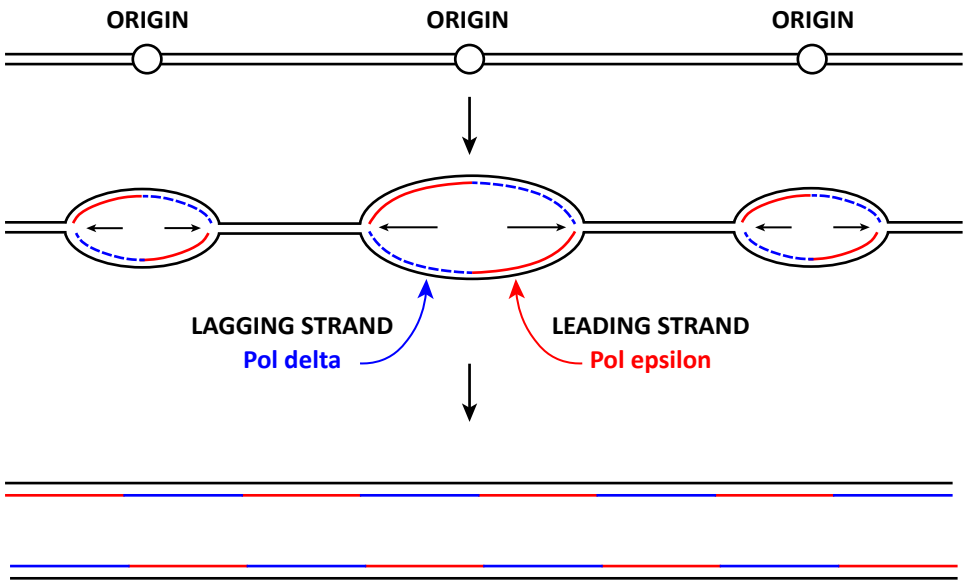


Figure 1-1 Mechanism of eukaryotic DNA replication
Eukaryotic DNA replication starts at multiple points along each chromosome termed origins. Once licensed to fire, replication from each origin proceeds in a bidirectional manner with continuous synthesis on the leading strand by polymerase (Pol) epsilon and discontinuous synthesis on the lagging strand by Pol delta. Each replication fork proceeds until it encounters an oncoming fork, at which point replication terminates. The result of replication is two daughter DNA molecules, each consisting of one strand derived from the original parent.

at that site since there are over 12,000 potential ACS sites in the *S. cerevisiae* genome and only around 400 are used to initiate replication (Nieduszynski et al. 2006).

Origin identity in fission yeast *Schizosaccharomyces pombe* is less well-defined, although origins are marked by the presence of sequence-independent AT-rich regions up to 1kb in length (Clyne & Kelly 1995; Segurado et al. 2003). Furthermore, replacement of AT-rich regions in *S. pombe* ARS that are essential for replication initiation, with regions containing polyA/T stretches has little effect on origin usage (Okuno et al. 1999).

In higher eukaryotes there are no known replication origins that are dependent upon specific sequences; rather origins are marked by features of the genomic landscape. In human, mouse and *Drosophila*, origins have been found to cluster in GC-rich regions and in regions within or near to CpG islands (Necsulea et al. 2009; Sequeira-Mendes et al. 2009). Bioinformatic analyses have identified G-rich clusters termed origin G-rich repeated elements (OGREs) around origins in the higher eukaryotes. Certain guanine nucleotides within OGREs are capable of forming G-quadruplexes, non-canonical DNA structures which have been identified in human origins (Besnard et al. 2012; Cayrou, Coulombe, et al. 2012; Cayrou, Gregoire, et al. 2012) and are thought to play a role in replication initiation (Valton et al. 2014).

1.2.3 Establishing the pre-replication complex

The initial step in replication initiation, termed licensing, is the assembly of the pre-replication complex (pre-RC) at all origins. This process begins with the recognition of, and the subsequent loading to origins of the origin recognition complex (ORC). ORC is a hetero-hexamer composed of Orc1-Orc6. In *S. cerevisiae* and *S. pombe*, it is thought that ORC remains bound to chromatin throughout the cell cycle (Newlon 1997; Lygerou & Nurse 1999), whilst in humans ORC binding occurs only during G1 (Tatsumi et al. 2003). Of the six ORC proteins, Orc1 is the first to bind DNA upon assembly of the Pre-RC. The cell cycle phase-dependent binding of ORC in humans is a result of a tightly regulated process affecting Orc1 expression levels, with accumulation of Orc1 limited to mid to late G1, followed by ubiquitin-mediated degradation of Orc1 at the G2-M transition decreasing both ORC stability and its ability to be recruited to origins. Despite

the absence of Orc1 from chromatin outside G1, other components of ORC remain chromatin-associated (Ladenburger et al. 2002; Tatsumi et al. 2003; Siddiqui & Stillman 2007; Kara et al. 2015). This tight regulation of Orc1 expression, and subsequent levels of ORC stability, contributes to the once-per-cell-cycle nature of origin licensing (Mechali 2010; Yardimci & Walter 2014).

Once ORC has bound origins, it acts as a recruitment factor for loading of Cdc6, an AAA^+ ATPase, which interacts with ORC via Orc3 and Orc1, which in turn interact with the other ORC subunits, forming a ring-like structure around the origin (Sun et al. 2012; Yardimci & Walter 2014). The next step in Pre-RC assembly is the loading of the core motor of the main replicative helicase, the minichromosome maintenance complex (MCM) a hetero-hexamer of six related AAA^+ ATPases, MCM2-MCM7, which are initially recruited to ORC-Cdc6 via interactions between MCM3 and Cdc6 (Fernandez-Cid et al. 2013; Deegan & Diffley 2016). Further stability of the ORC-Cdc6-MCM complex is provided by Cdt1, which interacts with MCM principally via MCM2, but also via MCM5 and MCM6 (J. Sun et al. 2013) (Figure 1-2). Cdt1 plays an important role in alleviating the autoinhibitory effect of MCM6, which in the absence of Cdt1 prevents MCM loading onto ORC-Cdc6 (Fernandez-Cid et al. 2013; Riera et al. 2013). MCM loading is dependent upon both the presence of Cdc6 and subsequent ATP hydrolysis by Cdc6, which in turn leads to release of Cdt1 from the Pre-RC (Randell et al. 2006; Remus et al. 2009). MCM loading efficiency is also influenced by the chromatin architecture surrounding origins (Sugimoto et al. 2015).

1.2.4 Establishing the pre-initiation complex

In vitro studies have found that MCM is loaded as an inactive head-to-head double hexamer encircling double stranded DNA (dsDNA) via a process requiring ATP hydrolysis (Remus et al. 2009; Evrin et al. 2009; Gambus et al. 2011), although the precise mechanism through which the two MCM hexamers are sequentially loaded remains unknown (Yardimci & Walter 2014; Deegan & Diffley 2016). The progression of replication once the pre-RC complex is formed requires the association of further replication factors with MCM to form the pre-initiation complex (pre-IC), ultimately resulting in the active form of the replicative helicase, the Cdc45-MCM-GINS (CMG)

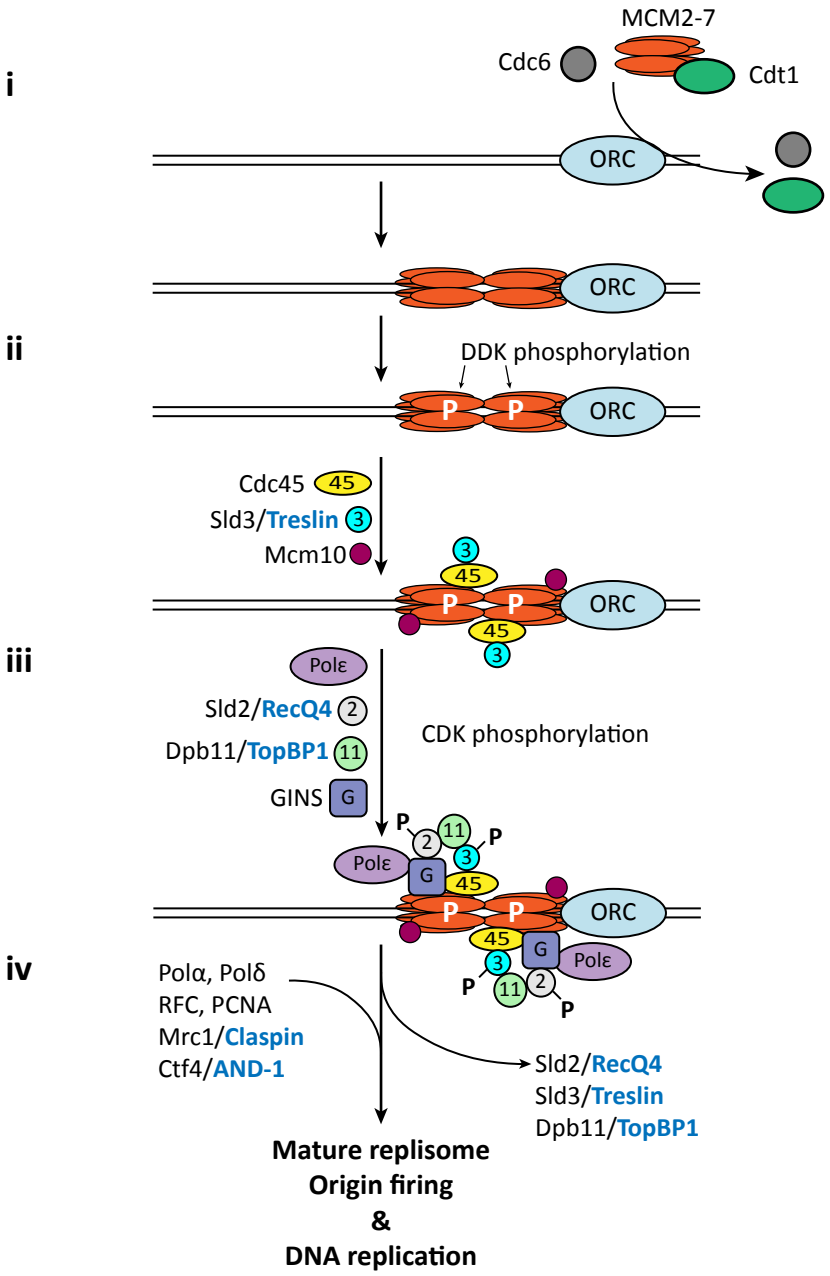


Figure 1-2 Eukaryotic DNA replication origin activation
Steps involved in the activation and firing of eukaryotic origins. i) recruitment of the MCM replicative helicase to the origin recognition complex (ORC) by Cdt1 and Cdc6. ii) MCM phosphorylation by DDK causing recruitment of Cdc45, Sld3/Treslin and Mcm10. iii) CDK mediated phosphorylation resulting in recruitment of polymerase epsilon, Sld2/RecQ4, Dpb11/TopBP1 and the tetrameric GINS complex. iv) Subsequent steps recruit the factors necessary to assemble the mature replisome. Putative human homologues labelled in blue. Adapted from O'Donnell et al. (2013).

complex. In *S. cerevisiae*, the loading and activation of the CMG complex is well characterized and requires further firing factors, including Sld2, Sld3, Sld7, Dpb11, and Pol ε, along with the action of two S-phase promoting kinases, cyclin-dependent kinase CDK and Dbf4-dependent Cdc7 kinase DDK (Labib 2010; On et al. 2014) (Figure 1-2).

The mechanisms of pre-IC assembly and the roles of CDK and DDK in CMG activation in mammalian cells are less well understood than in yeast systems. Nevertheless several homologues of yeast proteins have been identified in metazoan systems, including TopBP1, RecQ4 and Treslin/Ticrr, homologues of Dpb11, Sld2 and Sld3 respectively (Tanaka & Araki 2013). TopBP1 and Treslin both bind to the MCM hexamer and MCM10 is recruited to the Pre-IC and is required for helicase activation following multiple phosphorylations by CDK and DDK of the MCM helicase subunits and RecQ4 (Kanke et al. 2012; Im et al. 2015). Importantly, MCM10 has been shown to stimulate CMG activity ~30-fold (Langston et al. 2017).

1.2.5 Priming replication and the replisome

Bidirectional unwinding of origin DNA by the MCM helicase starts during CMG assembly and continues following CMG activation by MCM10 exposing single-stranded DNA (ssDNA) that is coated with RPA (Sun et al. 2015; Burgers & Kunkel 2017; Douglas et al. 2018). Exposed origin ssDNA is used as a template for priming by the primase subunit of tetrameric Polymerase alpha (Polα), which is recruited to the replisome by MCM10 and And-1/Ctf4 (Zhu et al. 2007). Polα primes replication by polymerizing a short RNA (~10-12 nucleotides)-DNA (~20 nucleotides) primer; this is an important step since neither of the two major replicative polymerases, Polymerase delta (Polδ) and Polymerase epsilon (Polε) are able to initiate nucleotide polymerization *ab initio* (Frick & Richardson 2001).

The recruitment of both Polδ and Polε to the replisome is mediated by their interactions with the molecular scaffold Proliferating Cell Nuclear Antigen (PCNA). PCNA is loaded at replication origins by replication factor C (RFC) (Tsurimoto & Stillman 1990) and its presence increases the processivity of *S. cerevisiae* Polδ ~100-fold and Polε ~6-fold *in*

vitro (Chilkova et al. 2007). PCNA also interacts with several other proteins at the replication fork to help mediate process such as Okazaki fragment maturation, mismatch repair, nucleotide excision repair and translesion synthesis (Boehm et al. 2016).

Following priming, replication proceeds with continuous synthesis on the leading strand by Polε, which in *S. cerevisiae* is recruited to origins via interactions with GINS, Dpb11 and Sld2 and discontinuous synthesis on the lagging strand by Polδ (Pursell et al. 2007; Nick McElhinny et al. 2008; Muramatsu et al. 2010; Miyabe et al. 2011; Langston et al. 2014; Daigaku et al. 2015).

1.3 Replication stress

Replication stress is the term used to describe when impediments from endogenous and exogenous sources cause slowing, stalling or collapse of replication fork progression and the DNA damage that results from such erroneous replication, including the exposure of single-stranded DNA and DNA double strand breaks (Mazouzi et al. 2014; Zeman & Cimprich 2014; Macheret & Halazonetis 2015; O'Driscoll 2017).

1.3.1 Causes of replication stress

Replication stress can be caused by a wide range of different phenomena, mostly arising from intrinsic cellular metabolic processes and features relating to DNA sequence and topology but also from exogenous sources and from the consequences of mutations in genes regulating DNA metabolism and the cell cycle (Zeman & Cimprich 2014).

Replication fork progression may be impeded by a variety of endogenously-arising and exogenously-derived unrepaired DNA lesions, including, but not limited to abasic sites, UV-induced lesions and aromatic DNA adducts (Lopes et al. 2006; Ciccia & Elledge 2010; Zeman & Cimprich 2014; Kidane et al. 2014). Impediments to replication also arise due to the presence of non-B form secondary structure within DNA, formed within AT-rich regions (Zhang & Freudenreich 2007) and within GC-rich regions capable of forming G-

quadruplexes (G4) (Burge et al. 2006; Lopes et al. 2011). Hairpin and cruciform DNA structures caused by inverted repeat sequences also present a problem for active replication forks (Voineagu et al. 2008; Boyer et al. 2013; Inagaki et al. 2013). DNA replication stress may also be caused by structure-forming tri-nucleotide repetitive DNA sequences that cause intramolecular slippage of the replicative polymerases (Samadashwily et al. 1997); indeed several human diseases have been found to be caused by the expansion of nucleotide repeats from three up to twelve nucleotides in length, located within the coding regions, 3'- and 5'-untranslated regions, intronic sequences and promoter regions of genes (Mirkin 2007).

Eukaryotic telomeres and subtelomeric regions are particularly prone to replication stress since they contain G-rich repetitive sequences that are prone to forming secondary structures including G4, t-loops, heterochromatin-like structures as well as structures that have not been seen *in vivo* but are presumed possible from *in vitro* studies such as triple helices, four-way junctions and D-loops (Gilson & Geli 2007).

Both DNA replication and RNA transcription compete for the same template. To avoid collisions between replication and transcription machinery, cells separate the two processes spatially and temporally (discussed in section 1.3.2). Despite this, head-on collisions between replication and transcription machinery travelling in opposing directions and codirectional collisions caused by replication and transcription on the same template do occur; these collisions have been found to cause fork stalling resulting in genome instability in eukaryotes (Deshpande & Newlon 1996; Prado & Aguilera 2005; Azvolinsky et al. 2009; Garcia-Rubio et al. 2018). Replication fork stalling as a result of transcription may not be solely derived from steric hindrance. The unwinding of DNA necessary for replication and transcription generates positive supercoiling ahead of each set of machinery. This leads to topological stress between opposing replication and transcription machinery which prevents the progression of the replication fork (Olavarrieta et al. 2002; Bermejo et al. 2011; Bermejo et al. 2012; Keszthelyi et al. 2016).

Opposing transcription and replication pose a further problem to replisomes due to the formation of three-stranded structures composed of an RNA-DNA duplex and free

single-stranded DNA known as R-loops. During transcription R-loops may be formed by the interaction of the nascent transcript with exposed DNA at the transcription bubble behind the transcription machinery. R-loops form preferentially in negatively supercoiled DNA such as that found behind RNA polymerases and their stabilization at such positions is aided by the excess positive supercoiling generated by the oncoming replisome (Zeman & Cimprich 2014; Brambati et al. 2015). Transcription of long human genes over 800kb in length can take longer than one full cell cycle. In such instances, transcription starts during G2/M of one cell cycle and proceeds throughout an entire cell cycle, terminating in the subsequent G1 or early S phase of the next cell cycle. Replication through long genes that require longer than one cell cycle to complete transcription increases the chances of replication-transcription collision and R-loop formation at these loci (Helmrich et al. 2011). The failure to resolve R-loops causes DNA damage and chromosome breakage (Stirling et al. 2012).

It is thought that collisions between transcription and replication machinery may be one of the sources of replication stress seen in cancer cells that overexpress or have constitutively activated oncogenes. In such instances it is likely that elevated transcription levels due to excessive oncogene expression increases the chances of collision with an oncoming replication fork. Furthermore, the increase in replication caused by oncogene overexpression may result in a depletion of the nucleotide pool, hindering replication fork progression, thereby increasing genome instability (Bester et al. 2011). Whilst the exact causes of oncogene-induced replication stress are not known, the appearance of S-phase-specific ssDNA foci thought to be due to polymerase-helicase uncoupling, and the colocalization of markers of DNA damage with PCNA staining following Cyclin E overexpression, implicate aberrant oncogene expression in replication-specific DNA damage (Bartkova et al. 2006).

DNA breaks formed at genomic regions in which there is frequent concomitant replication and transcription such as in the long genes described above often map to common fragile sites (CFSs) (Helmrich et al. 2006; Helmrich et al. 2011). CFSs can be defined as hotspots for replication stress-induced DNA breaks seen in a majority of cells. To date, over 200 CFSs have been identified in human lymphocytes, ranging in size from

just under 1 Mb to over 10 Mb (Mazouzi et al. 2014; Bhowmick & Hickson 2017). CFSs can be reproducibly induced in the presence of agents that specifically impede replication such as aphidicolin, an inhibitor of replicative DNA polymerases (Glover et al. 1984; Arlt et al. 2006) and hydroxyurea (HU), an inhibitor of ribonucleotide reductase (Lundin et al. 2002). CFSs are often found around genomic loci containing AT-rich stretches that affect helix flexibility and those regions capable of forming secondary structure described above (Arlt et al. 2006; Bhowmick & Hickson 2017).

CFS expression also correlates with regions of highly condensed chromatin, marked by histone hypoacetylation. It is thought that the increased compaction of DNA at these points may be inhibitory to the establishment and activation of replication complexes at sites that otherwise may serve as replication origins and that the paucity of replication in these areas may be the source of replication stress (Jiang et al. 2009; Ozer & Hickson 2018). Similarly, human cancers often have elevated mutation rates surrounding regions of heterochromatic histone modifications (e.g. H3K9me3 and H4K20me3) although whether the high incidence of DSBs in these regions and in regions of hypoacetylation is due to the repressive nature of the chromatin landscape on replication or the inhibitory effect of heterochromatin on DNA repair remains to be elucidated (Schuster-Böckler & Lehner 2012; Tubbs & Nussenzweig 2017).

1.3.2 Preventing replication stress

Cells have evolved several mechanisms to help prevent the occurrence of events that may lead to replication stress and to resolve replication stress when it arises. As described in section 1.3.1, clashes between replication and transcription machinery are common sources of replication stress in eukaryotes. Transcription and replication are often temporally distinct, with the majority of transcription occurring during G1-phase of the cell cycle and replication occurring during S-phase (Bertoli, Skotheim, et al. 2013). This is principally regulated by the E2F family of transcription factors. Before the onset of S-phase, E2F6 transcriptional repressors replace activating E2Fs at promoters to inhibit transcription during DNA replication. In response to replication stress, the checkpoint kinase Chk1 phosphorylates E2F6 causing E2F6 to dissociate from promoters

and to be replaced by activating E2F transcription factors to activate transcription of DNA repair and nucleotide synthesis genes (Bertoli, Klier, et al. 2013; Bertoli et al. 2016).

Transcription-replication clashes often result in the formation of R-loops. It is therefore important that cells have pathways in place to suppress R-loop formation and to remove R-loops once formed. One of the principle mechanisms by which R-loops are repaired is RNase H1-mediated degradation of the RNA in a DNA:RNA hybrid in a process requiring replication protein A (RPA) which recognizes ssDNA and is thought to recruit RNase H1 to R-loops (Parajuli et al. 2017; Nguyen et al. 2017). Although less well-understood, DNA Topoisomerase I (Top1) also plays a role in the processing of R-loops, particularly in the context of genes that are difficult to transcribe (Tuduri et al. 2009; Manzo et al. 2018).

Transcription and replication can be spatially separated to avoid replication stress caused by transcription-replication collisions in highly transcribed genes such as eukaryotic ribosomal DNA (rDNA). This spatial separation prevents replication over actively transcribed templates and is achieved by the presence of sequence-specific replication fork barriers (RFBs) at the 3' ends of pre-ribosomal RNA coding regions that are tightly bound by proteins to arrest replication fork progression. This prevents head-on collisions between the incoming replication fork and transcription machinery travelling in the opposite direction. Replication of these regions is then completed by a replication fork originating from a region flanking the rDNA repeat travelling co-directionally with transcription (Brewer & Fangman 1988).

The majority of replication fork barriers discovered in yeasts and higher eukaryotes are polar, i.e. they stop replication from one direction only, however some human rDNA RFBs are bidirectional causing arrest of incoming replication forks from either direction (Akamatsu & Kobayashi 2015; Gadaleta & Noguchi 2017).

1.3.3 Restarting stalled forks

A failure to restart and process replication forks that stall due to replication stress-causing factors is a major source of genetic instability. Stalled forks can be defined as

replication forks whose progress is impeded by the presence of an inhibitory factor but whose activity can be resumed following removal of the impediment without further intervention (Lambert & Carr 2013).

Some stalled forks are restarted by repriming downstream of the lesion, leaving a stretch of single stranded DNA that is subsequently repaired by postreplicative repair (PRR) mechanisms (Yeeles et al. 2013). Stalled forks may also be rescued by a converging replication fork that may have resulted from the programmed firing of dormant origins in response to replication stress (Woodward et al. 2006; Ge et al. 2007). Stalled forks may also activate pathways that allow translesion polymerases to replicate over the template containing the damaged bases in a process termed translesion synthesis (TLS). Such polymerases have larger active sites than the major replicative polymerases that allow for the accommodation of the helix distortions caused by the damaged bases, but the larger active site also reduces the fidelity of translesion polymerases when replicating undamaged bases (Hubscher & Maga 2011; Zahn et al. 2011).

1.3.4 The Intra-S phase response to replication stress

One of the key features in triggering the DNA damage response to replication stress is the generation of single-stranded DNA (ssDNA) caused by the continued unwinding of duplex DNA by the MCM replicative helicase having become uncoupled from the replicative polymerase (Byun et al. 2005). ssDNA is recognized, and bound by RPA which recruits a key protein kinase, ataxia telangiectasia mutated and Rad3-related (ATR), to sites of replication stress via interactions with ATR-interacting protein (Zou & Elledge 2003). Once activated at sites of ssDNA following phosphorylation by TopBP1, ATR phosphorylates hundreds of targets, the most important of which is its principle downstream effector, the checkpoint kinase Chk1 which can act locally and globally to help overcome replication stress (Mazouzi et al. 2014).

Following HU-induced replication stress, the *S. cerevisiae* Chk1 homologue, Rad53 acts directly on the CMG helicase to prevent further unwinding of duplex DNA. This prevents CMG from advancing away from the site of DNA incorporation, thereby preventing the formation of excess ssDNA (Gan et al. 2017).

Chk1 plays an important role in inhibiting the global late firing of origins in response to replication stress whilst at the same time allowing the firing of dormant origins in the vicinity of stalled replication forks under low replication stress conditions (Ge et al. 2007; Ge & Blow 2010). The suppression of global origin firing by ATR-Chk1 signalling protects cells by ensuring that the level of ssDNA generated by DNA replication does not exceed that which can be bound by the cellular RPA pool, since RPA depletion causes widespread DNA breaks (Toledo et al. 2013).

A reduction in global origin firing in favour of rescuing replication in the vicinity of stalled replication forks is also mediated by the activity of ATR on a member of the Fanconi anaemia DNA repair pathway, FANCI, which along with its partner FANCD2 interacts with the replisome via interactions with the MCM helicase (Lossaint et al. 2013). Under normal growth FANCI plays a role in activating DDK, which in turn promotes origin firing as described in section 1.2.4. However, under conditions of replication stress, ATR phosphorylates FANCI, which causes a shift in the pool of FANCI from dormant origins to stalled replication forks where it promotes replication fork protection, repair and restart (Chen et al. 2015).

1.3.5 Restarting collapsed replication forks by homologous recombination

If a stalled fork is not correctly stabilized following the onset of replication stress and requires the involvement of a replication restart mechanism to resume, it is referred to as a collapsed fork (Lambert & Carr 2013). This collapse may cause the dissociation of some components of the replisome (Katou et al. 2003; Lucca et al. 2004; Cobb et al. 2005), although it is more likely that the replisome remains intact although the replicative helicase may continue to separate the DNA duplex ahead of the collapsed fork (Lopes et al. 2006; De Piccoli et al. 2012; Dungrawala et al. 2015).

One of the main mechanisms by which cells protect and restart collapsed replication forks is the use of homologous recombination (HR) (Carr & Lambert 2013). In human cells, this response is mediated by BRCA2 which binds to RPA-coated ssDNA and promotes the displacement of RPA by the RAD51 recombinase to establish the

nucleoprotein filament required to initiate the search for homology (Jensen et al. 2010; Thorslund et al. 2010). Collapsed replication forks may be restarted by homologous recombination either with or without DSB formation in mammalian cells (Lundin et al. 2002), although this process can be error-prone if the incorrect template for restart is selected (Carvalho et al. 2015; Zou et al. 2017). Similar findings in yeast (Mizuno et al. 2009; Mizuno et al. 2013) suggest that error-prone HR-mediated fork restart mechanisms are conserved throughout evolution.

1.4 Ribonucleotides in genomic DNA

1.4.1 Introduction to genomic ribonucleotides

The maintenance and heritability of genetic information is dependent upon the highly accurate duplication of genomic DNA (gDNA) prior to cell division. The use of deoxyribonucleotides (dNTPs) as the building blocks for genetic storage offers an advantage over ribonucleotide (rNTP)-based genomes due to the increased stability of dNTP polymers. Polynucleotide chains are linked via chemically stable phosphodiester bonds that form between the 5'-phosphate and 3'-hydroxyl of adjacent nucleotides. Phosphodiester bonds involving ribonucleotides are more susceptible to hydrolysis than bonds between adjacent dNTPs due to the presence of the 2'-hydroxyl group on the ribose sugar moiety of rNTPs (Li & Breaker 1999). It is therefore important that cells have mechanisms to detect and repair any misincorporated ribonucleotides. The following sections will discuss the source of genomic ribonucleotides and the pathways that cells have evolved to remove them.

1.4.2 Sources of genomic ribonucleotide incorporation

Ribonucleotides may be present in the genome as a result of programmed inclusion or as the result of erroneous incorporation by DNA polymerases, in addition to stochastic events such as the spontaneous hydrolysis of the deoxyribose sugar in DNA to ribose by free radicals to generate an embedded genomic ribonucleotide (Randerath et al. 1992).

As described in section 1.2, the major replicative polymerases, Polδ and Polε, are incapable of initiating *de novo* replication events and rely upon the primase activity of Polα. Polα initiates nucleotide chain extension on both the leading strand and lagging strands by inserting ~8-10 nucleotide long RNA primers. Due to the requirement for priming to initiate replication at the start of every Okazaki fragment (OF) on the lagging strand, eukaryotic cells incorporate an RNA primer every ~200 bases, meaning that replication of the entire 3×10^9 human genome requires the incorporation of ~1.5×10^8 ribonucleotides into the nascent DNA (Williams & Kunkel 2014). RNA primers are subsequently removed from gDNA during OF maturation, a process similar to ribonucleotide excision repair (described in section 1.4.3) but without the requirement for RNase H2 (Qiu et al. 1999). It is likely that this process involves strand displacement synthesis by Pol δ followed by cleavage of the resultant flap by FEN1 (Garg et al. 2004) or in the absence of FEN1 by DNA2 (Balakrishnan & Bambara 2013).

Replicative DNA polymerases have strong selectivity for deoxyribonucleotides over ribonucleotides, largely mediated by the presence of well-conserved steric gate residues at their active sites that clash with the 2'hydroxyl group on the ribose moiety of rNTPs (Brown & Suo 2011). In spite of this structural discrimination, several studies have shown that *in vivo*, and in *in vitro* studies under physiologically relevant conditions, eukaryotic replicative polymerases incorporate rNTPs at a rate of 1 rNTP/650 dNTPs (*S. cerevisiae* Polα); 1/5000 (*S. cerevisiae* Polδ); 1/1250 (*S. cerevisiae* Polε); 1/500 (human Polα); 1/20000 (bovine Polα) and 1/7600 (mean for all replicative polymerases in mouse) (Kuchta et al. 1992; Richardson et al. 2000; Nick McElhinny, Watts, et al. 2010; Reijns et al. 2012). The rate of rNTP incorporation by human Polδ and Polε (1/2000 and 1/1500 respectively) indicate that it is possible that replication of mammalian genomes results in the stable incorporation (not including RNA primers) of more than 3×10^6 rNTPs (Goksenin et al. 2012; Clausen et al. 2013; Williams & Kunkel 2014). That polymerases incorporate ribonucleotides despite possessing structural barriers to increase discrimination for the correct sugar can possibly be explained by the fact that the cellular pool of rNTPs is greater than that of dNTPs in both yeast (Nick McElhinny, Watts, et al. 2010) and mammalian cells (Traut 1994; Ferraro et al. 2010).

1.4.3 Removing misincorporated ribonucleotides from genomic DNA

RNA:DNA hybrids are formed during several cellular process including DNA replication (Nick McElhinny, Watts, et al. 2010; Clausen et al. 2013), transcription (Garcia-Benitez et al. 2017) and telomere elongation (Forstemann & Lingner 2005). The majority of repair of such RNA:DNA hybrids is initiated by the activity of a family of enzymes called ribonucleases H (RNases H). In both prokaryotes and eukaryotes, two classes of RNase H have been described, type 1 and type 2 (Cerritelli & Crouch 2009). RNases H1 and H2 differ in their overall structure and their substrate specificity. RNase H1 is monomeric and requires the 2'-OH groups of a stretch of at least 4 ribonucleotides for efficient catalytic activity (Nowotny et al. 2007). The predominant RNase activity in mammalian cells is derived from RNase H2 (Busen 1980), a heterotrimer consisting of three subunits: the catalytic subunit, RNase H2A, and two accessory subunits, RNase H2B and RNase H2C (Jeong et al. 2004; Crow, Leitch, et al. 2006; Shaban et al. 2010; Reijns et al. 2011). RNase H2 has been shown to act on single genomic ribonucleotides (Eder et al. 1993; Rychlik et al. 2010) and on longer stretches of genomic RNA (Chon et al. 2009). Since the focus of this thesis is on the misincorporation of ribonucleotides by replicative polymerases during DNA replication, and since the chances of polymerases incorporating two or more successive ribonucleotide are presumed to be very rare, this section will concentrate on the catalytic activity of RNase H2 since it is the enzyme responsible for the removal of single ribonucleotides embedded in gDNA. Further details relating to the structure and function of RNase H2 will be discussed in more detail in section 1.5.

The principle pathway by which single ribonucleotides misincorporated by polymerases are removed is called ribonucleotide excision repair (RER). RER starts with the recognition of the embedded ribonucleotide by RNase H2 which then cleaves the 5'-phosphodiester bond of the ribonucleotide (Eder et al. 1993). The resultant nick allows for strand displacement synthesis by Polδ (and less efficiently by Polε) generating a flap which acts as a substrate for flap removal by FEN1 and/or DNA2 with DNA Ligase 1 (LIG1) sealing the gap between the newly synthesized and the original DNA strands (Sparks et al. 2012) (Figure 1-3). The requirement for RNase H2 and FEN1 in RER were first

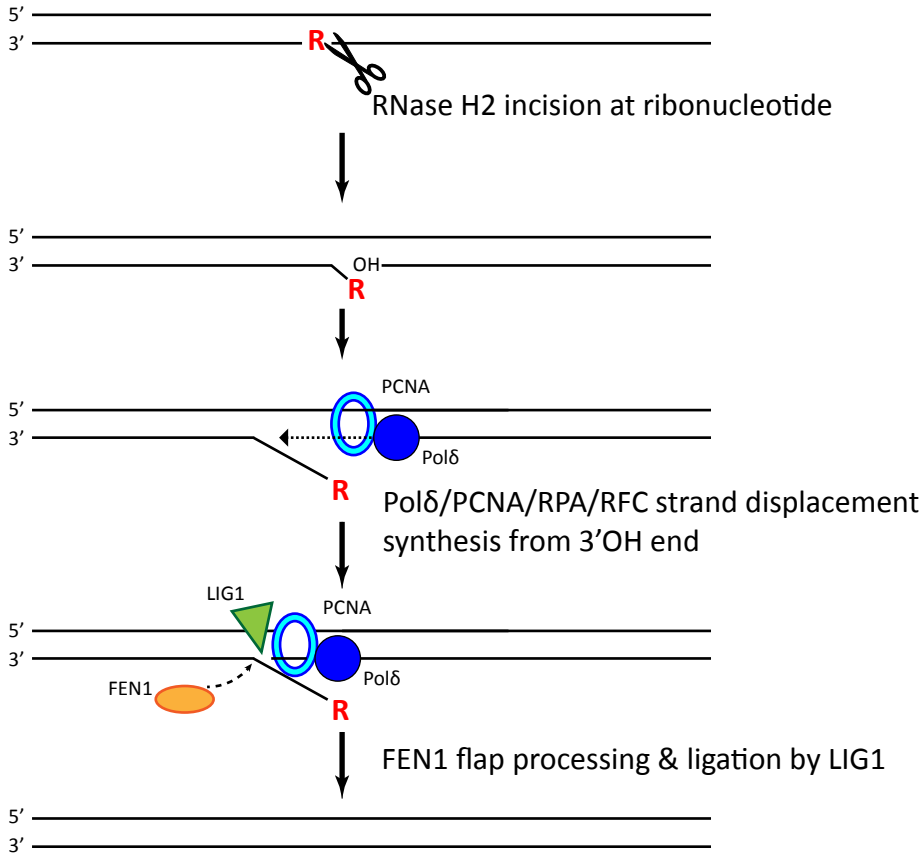


Figure 1-3 Ribonucleotide excision repair
A genomic ribonucleotide (red R) is recognised by RNase H2 which makes an incision on the 5’ side of the ribonucleotide leaving a 3’-OH end. Strand displacement synthesis requiring Polδ, PCNA, the clamp loader, RFC and RPA creates a ribo-terminating flap which is a substrate for FEN1. FEN1 cleaves the flap and the remaining gap is sealed by DNA ligase 1. Adapted from Klein (2017).

identified in a study in *S. cerevisiae* in which deletion RNase H2A resulted in decreased incision at ribonucleotides and deletion of RAD27/FEN1 decreased ribonucleotide liberation (Rydberg & Game 2002).

RER can be reconstituted *in vitro* using RNase H2, FEN1, Polδ and LIG1 (Sparks et al. 2012). All four enzymes have been shown to interact with PCNA, raising the possibility that the RER machinery may travel with the replication fork. Interaction with PCNA stimulates the endonuclease activity of FEN1 (Li et al. 1995) but has no effect on the ligation efficiency of LIG1 (Montecucco et al. 1998). RNase H2 contains a motif for interaction with PCNA termed a PCNA-interacting protein (PIP) box which is located in RNase H2B (Chon et al. 2009), the presence of which is essential for the colocalization of RNase H2 with active replisomes in mammalian cells (Bubeck et al. 2011). Whilst PCNA greatly stimulates RER *in vitro*, the RNase H2 PIP-box is dispensable for catalysis *in vitro* (Sparks et al. 2012). Whilst these latter results may have been an artefact of the experimental design, they perhaps indicate that RNase H2 may be able to effect RER without interacting directly with PCNA. It seems that a repair system in which misincorporated ribonucleotides could be repaired by RER after the replisome has advanced away from the site of misincorporation would be favoured by evolution, rather than relying on damage to be repaired concomitantly with replication. Indeed, genomic ribonucleotides have been found to act as a marker of newly synthesized DNA to allow for strand discrimination in mismatch repair (MMR), and RNase H2 activity at such sites may generate the nicks required to initiate MMR (Ghodgaonkar et al. 2013; Lujan et al. 2013). These findings indicate that RER probably occurs independently of the interaction of RNase H2 with PCNA at the replication fork.

Recent work in *S. cerevisiae* has indicated a role for Topoisomerase 1 (Top1) in the repair of single embedded genomic ribonucleotides. In yeast, the absence of RNase H2 results in an increase in Top1-dependent 2-5 bp deletions (N. Kim et al. 2011). Top1 activity at ribonucleotides generates single strand breaks with 2',3'-cyclic phosphate termini. Top1 also nicks two nucleotides upstream of the cleaved ribonucleotide, generating a covalently linked Top1-DNA complex with a 2 nucleotide gap (Sparks & Burgers 2015). Subsequent Top1-mediated cleavage on the opposite strand generates irreversible

double strand breaks which can be repaired in an error-free manner by homologous recombination (Huang et al. 2017). The Top1-DNA complex may be repaired before double strand break formation by a process requiring Exo1 and the Sgs1 and Srs2 helicases (Niu et al. 2016). Interestingly, having generated a 2',3'-cyclic phosphate, Top1 is able to reverse this nick formation to provide once again a suitable substrate for conventional RNase H2-mediated repair (Sparks & Burgers 2015)

1.5 Ribonuclease H2

1.5.1 Ribonuclease H2 structure

As described in section 1.4.3, eukaryotic RNase H2 is a trimer consisting three subunits: one catalytic subunit, RNase H2A and two accessory subunits, RNase H2B and RNase H2C (Figure 1-4 A). In humans the three subunits are composed of 299, 312 and 164 amino acids respectively. The core domains of the three subunits stack in a linear manner, with RNase H2C flanked by the other two subunits. The C-terminal 17 residues of RNase H2A (283-299) form an α/β motif that forms part of a central β-barrel at the core of the holoenzyme that is required for binding to the RNase H2BC heterodimer. Mutational peptide array analysis revealed Tyr-287, Phe-288, Arg-291, Leu-293 and Ala-296 as the key residues required to mediate this interaction. Indeed, deletion of residues 248-299 of the C-terminal extension of RNase H2A not only prevented trimer formation following expression in *E. coli*, but also abolished catalytic activity when mixed with purified RNase H2BC *in vitro* (Reijns et al. 2011). Deletion of the N-terminal 23 residues of RNase H2A also prevents trimer formation following expression of all three subunits in *E. coli* (Figiel et al. 2011). The requirement for a stable interaction between all three subunits for catalytic activity is suggested in studies using purified human and *S. cerevisiae* RNase H2. In these studies it was found that all three subunits are necessary for catalytic activity and that RNase H2A alone is insufficient for catalysis (Jeong et al. 2004; Chon et al. 2009; Figiel et al. 2011).

The discovery that RNase H2B and RNase H2C form stable soluble dimers after expression in *E. coli*, and that excess RNase H2C accumulated in the insoluble fraction led Chon and colleagues to postulate that the RNase H2BC heterodimer may serve as a

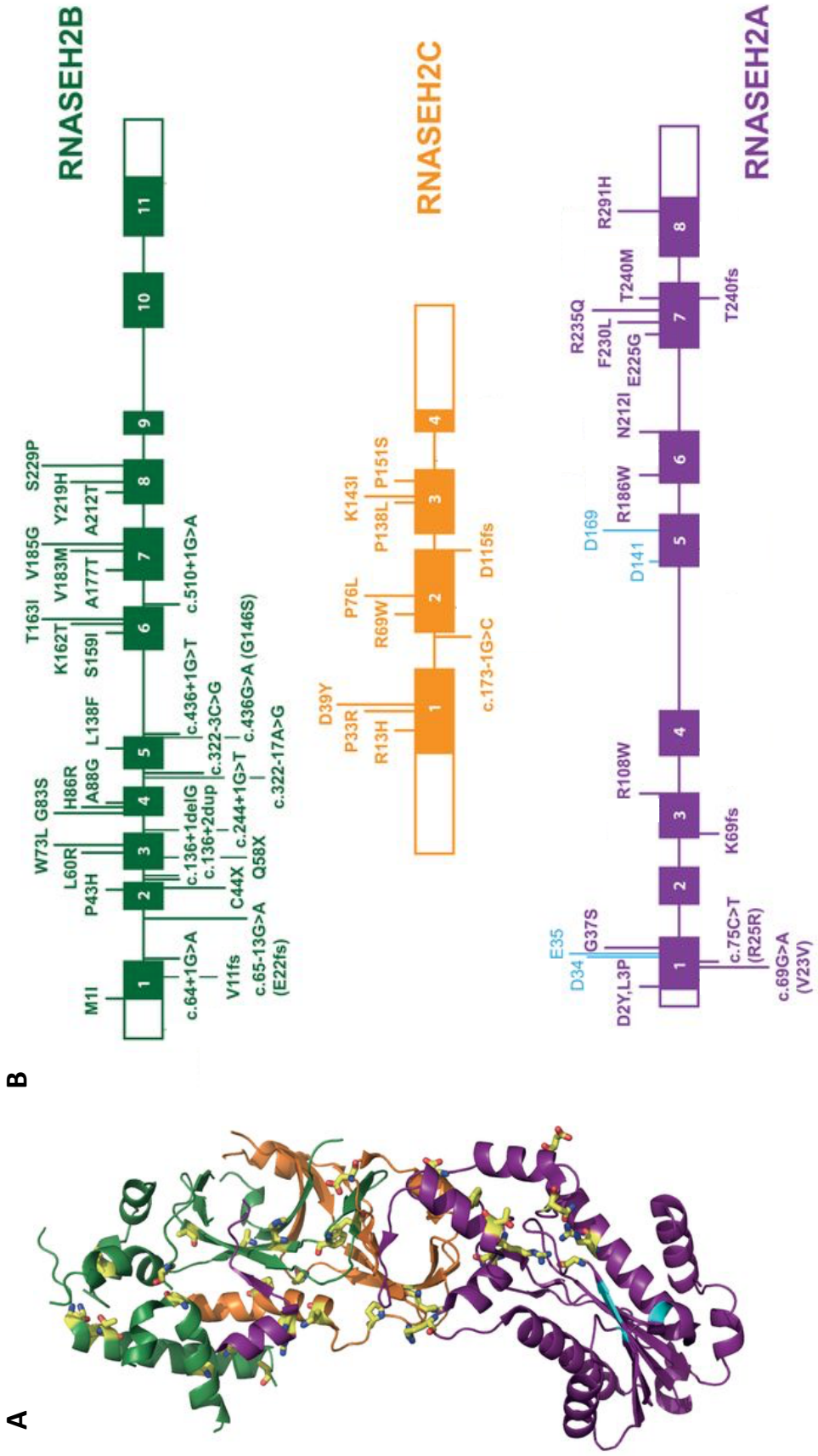

A
B
RNASEH2B
M1I
c.64+1G>A
V11fs
c.65-13G>A
(E22fs)
P43H
L60R
W73L
G83S
C44X
Q58X
c.136+1delG
c.136+2dup
c.244+1G>T
c.322-17A>G
H86R
A88G
L138F
c.322-3C>G
c.436G>A (G146S)
T163I
K162T
S159I
A177T
V183M
V185G
c.436+1G>T
c.510+1G>A
Y219H
A212T
S229P
RNASEH2C
D39Y
P33R
R13H
P76L
R69W
K143I
P138L
P151S
c.173-1G>C
D115fs
RNASEH2A
E35
D34
G37S
D2Y,L3P
c.75C>T
(R25R)
c.69G>A
(V23V)
K69fs
R108W
D169
D141
R186W
N212I
R235Q
F230L
E225G
T240M
T240fs
R291H

Figure 1-4 RNase H2 structure and AGS disease mutations
A. Crystal structure of human RNase H2 with RNase H2A shown in purple, RNase H2B shown in green and RNase H2C shown in green. The side chains of residues mutated in AGS are shown as yellow sticks. Catalytic site residues in RNase H2A are shown in cyan. B. Schematic diagrams of the *RNASEH2A*, *RNASEH2B* and *RNASEH2C* genes, coloured as in A showing the locations of known mutations AGS disease mutations. Filled boxes indicate exons, horizontal lines indicated introns, open boxes indicate untranslated regions. Missense mutations are indicated above each gene and splice site, frameshift and nonsense mutations are indicated below each gene. Mutations in catalytic site residues within RNase H2A are shown in cyan. Figures from Reijns & Jackson (2014).

nucleation site for trimer formation (Chon et al. 2009). Indeed, structural studies of RNase H2 showed an important role for RNase H2C in mediating the interactions between the other two subunits via a C-terminal kinked helix formed by residues 143-160 which interact with both RNase H2A and RNase H2B. Furthermore, two central β-strands in RNase H2C (residues 118-126 and 128-134) form part of two β-sheets found within a key central β-barrel architecture that mediates interactions between all three subunits. Accordingly, the C-terminal residues 117-164 of RNase H2C are essential for trimer formation and catalytic activity (Reijns et al. 2011). Interactions between RNase H2A and the RNase H2BC heterodimer appear to be mediated principally by interactions between an α-helix and three loops within RNase H2A and four regions within RNase H2C, suggesting an important role for RNase H2C in RNase H2 complex stability (Figiel et al. 2011).

1.5.2 RNase H2 and disease

RNase H2 is presumed to be essential in human cells expressing normal levels of p53 (this study) and loss of RNase H2 is embryonic lethal in mice (Hiller et al. 2012a; Reijns et al. 2012). Hypomorphic mutations in RNase H2 cause Aicardi-Goutières Syndrome (AGS), a rare inherited autoimmune disease resulting in upregulation of type 1 interferon signaling (Reijns & Jackson 2014; Crow et al. 2015). Although AGS is a monogenic disorder, AGS patients have been described with mutations in seven different genes: RNASEH2A, RNASEH2B, RNASEH2C (Crow, Leitch, et al. 2006), TREX1 (Crow, Hayward, et al. 2006), SAMHD1 (Rice et al. 2009), ADAR1 (Rice et al. 2012) and IFIH1 (Rice et al. 2014), all of which are involved in nucleic acid metabolism or signaling. AGS is a phenotypically variable disease with some cases being diagnosed several months after birth whilst other cases are evident at birth and can result in premature death (Rice et al. 2007).

Defects in AGS-causing genes are thought to lead to an intracellular accumulation of aberrant nucleic acid species. The autoimmune response seen in AGS is triggered by the sensing of these nucleic acids by cyclic GMP-AMP synthase (cGAS) and its adaptor STING which lead to an upregulation of transcription of interferon-stimulated genes

(Mackenzie et al. 2016). TREX1 is a 3' DNA exonuclease that functions within the host immune response and prevents the activation of cGAS by degrading reverse transcribed DNA from endogenous retroelements (Li et al. 2017). SAMHD1 is thought to prevent activation of the immune response by preventing the interaction of innate immune sensors with intracellular nucleic acid species (Martinez-Lopez et al. 2018). ADAR1 is thought to prevent the accumulation of double stranded RNA (dsRNA) transcribed from genomic Alu repeat elements (Rice et al. 2012). Whilst AGS is thought to be caused by loss of function mutations in TREX1, SAMHD1, ADAR1 and the three RNase H2 subunits, AGS can be caused by a gain of function mutation in IFIH1. IFIH1 is a cytosolic dsRNA sensor and gain of function mutations in this gene cause IFIH1 to bind more avidly to dsRNA resulting in a robust upregulation in interferon signaling (Rice et al. 2014).

Mutations in subunits of RNase H2 account for over 50% of reported cases of AGS, with mutations in RNase H2B accounting for 68% of those AGS patients with RNase H2 mutations (RNase H2A: 9%; RNase H2B: 68%; RNase H2C: 23%) (Crow et al. 2015). Several of these mutations have been mapped to the structures of human RNase H2 (Figure 1-4 B) and are predicted to affect RNase H2 activity in three ways: i) by decreasing stability of the trimer due to mutations either within the hydrophobic core or in residues at the interface between adjacent subunits; ii) by decreasing the efficiency of substrate hydrolysis due to mutations within the enzyme active site; iii) by decreasing the ability of RNase H2 to interact with putative protein partners due to mutations on the surface of the trimer (Figiel et al. 2011; Reijns et al. 2011).

1.6 Polymerase Usage Sequencing

The assignment of Polε and Polδ as the major replicative polymerases of the leading and lagging strand respectively was carried out using mutant alleles of each polymerase that introduced a biased mutation spectrum in *S. cerevisiae* (Pursell et al. 2007; Nick McElhinny et al. 2008). Further work using an equivalent mutator allele of Polδ found that Polδ is also responsible for the majority of replication on the lagging strand in *S. pombe*; in the same study Polε was assigned as the leading strand

polymerase by using a Polε mutant allele that incorporated elevated levels of ribonucleotides into DNA upon replication in the absence of RNase H2 (Miyabe et al. 2011). Human tumour cells containing mutations in the proofreading exonuclease domain of Polε exhibit a strand-specific mutational bias (Shinbrot et al. 2014) suggesting that the role of Polε as the major leading strand polymerase and the division of labour at the eukaryotic replication fork are evolutionarily conserved.

The Carr laboratory has developed a system termed Polymerase Usage Sequencing (Pu-seq) to determine if this division of labour is true genome-wide in *S. pombe* (Daigaku et al. 2015). The Pu-seq system uses strains of *S. pombe* containing mutations in either Polδ or Polε that incorporate elevated levels of ribonucleotides upon genome replication. In addition, the catalytic subunit of RNase H2 (RNase H2A), encoded by the gene *rnh201*, is knocked out in these strains, preventing the repair of misincorporated ribonucleotides by ribonucleotide excision repair (RER). In the absence of RER, genomic ribonucleotides act as a molecular footprint of the mutant polymerases.

The presence of a ribonucleotide in genomic DNA renders the DNA duplex at that point more susceptible to alkaline hydrolysis. By treating gDNA extracted from Pu-seq strains of *S. pombe* with alkali and using a deep-sequencing methodology on gDNA extracted from ~10^8 cells to map the 5'-end of ribo-terminating fragments to genomic loci, the relative contributions of each polymerase to genome-wide replication can be determined (Figure 1-5A). The Pu-seq method identifies the location of origins, marked by the point at which synthesis on one strand switches from being carried out predominantly by one of the two replicative polymerases to being carried out predominantly by the other (with a concomitant, but inverse, switch seen on the opposite strand) (Figure 1-5). Furthermore, the Pu-seq method offers a direct measure of origin efficiency, as well as an indirect inferral of replication timing (Figure 1-5C) (Daigaku et al. 2015; Keszthelyi et al. 2015). This is in contrast to other methods of monitoring genome-wide replication dynamics that generally produce replication timing data directly, from which origin efficiency is inferred (occasionally with not much accuracy) (Peace et al. 2016; Marchal et al. 2018). There exists therefore a certain complementarity between our Pu-seq system and other genome-wide systems.

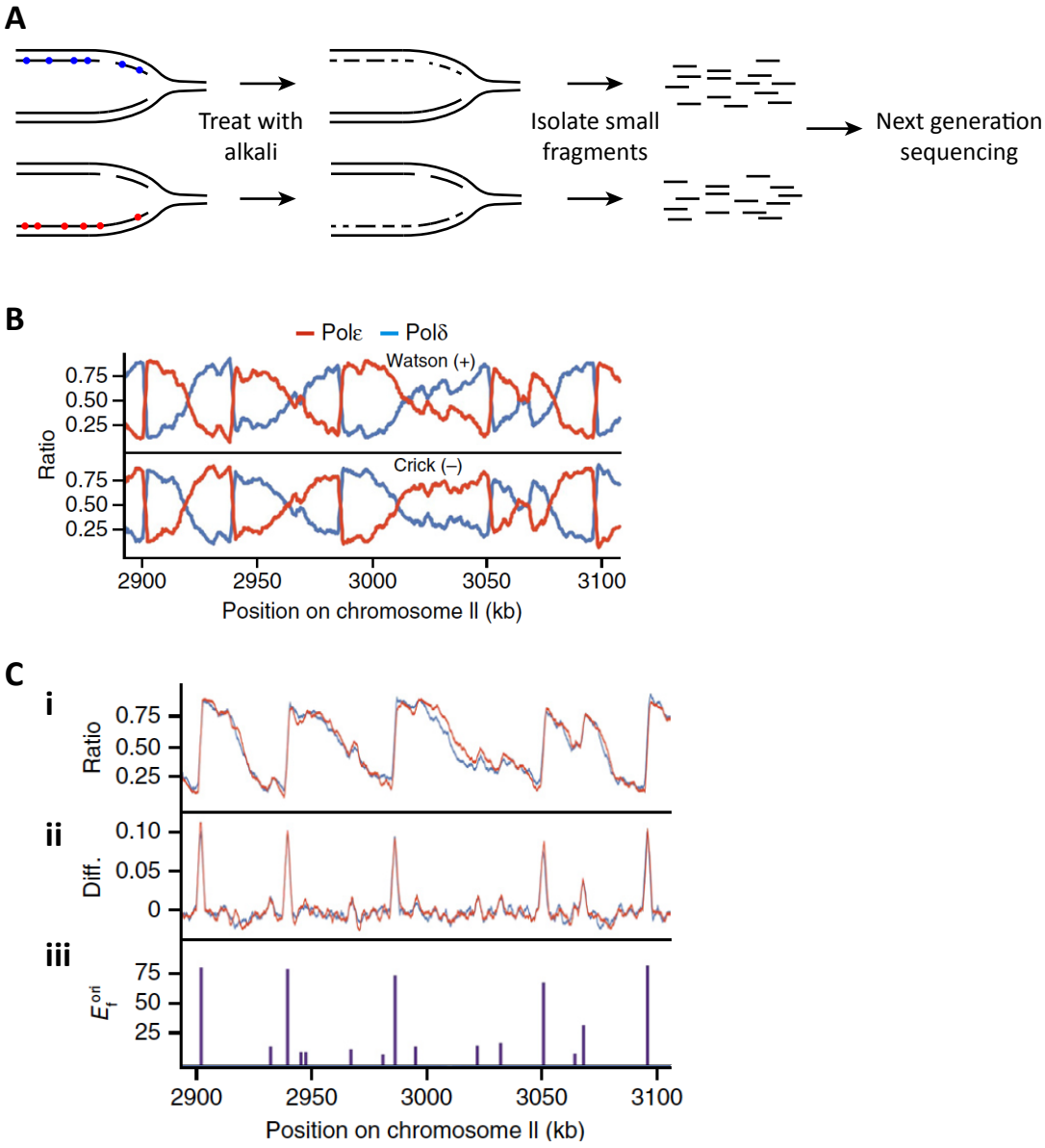


Figure 1-5 The principle of polymerase usage sequencing

A. Strains of *S. pombe* containing riboincorporating mutant polymerase delta or polymerase epsilon and lacking RNase H2 are allowed to go through replication. Treatment of genomic DNA from these strains is treated with alkali, hydrolysing the DNA backbone at sites of misincorporated ribonucleotides (blue dots, Polδ; red dots, Polε). The fragmented DNA from ~10^8 cells is pooled and the sites of ribonucleotide incorporation determined from deep sequencing. B. Smoothed data showing the mean proportion of synthesis by Polδ (blue) and Polε (red) on each strand along a 200 kb stretch of *S. pombe* chromosome II. Origins of replication are marked by the point at which synthesis on one strand switches from being carried out by mostly Polδ to mostly Polε (or vice-versa) with a concommitant opposing switch seen at the same locus on the oppsite strand. C. Representative data from a Pu-seq experiment. i) Mean Polε usage on the Watson strand (blue) and Polδ on the Crick strand (red). ii) Differential of the polymerase usage plots in i. iii) Origin efficiencies calculated from Pu-seq data.

The Pu-seq method therefore allows for in-depth study of replication dynamics genome-wide in a variety of genetic backgrounds.

1.7 Genome editing using CRISPR-Cas9

1.7.1 A historical perspective of mammalian genome engineering

An early breakthrough in the field of genome editing was made in the mid 1980s with the discovery that targeted gene disruption could be effected by the homologous recombination-mediated introduction of exogenous DNA into mammalian genomes (Smithies et al. 1985; Thomas et al. 1986; Capecchi 1989). Despite the benefits of this system it suffered from low rates of integration and the risk of random integration at unintended sites elsewhere in the genome (Adli 2018).

The discovery that the efficiency of integration of exogenous DNA was greatly increased by the induction of a restriction endonuclease-induced double strand break at the target site (Rouet et al. 1994) provided a new breakthrough in genome editing. This method is however limited by the paucity of rare enzyme cutting sites within eukaryotic genomes. The discovery that precise genomic DNA mutagenesis could be achieved using novel hybrid restriction endonucleases termed zinc finger nucleases (ZFNs) helped to overcome this problem. ZFNs are engineered by fusing sequence-specific zinc finger DNA-binding domains with dimerization-dependent, sequence-independent FokI nuclease domains (Kim et al. 1996). DSB induction using ZFNs requires the engineering of two separate ZFNs that recognize adjacent sequences on opposite strands, separated by a 5-7bp spacer. Binding of the two ZFNs to their cognate sequences induces dimerization and activation of the two FokI nuclease domains and subsequent DSB formation at the target site (Bibikova et al. 2001).

ZFN technology was used for several years as the principal method genome editing in higher eukaryotes, however the technology was not widely adopted due to the inherent difficulties of designing zinc finger domains with a high enough level of specificity (Doudna & Charpentier 2014). In 2009 a new technique for genome editing was discovered using transcription activator-like (TAL) effectors derived from plant

pathogenic bacteria. TALs bind to specific DNA sequences via a series of around 34 amino acid repeats, with each repeat forming specific interactions with single base pairs. As with zinc finger domains, these proteins are fused with FokI nuclease domains to generate sequence-specific nucleases, termed TAL effector nucleases (TALENs) (Boch et al. 2009; Christian et al. 2010).

1.7.2 CRISPR-Cas9 as a genome editing tool in human cells

Whilst TALENs benefit from greater specificity in comparison to ZFNs (Ul Ain et al. 2015) they retain the complications associated with ZFNs of having to engineer novel DNA binding subunits for each target. The solution to this problem arrived with the discovery of the mechanism of action of the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CRISPR-associated (CRISPR-Cas) immune defence system of *Streptococcus pyogenes* (Jinek et al. 2012) and the subsequent discovery that this system can be used to induce targeted double strand breaks in human cells (Cong et al. 2013; Jinek et al. 2013; Mali, Yang, et al. 2013).

Prokaryotic CRISPR loci are composed of several Cas genes arranged in one or more operons, flanked by an array of unique spacer regions derived from foreign DNA, separated by several identical repeat sequences (Terns & Terns 2011; Wiedenheft et al. 2012). The *Streptococcus pyogenes* CRISPR-Cas system belongs to the type II CRISPR-Cas bacterial adaptive immunity system and uses the Cas9 RNA-guided endonuclease (RGEN) to cleave target DNA (Gasiunas et al. 2012). Each of the three CRISPR-Cas systems (types I, II and III) differs in their mechanisms of RNA processing and interference (Makarova et al. 2006; Bhaya et al. 2011) but only the type II system of *S. pyogenes* will be described here.

During the adaptive phase of immunity acquisition, 20bp sequences of foreign DNA (protospacers) are incorporated into the CRISPR array between the identical repeat sequences. Protospacer identification is determined by the presence of a sequence termed the protospacer-adjacent motif (PAM) immediately 3'-wards of a 20bp sequence, usually NGG, although NAG PAMs can target a sequence for cleavage by Cas9

(Hsu et al. 2013; Jiang et al. 2013). During the immune response phase, a precursor CRISPR RNA (pre-crRNA) is transcribed from a spacer sequence complementary to the target foreign DNA, and its adjacent repeat sequence. A second RNA molecule, termed trans-activating CRISPR RNA (tracrRNA), is transcribed and base pairs with a complementary region within the repeat sequence of the pre-crRNA. This RNA duplex is processed by the double-stranded RNA-dependent RNase III in a process requiring Cas9 to yield a mature CRISPR RNA (crRNA) hybridized to the tracrRNA, the 3' portion of which interacts with Cas9 (Deltcheva et al. 2011) (Figure 1-6). The mature crRNA then directs Cas9 to its complementary sequence within target DNA, whereupon Cas9 cleaves both strands of the target DNA, typically 3 bp upstream from the PAM (Ran et al. 2015).

Successful genome editing in human cells can be achieved by targeting Cas9 with a chimeric RNA formed by fusing the target sequence-containing portion of the crRNA at with the dsRNA scaffold portion of the tracrRNA (Jinek et al. 2012). This chimeric RNA is termed a single guide RNA (sgRNA) or more simply, guide RNA (gRNA). This discovery carried great importance as it allowed the DNA cleavage mechanisms of the *S. pyogenes* CRISPR-Cas system to be reconstituted in such a way that it required only two components to target DSBs at any site within a genome with an upstream NGG PAM: a Cas9 protein and a single RNA molecule, both of which can be encoded on a single plasmid (Jinek et al. 2012; Cong et al. 2013).

Despite the advantages offered by the simplicity of programming unique gRNAs for target sequences, the system is not entirely perfect and suffers from variable levels of off-target activity. Cas9 can tolerate several mismatches between gRNA and DNA and as such is prone to cutting at undesired off-target locations. Mismatch sensitivity is dependent upon the number, distribution and position of mismatches, with mutations in the PAM-distal region more tolerated than those in the PAM-proximal region. The first 8-14 nucleotides at the PAM-proximal end of the gRNA form a key seed region in which mismatches are less tolerated (Hsu et al. 2013).

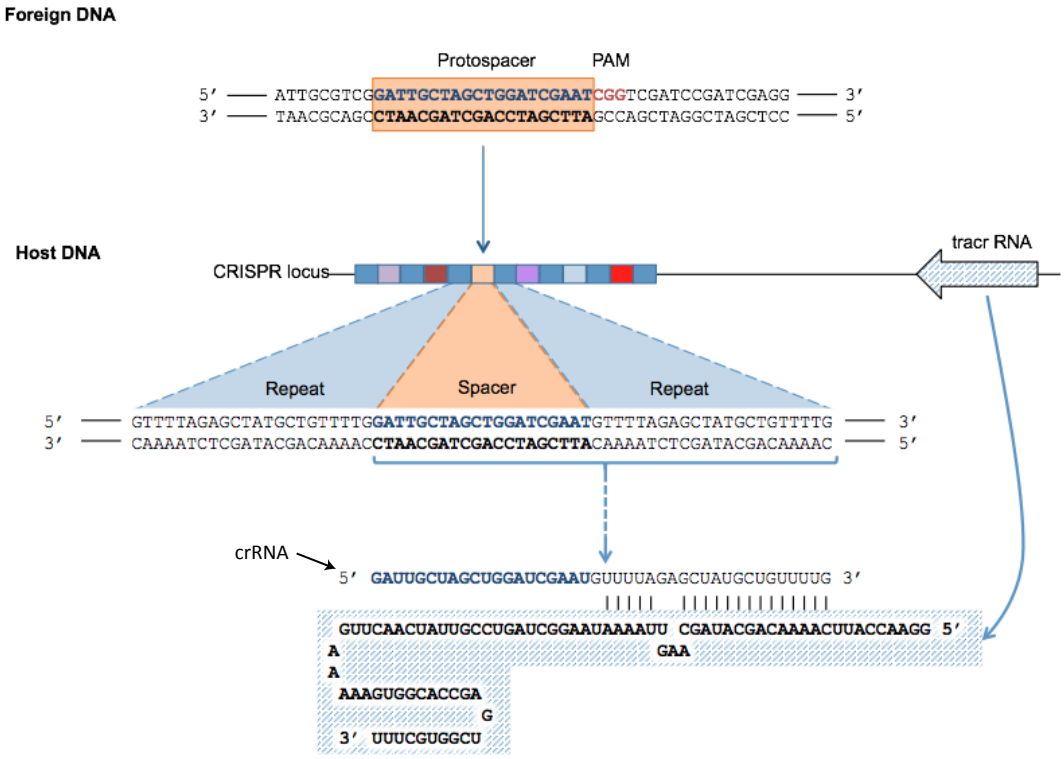


Figure 1-6 The CRISPR-Cas system of *S. pyogenes*
A 20 base pair section of foreign genomic DNA termed the protospacer is incorporated into the CRISPR locus in the *S. pyogenes* genome. The protospacer itself is marked by the presence of a NGG trinucleotide immediately downstream of the protospacer in the foreign genome. This NGG trinucleotide is termed a protospacer adjacent motif (PAM). Each 20 base pair protospacer is incorporated into the CRISPR array, separated by identical repeat sequences. Once incorporated, the 20 base pair sequence is termed a spacer. Each unique spacer is co-transcribed with its adjacent identical repeat sequence to form a crRNA which forms Watson-Crick base pairs with a trans-activating CRISPR RNA (tracrRNA) transcribed from a nearby locus. Processing by RNase III (blue dashed arrow) in a process requiring Cas9 results in the formation of a mature CRISPR RNA (crRNA) which directs Cas9 to its target sequence. .

1.7.3 Decreasing the off-target effects of Cas9

Several methods have been developed to help overcome the problems of off-target cleavage by Cas9:

i) Cas9 contains two nuclease domains, an HNH domain which cleaves the strand complementary to the gRNA sequence and a RuvC domain which cleaves the non-complementary strand (Anders et al. 2014) (Figure 1-7). Mutations which knock out nuclease function in either of these domains (D10A in RuvC & H840A in HNH) create Cas9 nickases, each capable of cleaving only one strand of the target DNA. Using offset paired nCas9 nickases targeted to adjacent 20bp sequences on opposite strands by two different gRNAs, can reduce off-target cleavage by 50 to 1,500-fold. This strategy offers a much higher level of specificity in comparison to wild type Cas9, since DSB formation is dependent upon two single-strand nicks being formed in close proximity to each other (F. Ann Ran et al. 2013; Mali, Aach, et al. 2013).

ii) The use of gRNAs bearing 2 or 3 nucleotide 5' truncations to target Cas9 reduces off target activity since these PAM-distal terminal nucleotides are thought to help compensate for mismatches in the PAM-proximal seed region (Fu et al. 2014).

iii) Mutating both Cas9 nuclease domains ($RuvC^{D10A}$ and HNH^{H840A}) results in a catalytically inactive Cas9 (dCas9) that retains its ability to be targeted to specific sequences by a gRNA (Qi et al. 2013). Fusing dCas9 to FokI endonuclease (dCas9-FokI) has been successfully used to make targeted DSBs in genomic DNA. Acting via a similar mechanism to ZFNs and TALENs, this system relies upon two fused dCas9-FokI monomers being targeted to adjacent half sites whereupon DSBs are induced following dimerization of the two FokI subunits. Whilst this system can greatly reduce off-target activity, it is restrained by the requirement for half sites to be situated either 15 or 25bp apart due to the helical twist of DNA (Guilinger et al. 2014; Tsai et al. 2014).

iv) Reduced off-target specificity has also been achieved through transfection of ribonucleoprotein (RNP) complexes consisting of recombinant Cas9 protein and *in vitro* transcribed gRNAs (Kim, Kim, Cho, Kim & J. Kim 2014). It is thought that the persistence of plasmid DNA following transfection results in excessive transcription of Cas9 and

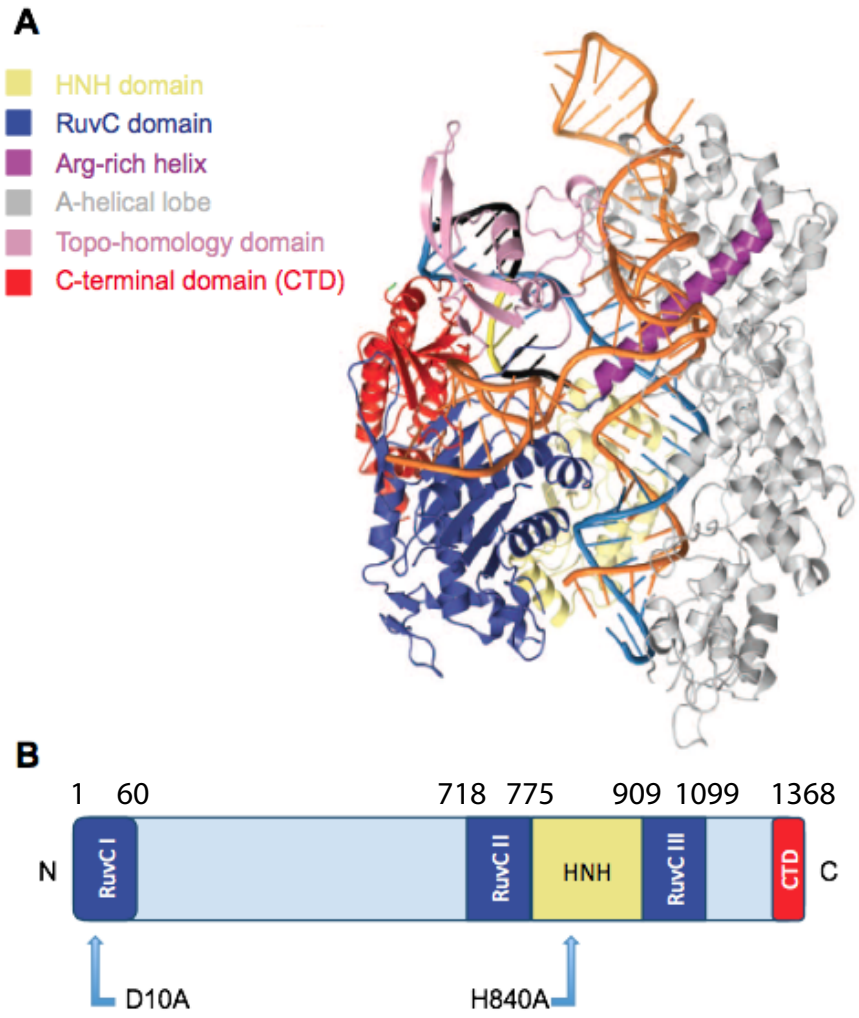


Figure 1-7 ***S. pyogenes*** **Cas9 structure**
A. Crystal structure of Cas9 in complex with gRNA and PAM-containing target DNA. Orange: gRNA; light blue: complementary DNA strand; black: non-complementary DNA strand; yellow: PAM (PDB: 4UN3). B. Domain architecture of Cas9 showing the location of the D10A and H840A mutations used to generate Cas9 nickases.

gRNA. The resultant persistence of Cas9 and gRNA within the cell increases the chances of off-target activity, whereas RNPs are rapidly degraded within the cells (Gaj et al. 2012). Furthermore, RNP-mediated delivery of CRISPR-Cas components increases cell survival in comparison to plasmid-mediated delivery (Kim, Kim, Cho, Kim & J. Kim 2014), since plasmid DNA can trigger cyclic GMP-AMP synthase activation and subsequent activation of the type I interferon pathway (L. Sun et al. 2013).

1.8 Summary and aims

The ability of cells to respond in an error-free and efficient manner to replication stress is key to the maintenance of genetic stability. Inaccurate restart of stalled and collapsed replication forks is a major source of a range of genetic aberrations that play a part in the progression of cancer. Knowledge of how cells respond to replication stress is therefore key to our understanding of the development of cancer and other genetic diseases. The development of a human Pu-seq methodology will allow detailed analysis of DNA replication dynamics in human cells in response to a variety of sources of replication stress, such as whether cells activate otherwise dormant replication origins in the vicinity of stalled or collapsed replication forks in response to induced replication stress. In addition, the human Pu-seq cell lines can be used to knock out other genes that are thought to be involved in programming replication timing and efficiency to see whether the absence of those genes alters the replication pattern.

The requirements for human Pu-seq cell lines are: i) establishing a method to downregulate ribonucleotide excision repair by targeting one of the subunits of the heterotrimeric enzyme RNase H2. ii) knocking in mutations to each of the two major replicative polymerases, polymerases δ and ε such that they incorporate excess ribonucleotides upon replication. iii) developing assays to assess RNase H2 activity and genomic ribonucleotide incorporation by the two replicative polymerases.

This thesis will describe progress made towards developing the cell lines necessary to carry out human Pu-seq.

Chapter 2

Materials and methods

2.1 Materials

All solutions were made using distilled water where appropriate. Materials and solutions were autoclaved at 125°C for 15 minutes when possible. Solutions and buffers were filter sterilized using a 0.2 μM filter (Nalgene) and stored at room temperature, −4°C or −20°C according to requirements.

LB Broth (Autoclaved)

10 g tryptone

5 g yeast extract

10 g NaCl

Adjusted to pH7.5 made up to 1 L with ddH_2O

LB Agar (Autoclaved)

7.5 g Agar

Made to 500 mL of prepared LB broth

TBE (5×) – for 1 L of 5× stock

54 g Tris base

27.5 g Boric acid

20 mL 0.5M EDTA pH8.0

Western blot transfer buffer (10×) – for 1 L of 10× stock

31 g Tris base

144 g Glycine

Made to 1 L using ddH_2O

Western blot transfer buffer (1×)

100 mL 10× transfer buffer

100 mL Methanol

800 mL ddH_2O

Tris-Glycine Electrophoresis buffer (Western blot running buffer)

31 g Tris base

144 g Glycine

100 mL 10% SDS

Made to 1 L using ddH_2O

Protein loading buffer (human)

40% Glycerol

240 mM Tris-HCl pH6.8

8% SDS

0.04% Bromophenol blue

5% beta-mercaptoethanol added just prior to use

Southern blot depurination buffer

0.25 M HCl

Southern blot denaturing buffer

1.5 M NaCl

0.5 M NaOH

Southern blot neutralizing buffer

1 M Tris-HCl

1.5 M NaCL

Made to pH7.4 with NaOH

10× SSC

1.5 M NaCl

0.15 M Sodium citrate

20% (w/v) TCA

100 g trichloroacetic acid in 500 mL ddH_2O

4× SDS-PAGE sample buffer (*S. pombe)*

250 mM Tris-HCl pH6.8

8% SDS

20% glycerol

0.4% Bromophenol blue

10% beta-mercaptoethanol

1× SDS-PAGE sample buffer (*S. pombe)*

1 vol 4× SDS-PAGE sample buffer

1 vol 1M Tris-HCl pH8

2 vol ddH_2O

2.5% beta-mercaptoethanol added just prior to use

2.2 Cloning and molecular methods

2.2.1 Polymerase chain reaction

PCRs were carried out using Phusion High-Fidelity DNA Polymerase (NEB) according to the manufacturer's guidelines. Generally, for amplification of plasmid template, initial denaturation was carried out at 98°C for 30 seconds followed by 30 rounds of denaturation at 98°C for 5 seconds, annealing at 3°C above the melting temperature of the primer with the lowest melting temperature for 15 seconds and extension at 72°C for 15 seconds/kb. Generally, for amplification of genomic DNA template, initial denaturation was carried out at 98°C for 30 seconds followed by ~30 rounds of denaturation at 98°C for 10 seconds, annealing at 3°C above the melting temperature of the primer with the lowest melting temperature (minimum 60°C) for 15 seconds and extension at 72°C for 30 seconds/kb.

2.2.2 Restriction digests

Restriction digests were carried out using enzymes from NEB and Thermo Fisher Scientific. Digestion reaction were incubated at the required temperatures for 15 minutes in the case of NEB HF enzymes or 1-2 hours for other enzymes. Digested fragments were run on 1% or 2% agarose gels (% w/v in 1× TBE) according to fragment size. The desired fragments were then excised and purified using Macherey-Nagel NucleoSpin Gel and PCR Clean-up kit.

2.2.3 Ligation reactions

2.2.3.1 Conventional ligation reactions

Conventional sticky-end and blunt-end ligation reactions were carried out using T4 DNA ligase (NEB) or Quick Ligation Kit (NEB) according to manufacturer's instructions. Briefly, reactions using T4 DNA ligase were carried out at room temperature for 10 minutes and reactions using Quick Ligation Kit were carried out at room temperature for 5 minutes. Following ligation, samples were placed on ice for 5 minutes before transformation into competent DH5α *E. coli* cells. Amplified sequences were sent for sequencing to check for the absence of unwanted mutations (GATC and Source Bioscience).

2.2.3.2 Homology-dependent cloning

Homology-dependent cloning was carried out using GeneArt Seamless Cloning and Assembly Kit (Invitrogen) according to the manufacturer's instructions. Briefly, fragments were designed such that adjacent fragments contained 15 bp of overlapping homology. 50 ng of cut vector was used in a reaction with 3-fold molar excess of insert and incubated at room temperature for 20 minutes. Cloning reactions were then on ice for 5 minutes before transformation into competent DH5α *E. coli* cells. Amplified sequences were sent for sequencing as previously described.

2.2.3.3 Assembly of pX330 plasmids

pX330-U6-Chimeric_BB-CBh-hSpCas9 (referred to as pX330 in this thesis) was a gift from Feng Zhang (Addgene plasmid # 42230). Cloning of pX330 plasmids was carried out as per the protocol provided by the Zhang laboratory (F Ann Ran et al. 2013). Briefly,

phosphorylated and annealed oligos containing the desired gRNA sequence were used in a dual restriction digest-ligation reaction with pX330, FastDigest BbsI (Thermo Fisher) and T7 DNA ligase (NEB) and incubated for 6 cycles of 37°C for 5 minutes and 21°C for 5 minutes. Following ligation, samples were treated with PlasmidSafe exonuclease (Epicentre) and transformed into competent DH5α *E. coli* cells.

Guide RNA sequences were selected using the online software at CRISPRdirect (http://crispr.dbcls.jp) or Benchling (www.benchling.com).

2.2.4 Design of gene targeting constructs

The gene targeting vectors used in this study were a kind gift from Professor Shunichi Takeda. Gene targeting plasmids were designed such that they contained homology arms of between ~750 and ~1250 bp in length. For each plasmid, the left homology arm was inserted at the AflII restriction site on the parental plasmid and the right homology arm was inserted at the ApaI restriction site.

The ROSA26 knockin construct used in this thesis was a kind gift from Dr. Helfrid Hochegger.

2.2.5 cDNA synthesis

cDNA was either synthesized by IDT DNA or amplified by reverse-transcription PCR from whole cell RNA extract using SuperScript III One-Step RT-PCR System with Platinum Tag High Fidelity (Invitrogen) according to manufacturer's instructions. Whole cell RNA was extracted using RNeasy Plus Minikit (Qiagen) according to manufacturer's instructions.

2.2.6 Site directed mutagenesis

Site directed mutagenesis was carried out using a two-step fusion PCR method. Firstly, two fragments were amplified using the wild-type template: one fragment was amplified using a wild type sequence forward primer and a reverse primer containing the desired mutation and the other fragment was amplified using a wild type sequence

reverse primer and a forward primer containing the desired mutation. The primers containing the point mutations were designed such that the melting temperature of the overlapping sequences was the same as the melting temperature of the wild type sequence primers. The full-length mutant fragment was amplified using equimolar amounts of each of the fragments generated in the first step as template with primers containing 15 bp extensions for the subsequent Seamless cloning step.

2.2.7 Bacterial transformation

100 μL of competent DH5α *E. coli* cells were transformed with 1 μL of previously generated miniprep plasmid or the full volume of ligation reaction. The DNA-*E. coli* mix was gently mixed by tapping and heat shocked at 42°C for 30 seconds then placed on ice for 2 minutes. For plasmids conferring resistance to ampicillin, transformed cells were streaked immediately onto LB agar plates supplemented with ampicillin. For plasmids conferring resistance to kanamycin, 1 mL LB broth was added and the tube was incubated at 37°C in a shaking benchtop incubator for 1 hour before streaking onto LB plates supplemented with kanamycin. LB plates were supplemented with ampicillin or kanamycin to a final concentration of 100 μg/mL.

2.2.8 Plasmid purification

Plasmid minipreps were purified from *E. coli* using NucleoSpin Plasmid Miniprep kit (Macherey-Nagel) or QIAprep Spin Miniprep kit (Qiagen). Plasmid midipreps were purified from *E. coli* using NucleoBond Xtra Midi kit (Macherey-Nagel) or Plasmid Midi Kit (Qiagen).

2.2.9 Human genomic DNA extraction

Human genomic DNA was extracted from cells using DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's instructions.

Table 2.1 DNA plasmids used in this study

NAME	DESCRIPTION	PARENT PLASMID	SOURCE
pCMV3xnls-I-SceI	Expresses I-SceI	Addgene plasmid #26477	Dr. Andy Porter
hCas9	Expresses human codon-optimised Cas9	Addgene plasmid # 41815	Prof. Feng Zhang
pHPRT-gRNA-1	Expresses gRNA targeting HPRT; ACTGCCTGACTAGGGATAAC	Addgene plasmid # 41824	Prof. Feng Zhang
pX330-U6-Chimeric_BB-CBh-hSpCas9	Co-expresses human codon-optimised Cas9 and gRNA	Addgene plasmid # 42230	Prof. Feng Zhang
Bsr-R_KI	Knockin vector conferring blasticidin resistance		Prof. Shunichi Takeda
His-R_KI	Knockin vector conferring histidinol resistance		Prof. Shunichi Takeda
Hygro-R_KI	Knockin vector conferring hygromycin resistance		Prof. Shunichi Takeda
G418-R_KI	Knockin vector conferring G418 resistance		Prof. Shunichi Takeda
Puro-R_KI	Knockin vector conferring puromycin resistance		Prof. Shunichi Takeda
Bsr-R_KI_PolyA	As Bsr-R_KI; introduces PolyA signal after inserted cDNA	Bsr-R_KI	This study
His-R_KI_PolyA	As His-R_KI; introduces PolyA signal after inserted cDNA	His-R_KI	Prof. Shunichi Takeda
Puro-R_KI_PolyA	As Puro-R_KI; introduces PolyA signal after inserted cDNA	Puro-R_KI	This study
POLD_ex1_gRNA	pX330 expressing gRNA targeting POLD1 exon 1; AGTCATGGTCTTCCCCGGCC	Addgene plasmid # 42230	This study
POLE_ex17 _gRNA	pX330 expressing gRNA targeting POLE exon 1; AACTGCACACGAGGCAAGCT	Addgene plasmid # 42230	This study
RNaseH2A_KO_gRNA	pX330 expressing gRNA targeting RNASEH2A exon 1; AGGCTCCTTGCGGCACACCG	Addgene plasmid # 42230	This study
RNaseH2A_Cterm-tag_gRNA	pX330 expressing gRNA targeting RNASEH2A terminal exon; GTTTAAGGAGAACCACACGT	Addgene plasmid # 42230	This study
RNaseH2B_Cterm-tag_gRNA	pX330 expressing gRNA targeting RNASEH2B terminal exon; TCCTGACTGTTAATGACTAC	Addgene plasmid # 42230	This study
RNaseH2B_Cterm-tag_gRNA	pX330 expressing gRNA targeting RNASEH2C terminal exon; CCTGTGCGTGAATCTGCAAC	Addgene plasmid # 42230	This study
ROSA26_gRNA	pX330 expressing gRNA targeting human ROSA26 locus	Addgene plasmid # 42230	Dr. Helfrid Hochegger
POLD_LM_KI_Bsr	POLD1 L606M cDNA knockin construct; blasticidin resistance	Bsr-R_KI_PolyA	This study
POLD_LM_KI_His	POLD1 L606M cDNA knockin construct; histidinol resistance	His-R_KI_PolyA	This study
POLD_LM_KI_Puro	POLD1 L606M cDNA knockin construct; puromycin resistance	Puro-R_KI_PolyA	This study
POLD_LG_KI_Bsr	POLD1 L606G cDNA knockin construct; blasticidin resistance	Bsr-R_KI_PolyA	This study
POLD_LG_KI_Puro	POLD1 L606G cDNA knockin construct; puromycin resistance	Puro-R_KI_PolyA	This study
POLD_LK_KI_Bsr	POLD1 L606K cDNA knockin construct; blasticidin resistance	Bsr-R_KI_PolyA	This study
POLD_LK_KI_Puro	POLD1 L606K cDNA knockin construct; puromycin resistance	Puro-R_KI_PolyA	This study
POLE_MG_KI_Bsr	POLE M630G cDNA knockin construct; blasticidin resistance	Bsr-R_KI	This study
POLE_MG_KI_Puro	POLE M630G cDNA knockin construct; puromycin resistance	Puro-R_KI	This study
POLE_MA_KI_Bsr	POLE M630A cDNA knockin construct; blasticidin resistance	Bsr-R_KI	This study
POLE_MV_KI_Bsr	POLE M630V cDNA knockin construct; blasticidin resistance	Bsr-R_KI	This study
POLE_MV_KI_Puro	POLE M630V cDNA knockin construct; puromycin resistance	Puro-R_KI	This study
POLE_MF_KI_Bsr	POLE M630F cDNA knockin construct; blasticidin resistance	Bsr-R_KI	This study
POLE_MF_KI_Puro	POLE M630F cDNA knockin construct; puromycin resistance	Puro-R_KI	This study
POLE_ML_KI_Bsr	POLE M630L cDNA knockin construct; blasticidin resistance	Bsr-R_KI	This study
POLE_ML_KI_Puro	POLE M630L cDNA knockin construct; puromycin resistance	Puro-R_KI	This study
RNaseH2A_KO_G418	RNASEH2A knockout repair template; hygromycin resistance	Hygro-R_KI	This study
RNaseH2A_KO_Hygro	RNASEH2A knockout repair template; G418 resistance	G418-R_KI	This study
RNaseH2A_Cterm-mAID_GFP_G418	RNASEH2A C-terminal mAID tagging; G418 resistance	G418-R_KI	Prof. Shunichi Takeda
RNaseH2A_Cterm-mAID_SMASh_G418	RNASEH2A C-terminal mAID-SMASh tagging; G418 resistance	G418-R_KI	This study
RNaseH2A_Cterm-mAID_SMASh_hygro	RNASEH2A C-terminal mAID-SMASh tagging; hygromycin resistance	Hygro-R_KI	This study
RNaseH2B_Cterm-AID_SMASh_G418	RNASEH2B C-terminal AID-SMASh tagging; G418 resistance	pBluescript	This study
RNaseH2C_Cterm-AID_SMASh_G418	RNASEH2C C-terminal AID-SMASh tagging; G418 resistance	pBluescript	This study
RNaseH2B_Nterm-SMASh_ROSA	rtTA_TRE3GS_N-terminal SMASh tagged RNASEH2B cDNA at ROSA26; Zeocin resistance	Takara 631168	Dr. Helfrid Hochegger/This study
RNaseH2C_Nterm-SMASh_ROSA	rtTA_TRE3GS_N-terminal SMASh tagged RNASEH2C cDNA at ROSA26; Zeocin resistance	Takara 631168	Dr. Helfrid Hochegger/This study

Table 2.2 Oligonucleotides used in this study

NAME	SEQUENCE
rNTP_Assay_18mer	/5rApp/CTGTAGGCACCA/iBiodT/CAAT/3ddC/
rNTP_Assay_9mer	/5rApp/CAGCAT/iBiodT/A/3ddC/
Universal_RA_Seq_FW	CATGCTCCAGACTGCCTT
Universal_RA_Seq_RV	GCCTTGACTTGAGGTTAG
Universal_LA_Seq_FW	TGCGCAACTGTTGGGAAG
Universal_LA_Seq_RV_V2	GCTATACGAAGTTATTAGGTCCC
POLD_RA_SMLS_RV	CACTAGTAGGCGCGCCCCTGGCCAACATGGCGAAA
POLD_RA_SMLS_FW	TGGGAAGCTTGTCGACCTTCCAACAGTTGGAGATTGACC
POLD_cDNA_SMLS_RV	GCTCGAGGGGGGGGCCTCACCAGGCCTCAGGTCCA
POLD_cDNA_SMLS_FW	AACTTGCCCAGCAGGATGGATGGCAAGCGGCG
POLD_LA_SMLS_RV	CCGCTTGCCATCCATCCTGCTGGGCAAGTTGGA
POLD_LA_SMLS_FW	ATTGGGTACCGGGCCAGAAAGCAGGTTACTATAAACGGC
L606G_SDM_FW	CTCCTCGggGTACCCGTCCATCATG
L606G_SDM_RV	GACGGGTACccCGAGGAGAAGTCCAG
L606K_SDM_FW	CTCCTCGaaGTACCCGTCCATCATG
L606K_SDM_RV	GACGGGTACttCGAGGAGAAGTCCAG
POLD_cDNA_Seq_1	ATGACGTCCTGTGCGTGAG
POLD_cDNA_Seq_2	TTGAGGGGCTCGATGACAG
POLD_cDNA_Seq_3	ACTCGATATCGAAGCTGAGCA
POLD_cDNA_Seq_4	AAGCCAGCGAGGATCTATGG
RNASE_ASSAY_TOP (DRD)	GATCTGAGCCTGGG(rA)GCT-3'fluorescein
RNASE_ASSAY_TOP (DNA)	GATCTGAGCCTGGGAGCT-3'fluorescein
RNASE_ASSAY_BOTTOM	5'Dabcyl-AGCTCCCAGGCTCAGATC
POLD_RT_PCR_FW	ACTCTTCATTCCAGTCCAAGCAGACGGG
POLD_RT_PCR_RV	GAGGACACGCCGAATCGGCACATGA
POLD_RT_PCR_SEQ	TCAGCTACCTGCTCAGTC
DB16_LA_FW	TGGATCCCCCGGGCTGCAGGCCTCGTGGGATACGTGATGGTGACAGAAGAGAAGAAGG
DB16_LA_RV	gttcgtatacGTCCTCGGGCACCTGTGCGTGAATCTGCAACAGcAGTCGCC
DB16_RA_FW	GCCCGAGGACGTATACGAACCAGAGCTTGAAATTC
DB16_RA_RV	TATCGATAAGCTTGATATCGTCGACTCCAAGTGTCTGC
DB17_Linker_FW	CACAGGTGCCCGAGGACGTAGGCGCCTCAGCGGCATCA
DB17_SMASh_RV	CTCCGCTTCCGTAATACAGGACTTCTCGGTCAGGGATGATAG
DB17_P2A_FW	CCTGTATTACGGAAGCGGAGCTACTAAC
DB17_Neo_RV	ATTTCAAGCTCTGGTTCGTATCAGAAGAACTCGTCAAGAAG
RNC_Cterm_pX330_FW	caccgCCTGTGCGTGAATCTGCAAC
RNC_Cterm_pX330_RV	aaacGTTGCAGATTCACGCACAGG
Neo_Seq_FW	CTGAATGAACTGCAAGACGAG
CTermSMASh_Seq1	AAGGCCGTGGACTTCATC
CTermSMASh_Seq2	CATTGATGCATGTGGCCAG
CTermSMASh_Seq3	TTGTGATAGGGTGGGTCAG
PURO_FW	GTCACCGAGCTGCAAGAAC
PURO_RV	GAGCCGCTCGTAGAAGGG
ZEO_FW	GACGACGTGACCCTGTTCAT
ZEO_RV	CTCCTCGGCCACGAAGTG

2.3 Western Blotting

2.3.1 Preparation of human whole cell extract

Human cells were grown to confluency in 6 well plates or T75 flasks, trypsinized as required and washed with PBS. Samples were pelleted and resuspended in 50 μL Western lysis buffer (9M urea, 50 mM Tris, pH7.5, 1% (v/v) beta-mercaptoethanol added just prior to use) and incubated at room temperature for 15 minutes then incubated on ice. Samples were lysed in a Bioruptor Pico sonication device (Diagenode) for 6 cycles of 30s on/30s off and spun at full speed for 10 minutes at 4°C. The supernatant was transferred to a new tube and 1 μL of supernatant used to assess protein yield by Bradford assay. Samples were mixed with an equal volume of 2× Protein loading buffer and incubated at 98°C for 2 minutes just prior to loading.

2.3.2 Preparation of *S. pombe* whole cell extract

5 OD_{595} of overnight culture was pelleted and resuspended in 200 μL 20% trichloroacetic acid (TCA). Glass beads (Sigma) were added up to the level of TCA and samples were treated in a multi-beads shocker (Yasui Kakui) for 20 × 60 seconds at 2.5 m/s. The bottom of the tube was punctured with a hot needle and the contents spun into a fresh tube. The samples were then spun at 13000 RPM in a microcentrifuge and the supernatant removed. The extracts were resuspended in 200 μL 1× SDS-PAGE sample buffer and boiled for 5 minutes. Samples were spun at 13000 RPM before loading.

2.3.3 Protein electrophoresis and immunoblotting

Whole cell protein extracts were resolved by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). The samples were resolved through a 5% acrylamide stacking gel at 80V for ~15 minutes and separated on a 12% acrylamide resolving gel at 180V for ~30 minutes. The stacking gel consisted of 5 % acrylamide from a 30% acrylamide stock (National Diagnostics), 0.125M Tris pH6.8, 0.1 % (w/v) SDS, 0.1% (w/v) ammonium persulphate and 0.1 % (v/v) N,N,N',N'-Tetramethylethylenediamine (TEMED). The resolving gels were made up with 12% acrylamide from a 30% acrylamide

stock (National Diagnostics), 0.375M Tris pH8.8, 0.1% (w/v) SDS, 0.1% (w/v) ammonium persulphate and 0.04% (v/v) TEMED.

Proteins were transferred from the gel to Hybond nitrocellulose membrane (GE Healthcare) via wet transfer for 90 minutes on ice at 90V constant in 1× transfer buffer. The membrane was blocked in 5% milk PBST (Marvel dried skimmed milk in phosphate buffered saline, 0.1% Tween (Sigma) for 1 hour at room temperature and incubated at 4°C overnight in 5% milk PBST with an appropriate dilution of primary antibody. The membrane was then washed with PBST for 2 × 5 minutes and incubated for 20-60 minutes in 5% milk PBST with an appropriate dilution of secondary antibody. Excess secondary antibody was washed off with PBST for 2 × 5 minutes and bound antibody was detected by chemiluminescence using ECL Western Lightning Plus (Perkin Elmer). Emission was captured using autoradiograph film and developed.

Table 2.3 Primary antibodies used in this study

Antibody	Host species	Supplier	Dilution
RNase H2A	Rabbit	Bethyl A304-149A	1:1000
RNase H2A	Rabbit	Bethyl A304-150A	1:1000
RNase H2B	Mouse	Invitrogen MA5-23523	1:2000
RNase H2C	Rabbit	Atlas HPA065375	1:2000
FLAG	Mouse	Invitrogen MA1-91878	1:500
GFP	Mouse	Roche 11814460001	1:1000
Tubulin	Mouse	Sigma T5168	1:25000
Histone H2Z	Rabbit	Merck 07-594	1:1000

2.4 Human cell culture

2.4.1 Maintenance of cell lines

Primary human cell lines were cultured at 37°C with 5% CO_2 in Gibco RPMI-1640 medium supplemented with 15% foetal calf serum (FCS), 2 mM L-glutamine, 100 U/mL penicillin, 100 μg/mL streptomycin and 1 mM sodium pyruvate. TK6 cells were cultured

at 37°C with 5% CO_2 in Gibco RPMI-1640 medium supplemented with 5% horse serum, 2 mM L-glutamine, 100 U/mL penicillin, 100 μg/mL streptomycin and 1 mM sodium pyruvate. RPE1 cells were cultured at 37°C with 5% CO_2 in DMEM-F12 (Sigma) supplemented with 10% FCS, 100 U/mL penicillin and 100 μg/mL streptomycin and 1 mM sodium pyruvate.

HT1080 *HP-I-RT* cells were cultured at 37°C with 5% CO_2 in Gibco DMEM supplemented with 10% FCS, 100 U/mL penicillin, 100 μg/mL streptomycin, 2 mM L-glutamine and 1 mM sodium pyruvate. To avoid spontaneous 6-TG resistant colonies cells were grown in HAT (5mM Sodium Hypoxanthine, 20μM Aminopterin, 0.8mM Thymidine, Invitrogen). Following I-SceI or Cas9-mediated knockout of *HPRT*, HT1010 *HP-I-RT* cells were grown in Gibco DMEM supplemented with 10% FCS, 100 U/mL penicillin, 100 μg/mL streptomycin, 2 mM L-glutamine, 1 mM sodium pyruvate and HT (5mM Sodium Hypoxanthine, 0.8mM Thymidine, Invitrogen).

2.4.2 Plasmid Transfection

2.4.2.1 Lipofectamine 2000

HT1080 cells were seeded 24 hours prior to transfection in 24 well plates and grown to ~80% confluency. 100 μL serum-free medium was mixed with the appropriate volume of Lipofectamine 2000 solution and incubated with 1 μg plasmid DNA made up to 100μl with serum-free medium at room temperature for 5 minutes. 200μl of solution was added to each well and samples incubated for 24 hours prior to checking for expression of GFP or expanded into a larger volume after 48 hours for 6-TG assays.

2.4.2.2 FuGENE HD

HT1080 cells were seeded 24 hours prior to transfection in 24 well plates and grown to ~80% confluency. Varying quantities of FuGENE HD reagent were added to 2 μg plasmid DNA made up to 50 μL with culture medium to give the desired ratio of FuGENE HD (μL):DNA (μg). Each solution was added immediately to cells and samples were incubated for 24 hours prior to checking for expression of GFP or expanded into a larger volume after 48 hours for 6-TG assays.

2.4.2.3 Xfect

HT1080 cells were seeded 24 hours prior to transfection in 24 well plates and grown to ~60% confluency. The appropriate mass of plasmid DNA was suspended in 50 µL Xfect buffer to which was added 0.3 µL Xfect polymer/µg DNA. The solution was added to cells following 10 minutes incubation at room temperature. 4 hours post-transfection cells were washed with PBS and culture medium replaced with fresh medium. Samples were checked for expression of GFP 24 hours after transfection or expanded into a larger volume after 48 hours for 6-TG assays.

2.4.2.4 Amaxa nucleofection

HT1080 cells were passaged 48 hours prior to transfection. Prior to transfection, cells were trypsinized, washed with PBS and 10^6 cells suspended in 100 µL Nucleofector Solution T with 2 µg DNA. Samples were transferred to a cuvette and treated with Nucleofector program L-005. Immediately after transfection, 500 µL pre-warmed culture medium was added to the cuvette and samples transferred to 6 well plates. 24 hours after transfection, samples were checked for expression of GFP or expanded into a larger volume after 48 hours for 6-TG assays. For knockin and knockout experiments, samples were cultured for 48 hours after transfection then transferred into 96 well plates in selective media with the appropriate antibiotic(s) to isolate successfully targeted clones.

2.4.2.5 Neon electroporation

To transfect TK6 cells, 10^6 cells and 1 µg DNA were suspended in 10 µL solution R and electroporated using the 10 µL Neon tip and the following settings: 1350 V, 10 ms, ×3 pulse. Following transfection, cells were cultured in 2 mL 10% horse serum for 48 hours before transferring into 96 well plates in selective media supplemented with the appropriate antibiotic(s) to isolate successfully targeted clones.

To transfect RPE1 cells, 10^6 cells and 1.5 µg DNA were suspended in 30 µL solution R and electroporated three times using the 10 µL Neon tip and the following settings: 1350 V, 20 ms, ×2 pulse. Following transfection, cells were transferred into 2 mL culture medium and cultured for 48 hours before transferring into 96 well plates in

selective media supplemented with the appropriate antibiotic(s) to isolate successfully targeted clones.

2.4.3 HT1080 6-TG resistance colony formation assay

48 hours after transfection with plasmids encoding either I-SceI or Cas9 and a gRNA targeting the *HPRT* gene, the appropriate number of cells was transferred to a 10 cm^2 dish in 10 mL growth medium. 5 days after seeding the growth medium was replaced with medium containing 6-thioguanine (6-TG) at a final concentration of 15 μg/mL. After 10 days growth in 6-TG, plates were stained with methylene blue (0.4% w/v in 50% methanol) and 6-TG-resistant colonies counted. All plasmids used for this assay were a kind gift from Dr. Andy Porter.

2.5 *S. pombe* methods

2.5.1 *S. pombe* growth media

Yeast Extract (YE) media

5 g/L yeast extract

30 g/L glucose

Supplemented as necessary with:

0.2 g/L Adenine

0.1 g/L each of leucine, uracil, histidine and arginine.

YE Agar (YEA) plates

As for YE media plus:

12.5 g/L agar

Extremely Low Nitrogen (ELN) plates

27.3 g/L Formedium EMM broth (w/o nitrogen)

0.05 g/L Ammonium chloride

0.2 g/L Adenine

0.1 g/L each of leucine, uracil, histidine and arginine

25 g/L Agar

2.5.2 Lithium acetate *S. pombe* transformation

10^8 cells from an overnight culture grown in YE were pelleted and washed once with ddH_2O and once with 1 mL LiAc-TE (0.1 M LiAc, 10 mM Tris-HCl, 1 mM EDTA, pH7.5). Cells were then resuspended in 100 μL LiAc-TE to which was added 2 μl of salmon sperm DNA (Invitrogen) and 10 μg of PCR product. This mixture was incubated at room temperature for 10 minutes after which 260 μL of 40% PEG-4000/LiAc-TE was added. The cell suspension was then incubated at 30°C for ~45 minutes. 43 μL of dimethyl sulfoxide (DMSO) was added and the cells heat shocked at 42°C for 5 minutes. Cells were then pelleted by centrifugation and the pellet washed with 1 mL of ddH_2O and resuspended in 200 μL ddH_2O. Samples were then plated onto plates containing the relevant supplements.

2.5.3 Genomic DNA extraction from *S. pombe*

Cells were grown overnight in YE and 5 OD_{595} cells were spun down. The pellet was resuspended in 1mg/ml zymolyase in CSE and incubated at 37°C for 30 minutes. The cells were pelleted again and resuspended in 450 μL 5× TE to which was added 50 μL 10% SDS, then incubated for 5 minutes at room temperature. 150 μL of 5 M KAc was added and the sample was incubated on ice for 10 minutes. The lysed cells were spun at 13000 RPM for 10 minutes and the supernatant transferred to a fresh tube. 1 volume of isopropanol was added and the sample was spun at 13000 RPM at 4°C for 10 minutes. The supernatant was removed and the pellet washed with 500 μL ethanol and spun again for 5 minutes. The supernatant was removed and the pellet resuspended in 200 μL ddH_2O and 5 μL RNase A.

2.5.4 Genetic cross and random spore analysis

h+ and h- *S. pombe* were mixed in 5 μL ddH_2O and plated onto an ELN plate. The cells were incubated for 2 days at 25°C and crossing efficiency verified by microscopy. A loopful of the crossed cells was suspended in 1 mL ddH_2O and 2 μL of β-galactosidase from *Helix pomentia*. The spores were incubated overnight on a rotating wheel at room

temperature. The following day, ~1000 spores were plated onto YEA and the plates incubated at 30°C for 5 days.

2.5.5 *S. pombe* RNase H2 activity assay

To test the integrity and to estimate the yield of extracted gDNA, 1 μL of DNA was mixed with 9 μL H_2O and 2 μL DNA loading dye. 6 μL of this mixture was run on a 1% agarose gel alongside 6 μL DNA ladder. DNA concentration was calculated by comparing the signal intensity of each gDNA sample with a band of known concentration in the DNA ladder using ImageJ. To assess the extent of ribonucleotide presence in extracted gDNA, equal concentrations of untreated gDNA from different strains were run on a 1% agarose gel alongside 3-fold greater quantities of gDNA treated with 0.3 M KOH at 55°C for 1 hour. RNase H2 activity was inferred from the extent of alkaline hydrolysis.

2.6 *A. thaliana* tRNA ligase (AtRNL) reactions

2.6.1 AtRNL activity assay

Substrates were prepared by treating a ribonucleotide-containing 70mer oligonucleotide or a DNA-only 70mer oligonucleotide with 0.3 M KOH at 55°C for 2 hours. 10 μL reactions were set up containing 10 pmol of alkali-treated substrate and 50 pmol of internally biotinylated oligo (5'-rAppCTGTAGGCACCA-iBiodT-CAATddC-3' or 5'-rAppCAGCAT-iBiodT-AddC-3'), 1× tRNA ligase buffer (10 mM Tris–HCl pH7.5, 100 mM KOAc, 0.3 mM spermine, 6 mM $Mg(OAc)_2$, 0.5 mM DTT, 0.5% Triton X-100), 14% polyethylene glycol, and *A. thaliana* tRNA ligase (1 or 2 μL) and incubated at room temperature for 30 minutes or 2 hours. 1 μL of reaction mixture was diluted in 4 μL ddH_2O and combined with 5 μL 2× loading buffer (90% formamide, 0.5% EDTA, 0.1% xylene cyanol, 0.1% bromophenol blue) and run on a 10 mL 15% denaturing acrylamide gel at 170V for 20 minutes and 250V until the dye front neared the bottom of the gel.

10 mL 15% denaturing acrylamide gel

4.8 g urea

3.75 mL 40% acrylamide

100 µL 10% APS

2 mL 5× TBE

4 µL TEMED

Make up to 10 mL with ddH_2O

2.6.2 *S. pombe* AtRNL ribonucleotide incorporation assay

≥ 3 µg *S. pombe* gDNA was treated in 16.67 µL 0.3 M NaOH at 55°C for 2 hours, then neutralized with 5 µL 1 M HCl and 1 µL Tris base (2 M, pH7). 500 ng alkali-treated gDNA was used in a ligation reaction with 50 pmol of internally biotinylated oligonucleotide, 1× tRNA ligase buffer (10 mM Tris–HCl pH7.5, 100 mM KOAc, 0.3 mM spermine, 6 mM $Mg(OAc)_2$, 0.5 mM DTT, 0.5% Triton X-100), 14% polyethylene glycol, and 2 µL *A. thaliana* tRNA ligase and incubated at room temperature for 2 hours. Reaction mixtures were cleaned up using Nucleospin Gel and PCR Clean-up columns (Macherey-Nagel) or Illustra G-50 columns (GE healthcare). Samples were resolved on a 400 mL 1% Mega Base Rapid Agarose (Cambridge reagents) gel at 90V overnight. After running, the gel was washed in Southern blot depurination buffer for 10 minutes then washed in Southern blot denaturing buffer for 45 minutes and finally washed twice in Southern blot neutralisation buffer for 30 minutes then 15 minutes.

DNA was transferred onto Genescreen nylon membrane (Perkin Elmer) by capillary transfer overnight using 10× SSC. The membrane was allowed to dry and the DNA was crosslinked to the membrane using a Stratalink UV crosslinker at 1200 × 100µJ/cm^2. The membrane was then washed for 2 × 5 minutes in blocking buffer (0.5% SDS, 0.2% (w/v) I-Block reagent (Tropix) in 1× PBS) then incubated for 10 minutes in blocking buffer. The membrane was then incubated in streptavidin-alkaline phosphatase, diluted 1:20000 in blocking buffer for 20 minutes with shaking. The membrane was then washed for 5 minutes in blocking buffer then 6 × 15 minutes in wash buffer (0.5% SDS in 1× PBS). The membrane was next washed for 2 × 2 minutes in 1× assay buffer (dilute 10× assay buffer 1:10 in ddH_2O; 10x assay buffer (200 mM Tris-HCl pH9.8, 10 mM $MgCl_2$)). Samples were

visualized using CDP-Star chemiluminescent substrate (GE Healthcare) in an ImageQuant LAS 4000 imager (GE Healthcare).

2.7 RNase H2 activity assays

2.7.1 Whole cell extract preparation

~6 × 10^6 cells were pelleted and washed once with PBS. Pellets were resuspended in 50 μL lysis buffer (50 mM Tris pH8, 280 mM NaCl, 0.5% NP-40, 0.2 mM EDTA, 0.2 mM EGTA, 10% glycerol (v/v), 1 mM DTT, 1× Roche Complete Protease Inhibitor Tablet) and incubated on ice for 10 minutes. An equal volume of cytoplasmic buffer (20 mM HEPES pH7.9, 10 mM KCl, 1mM EDTA, 10% glycerol (v/v) 1 mM DTT, 1× Roche Complete Protease Inhibitor Tablet) was added, incubated on ice for 10 minutes and extracts were cleared by centrifugation for 5 minutes at 13000 RPM at 4°C. Final protein concentration was determined by Bradford assay.

2.7.2 Fluorometric substrate preparation

10 μM of 3'-fluorescein labelled oligonucleotides (GATCTGAGCCTGGG(rA)GCT-3'fluorescein for RNase H2-specific activity or GATCTGAGCCTGGGAGCT-3'fluorescein for non-RNase H2-mediated activity; IDT DNA) were annealed to 10 μM of complementary 5'-dabcyl labelled oligonucleotides (5'Dabcyl-AGCTCCCAGGCTCAGATC; IDT DNA) in 60 mM KCl, 50 mM Tris-HCl pH8 by denaturing at 95°C followed by gradual cooling to room temperature.

2.7.3 RNase H2 fluorometric assays

100 μL reactions were set up using the desired mass of whole cell protein extract with 250 nM substrate in 60 mM KCl, 50 mM Tris-HCl pH8, 10 mM $MgCl_2$, 0.01% bovine serum albumin, 0.01% Triton X-100. Reactions were carried out in black, flat bottomed 96-well plates (Nunc). For end-point assays, reactions were carried out at room temperature for 90 minutes and quenched with a final concentration of 20 mM EDTA. For time course assays, reactions were carried out at room temperate with fluorescence readings taken at 10-minute intervals. Fluorescence intensity was read using a

CLARIOstar plate reader (BMG labtech) using a 480 nm excitation filter and 535 nm emission filter.

2.8 Alkaline comet assay

Frosted slides (Fisher) were pre-coated with 0.6% agarose (w/v in 1× PBS). Cells were suspended in pre-chilled PBS and mixed with an equal volume (150 μL) of 1.2% Low Melting Point agarose (Sigma) that had been pre-incubated at 42°C. The cell suspension (~3×10^5 cells/mL) was dispersed onto the pre-coated slides and maintained at 4°C in the dark until set. All subsequent steps were carried out in the dark. After gelling, slides were immersed in lysis buffer (2.5 M NaCl, 10 mM Tris-HCl, 100 mM EDTA pH 8.0, 1% Triton X-100, 1 % DMSO; pH 10) for 1 hour then washed for 2 × 10 minutes with pre-chilled ddH_2O. Slides were then incubated in alkaline electrophoresis buffer (50 mM NaOH, 1 mM EDTA, 1 % DMSO) for 45 minutes. Electrophoresis was carried out at 12V for 25 minutes followed by neutralization with 0.4M Tris pH7 for 1 hour. DNA was stained with SYBR Green (1:10000 dilution) and mean comet tail length of 50 cells scored using Comet Assay IV software (Perceptive Instruments).

Chapter 3

Establishing a method to knock out RNASEH2A using CRISPR-Cas9

3.1 Introduction

Previous work in the Carr laboratory has established a system called Polymerase Usage Sequencing (PU-Seq) to monitor genome-wide replicative polymerase usage in *S. pombe* (Daigaku et al. 2015). This system requires two different strains of *S. pombe* containing mutations in either of the two major replicative polymerases, polymerase delta (Pol δ) or polymerase epsilon (Pol ε), that result in higher then wild type levels of ribonucleotide (rNTP) incorporation during DNA replication. In each of these strains, RNase H2A, the catalytic subunit of RNase H2, the enzyme responsible for the removal of the majority of misincorporated rNTPs, is knocked out. As a result, rNTPs remain embedded within the genome and serve as a molecular footprint of the mutant DNA polymerases. In *S. pombe*, RNase H2A is encoded by the gene *rnh201*.

The overall aim of this project is to set up a system to monitor replicative polymerase genome-wide in human cells. To set up this system in human cells, two criteria must therefore be met: 1) to establish mutations in both Pol δ and Pol ε that will result in higher than wild type levels of rNTP incorporation in either the lagging strand or leading strand respectively and 2) to prevent the removal of misincorporated rNTPs by RNase H2. In the first instance it was decided to attempt to knock out RNase H2A in human cells using CRISPR-Cas9.

As described in section 1.7.2 CRISPR-Cas systems provide bacteria and archaea with adaptive immunity against foreign DNA. The system relies upon the endonuclease activity of a CRISPR-associated (Cas) protein which is targeted to viral and plasmid DNA

by an antisense RNA molecule (Jinek et al. 2012). This system was recently repurposed to allow for precise genome editing in human cells via the induction of *Streptococcus pyogenes* Cas9-mediated double strand breaks (DSBs) (Mali, Yang, et al. 2013; Cong et al. 2013).

Due to the recent discovery of CRISPR-Cas9 for genome editing, the aim of this chapter was therefore to set up a working protocol for gene knockout using CRISPR-Cas9 and to use this protocol to attempt to regulate RNase H2 by knocking out RNase H2A in human cells.

3.2 Transfection optimization in HT1080 cells

3.2.1 DNA Transfection

Genome editing depends upon efficient transfection of the required components into cells. If the transfection method used is inefficient less DNA enters the cells and as such fewer cells express the genome editing machinery. As such, to prevent the delivery of plasmid DNA into cells being a major limiting factor in gene editing it is important to optimize transfection for each cell line to achieve the best possible transfection efficiency. For transfection optimization we used the HT1080 cell line which is a human fibrosarcoma cell line derived from a male patient. HT1080 cells have the benefit of a relatively stable karyotype with a modal chromosome number of 46 (range 44-48). In addition, the presence of one single X chromosome allows for accurate gene targeting optimization using a reporter gene located on the X chromosome (described in further detail in section 3.3.1). All of the HT1080 cells described in this project were a kind gift from Dr. Andy Porter.

To establish the most effective means of delivering DNA into HT1080 cells, four plasmid-based transfection methods were chosen: 1) FuGENE HD (Promega), 2) Xfect (Clontech), 3) Amaxa Nucleofection (Lonza) and 4) Lipofectamine 2000 (Invitrogen). These reagents were chosen to provide four different methods of delivering DNA into cells: non-liposomal, polymer-based, electroporation and lipid-based respectively.

To assess transfection efficiency using the different transfection reagents, cells were transfected with a plasmid expressing human codon optimized eGFP using a range of conditions as suggested by the manufacturers and transfection efficiency was scored as a percentage of cells expressing GFP:total number of cells in field of view, with at least 60 cells scored per experiment (Figure 3-1). Data shown are from one single experimental repeat.

For transfection with FuGENE HD ~5×10^5 HT1080 cells were transfected with 2 μg DNA with differing ratios of transfection reagent volume (1.5 μl reagent/1 μg DNA; 2 μl/μg; 2.5 μl/μg; 3 μl/μg; 3.5 μl/μg or 4 μl/μg). Transfection with FuGENE HD led to transfection efficiencies of 0% (1.5 μl reagent/1 μg DNA); 1% (2 μl/μg); 5% (2.5 μl/μg); 0% (3 μl/μg); 46% (3.5 μl/μg) and 76% (4 μl/μg) (Figure 3-1 A).

For transfection with Xfect ~5×10^5 HT1080 cells were transfected with 1 μg, 1.5 μg, 2 μg or 2.5 μg DNA with Xfect reagent added at a ratio of 0.3 μl reagent per μg DNA. Transfection with Xfect led to transfection efficiencies of 55% (1 μg DNA); 93% (1.5 μg DNA); 95% (2 μg DNA) and 98% (2.5 μg DNA) (Figure 3-1 B).

For transfection with Amaxa nucleofection, 10^6 HT1080 cells were transfected with 2 μg DNA using Kit T and program L-005. Transfection with Amaxa nucleofection led to a transfection efficiency of 80% (Figure 3-1 C).

For transfection with Lipofectamine 2000 ~5×10^5 HT1080 cells were transfected with 1 μg DNA with differing ratios of transfection reagent volume (1 μl reagent/1 μg,DNA; 2.5 μl/μg; 5 μl/μg; 7.5 μl/μg or 10 μl/μg). Transfection with Lipofectamine 2000 led to transfection efficiencies of 15% (1 μl reagent/1 μg,DNA); 86% (2.5 μl/μg); 65% (5 μl/μg; 42% (7.5 μl/μg) and 37% (10 μl/μg). Lipofectamine 2000-mediated transfection using 7.5 μl reagent/μg DNA and 10 μl reagent/μg DNA caused cell sickness as inferred from cell morphology so even if transfection efficiency had been higher using these two conditions, they would not have been suitable conditions for further transfections.

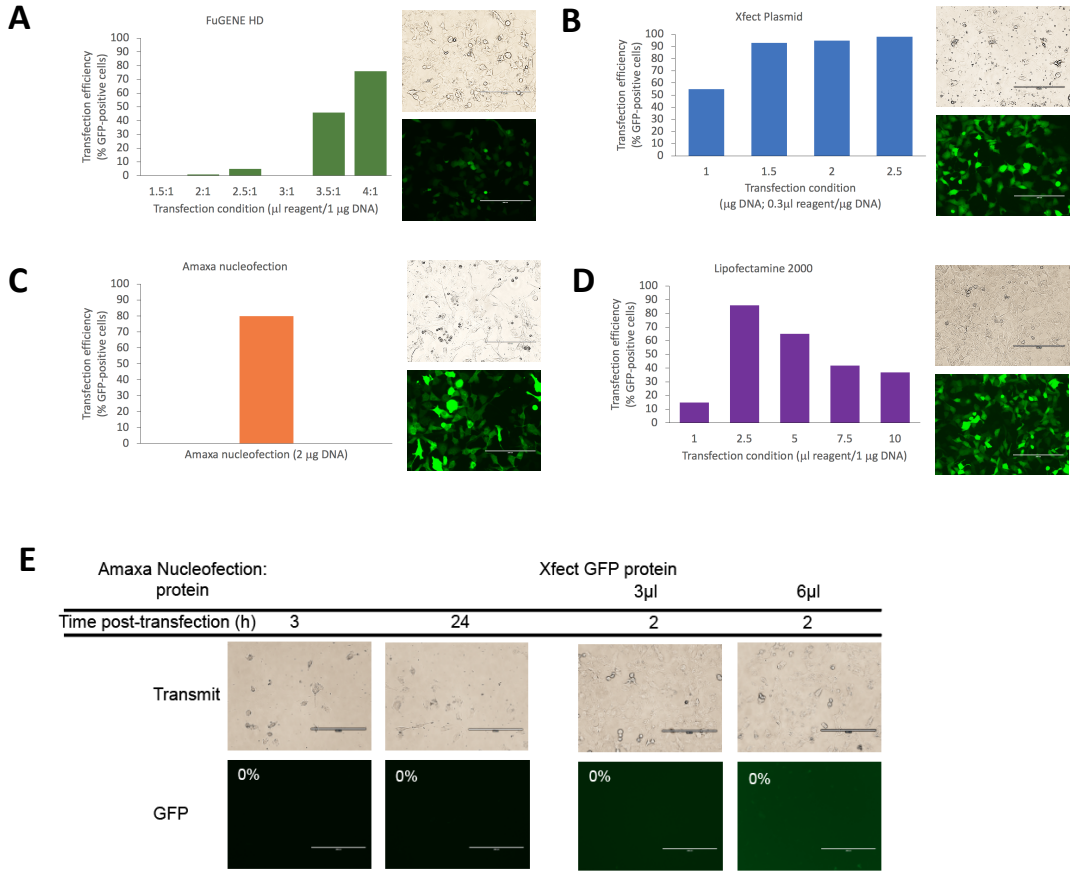


Figure 3-1 Transfection optimisation in HT1080 cells using a plasmid expressing eGFP and GFP protein

A-D. Transfection efficiency defined as a percentage of GFP-positive cells 24 hours after transfection with a plasmid expressing eGFP using a variety of commerically available transfection reagents and a range of reaction conditions as noted: FuGENE HD (A), Xfect (B), Amaxa Nucleuofection (C) and Lipfectamine 2000 (D). Images to the right of each chart show GFP-positive cells following transfection with optimal conditions for each reagent. E. Attempts to transfect cells with GFP protein yielded no GFP-positive cells using either Amaxa Nucleofection or Xfect and caused substantial cell death in the case of Amaxa nucleofection.

Optimal transfection efficiency (98%) was obtained by transfecting cells with Xfect using 2.5 μg of plasmid DNA (Figure 3-1B, right panel). Despite Amaxa nucleofection giving a lower transfection efficiency (80%) than Xfect or Lipofectamine 2000 (2.5μg DNA: 86%; Figure 3-1D) experiments using HT1080 cells described in chapters 4 and 6 were transfected using Amaxa nucleofection since this system does not require cells to be seeded the day before transfection, reducing experimental workflow by 1 day.

3.2.2 Protein Transfection

There is evidence that gene targeting achieved by transfection of a ribonucleoprotein (RNP) complex consisting of Cas9 protein and an *in-vitro* transcribed guide RNA is both efficient and accurate (Kim, Kim, Cho, Kim & J.-S. Kim 2014). Furthermore, RNP-mediated genome editing offers several advantages over plasmid-based delivery of CRISPR-Cas9: 1) Plasmid DNA transfected into cells can be randomly incorporated into the genome, or at on-target or off-target sites of Cas9-induced DSBs, allowing stable and persistent expression of Cas9 and/or gRNA. 2) Transfected DNA can trigger immune responses within cells and can cause stress to cells via activation of cyclic GMP-AMP synthases. 3) Plasmid DNA can persist within cells for several days after transfection, increasing the chances of off-target activity by Cas9. In contrast, both elements of the Cas9-gRNA RNP are degraded rapidly within cells, allowing only a small window for genome editing, reducing the chances of off-target effects (Kim, Kim, Cho, Kim & J.-S. Kim 2014).

In the event that we would try RNP-mediated genome editing, attempts were made to optimise protein transfection using recombinant GFP protein. For these experiments, transfection was attempted using Amaxa nucleofection (Lonza) and Xfect (Clontech). For transfection with Amaxa nucleofection (Figure 3-1E; left panel), 10^6 HT1080 cells were transfected with 10 μg GFP using programme L-005 and Kit T. For transfection with Xfect (Figure 3-1E; right panel), ~10^6 HT1080 cells were transfected with 2 μg GFP. For both methods no intracellular GFP could be seen 24 hours after transfection. Furthermore, nucleofection proved particularly toxic to cells. These results suggest

either it is not possible to transfect HT1080 with GFP, or that some component of the storage buffer or GFP itself is toxic to cells.

3.3 Using a reporter assay to test efficiency of gene targeting

3.3.1 The human HPRT gene functions as a reporter gene to assess gene targeting efficiency

Whilst transfection efficiency can be inferred by GFP plasmid transfection, this method suffers from the disadvantage that high levels of GFP within cells can prove toxic (Liu et al. 2001). To mitigate this problem and to take account of the requirement to optimize transfection in the context of gene targeting, we took advantage of a previously established reporter gene system (Gravells et al. 2015).

The reporter gene hypoxanthine phosphoribosyl transferase (*HPRT*) is located on the X chromosome and is involved in the salvage pathway of purine biosynthesis by catalyzing the transfer of the 5-phosphoribosyl group of 5-phosphoribosyl 1-pyrophosphate (PRPP) onto hypoxanthine to form inosine monophosphate (Figure 3-2A) (Stout & Caskey 1985; Caskey & Kruh 1979). Cells expressing functional HPRT can be selected in HAT medium (**H**ypoxanthine **A**minopterin **T**hymidine). In the presence of aminopterin which inhibits dihydrofolate reductase (DHFR), cell survival is dependent on functional HPRT to synthesize nucleotides via the salvage pathway (Figure 3-2B). Following targeted DSB formation within the *HPRT* gene and inaccurate repair by non-homologous end joining (NHEJ), cells that no longer have functional HPRT can be selected in medium containing 6-thioguanine (6-TG), a nucleoside analogue that is converted by HPRT into a toxic product (Figure 3-2C; Figure 3-2D).

Within exon 6 of the *HPRT* gene is a region that can be mutated to introduce the recognition sequence for the restriction endonuclease I-SceI such that there is only a single amino acid change in the resulting protein (ITLLSL>QTLLSL) that does not alter protein function (Gravells et al. 2015). Human HT1080 cells harbouring the I-SceI site within *HPRT* exon 6 are referred to as *HP-I-RT*. In these cells, the *HPRT* gene can be targeted with a DSB by I-SceI or by Cas9 (Figure 3-3A). Targeting efficiency can be

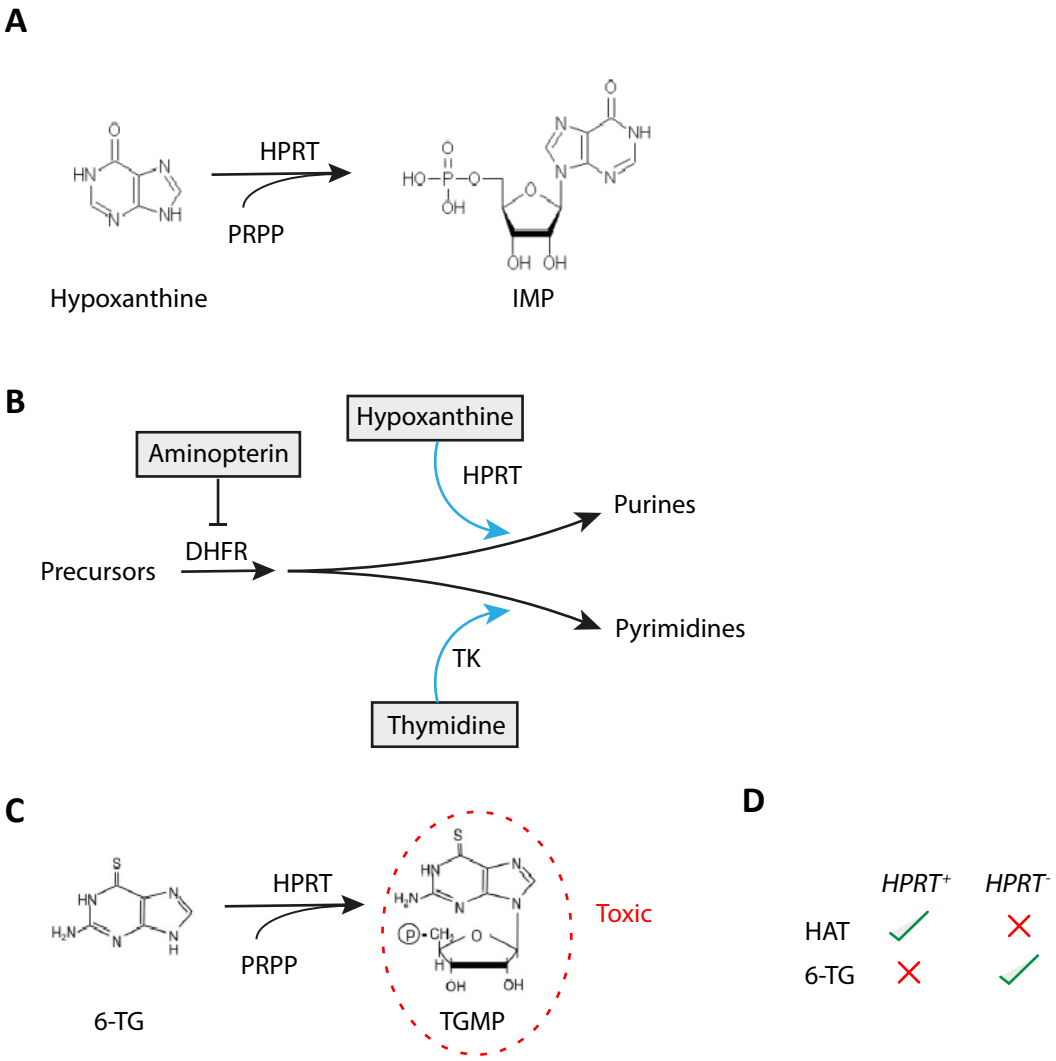


Figure 3-2 Explanation of the system used to determine whether cells contain a functional *HPRT* gene using 6-thioguanine and aminopterin

A. HPRT transfers the 5-phosphoribosyl group of 5-phosphoribosyl 1-pyrophosphate (PRPP) onto hypoxanthine to form inosine monophosphate. B. *De novo* synthesis (black arrows) of both purines and pyrimidines requires dihydrofolate reductase (DHFR). When DHFR is inhibited by aminopterin, purines and pyrimidines can still be synthesised via salvage pathways (blue arrows) that require hypoxanthine phosphoribosyl transferase (HPRT) and thymidine kinase (TK) respectively. C. HPRT transfers the 5-phosphoribosyl group of PRPP onto the purine analogue 6-thioguanine (6-TG) to form a cytotoxic product, thioguanine monophosphate (TGMP). D. Cells with functional HPRT grown in HAT medium are unable to synthesise nucleotides *de novo* but survive by using salvage pathways. In the absence of HPRT, cells grown in HAT medium die as DHFR is inhibited by aminopterin and salvage pathways are no longer functional. Cells with functional HPRT convert 6-TG into a cytotoxic product, TGMP and as such die when grown in 6-TG. In the absence of HPRT, cells grown in 6-TG survive as no TGMP is formed.

inferred from the proportion of cells that are able to grow in the presence of 6-TG following I-SceI- or Cas9-induced DSBs.

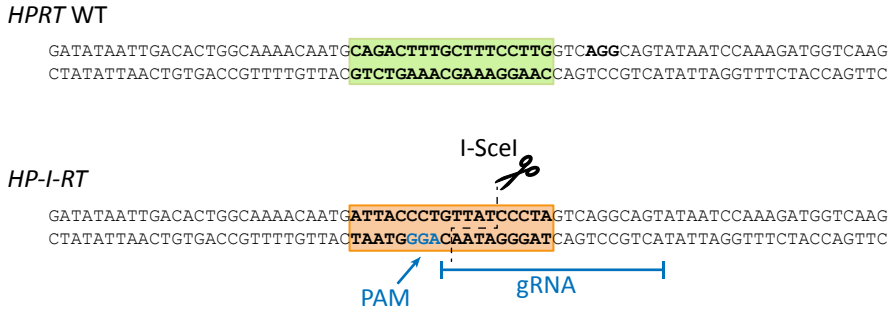
3.3.2 Inaccurate repair of I-SceI-mediated DSBs causes sensitivity to 6-TG

To test that this system would work in our hands, we carried out the I-SceI cleavage assay as described in Gravells et al., 2015. The initial gene targeting transfection optimization study involved transfecting HT1080 *HP-I-RT* cells with a plasmid expressing I-SceI using the same four commercially-available transfection reagents and conditions as described in section 3.2.1. 24 hours after transfection 500, 2000 and 5000 cells from each transfection condition were seeded into 10 cm dishes in growth medium without 6-TG. After a further 4 days, growth medium was supplemented with 6-TG (15 μg/ml). The 5-day delay between transfection and addition of 6-TG allows for depletion of existing HPRT protein in cells that have lost *HPRT* genetically following inaccurate repair of I-SceI-induced DSBs. After 10 days growth in 6-TG colony formation was assessed by staining with methylene blue (representative colony formation assay can be seen in figure 3-3B). Transfections were carried out in triplicate, seeding 500, 2000 and 5000 cells/10 cm dish from which mean cell survival (a measure of successful I-SceI-induced DSB formation and inaccurate repair by NHEJ) was calculated (Figure 3-4).

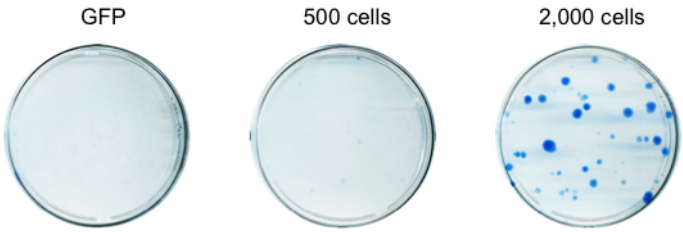
Transfection with FuGENE HD led to cell survival of 2.48% (1.5 μl reagent/1 μg,DNA); 2.01% (2 μl/μg); 1.9% (2.5 μl/μg); 1.89% (3 μl/μg); 7.13% (3.5 μl/μg) and 7.5% (4 μl/μg). Transfection with Lipofectamine 2000 led to cell survival of 2.03% (2 μl reagent/1 μg DNA); 11.11% (4 μl/μg); 6.98% (6 μl/μg); 3.97% (8 μl/μg) and 1.63% (10 μl/μg). Transfection with Amaxa nucleofection led to cell survival of 2.84%. Transfection with Xfect led to cell survival of 12.12% (1 μg DNA); 3.96% (1.5 μg DNA); 2.96% (2 μg DNA) and 3.95% (2.5 μg DNA).

For both FuGENE HD and Lipofectamine 2000, transfection efficiency of the plasmid expressing eGFP (Figure 3-1) correlates well with cell survival following I-SceI-induced DSB formation within *HPRT* and selection in 6-TG (Figure 3-2) across each of the

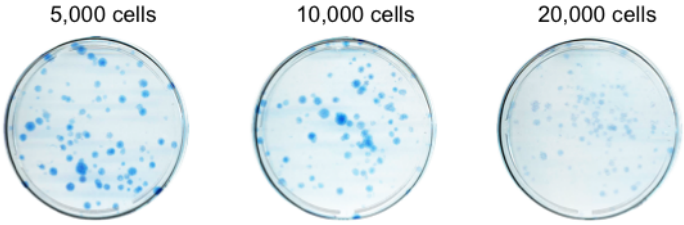
A



B



No. surviving colonies	0	7	52
% surviving colonies	0%	1.4%	2.6%



No. surviving colonies	104	116	TMTC
% surviving colonies	2.1%	1.2%	TMTC

Figure 3-3 Targeting *HPRT* exon 6 with I-SceI and CRISPR-Cas9

A. Mutation of the region shaded in green to the sequence shaded in orange introduces an I-SceI restriction enzyme site into exon 6 of the *HPRT* gene. The site of the I-SceI cut site is marked by the dashed black line and scissors. The guide RNA sequence used to induce double strand breaks by Cas9 is underlined in blue, with the corresponding PAM sequence highlighted in blue. B. Following transient expression of I-SceI, the stated numbers of HT1080$^{HP\text{-}I\text{-}RT}$ cells were plated in 10 cm² dishes and surviving colonies counted following growth for 10 days in 6-TG. Cells transfected with a plasmid expressing eGFP served as a negative control. Seeding 20000 cells per 10 cm² dish resulted in a colony density too high for accurate data to be obtained. Figure shows results from Lipofectamine 2000 (2 µl). TMTC = too many to count.

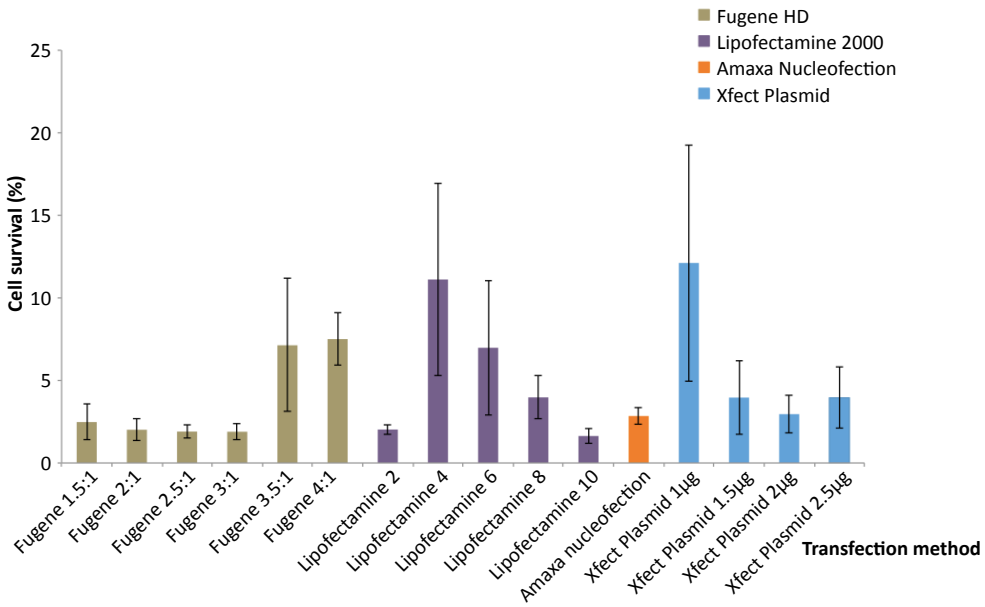


Figure 3-4 I-SceI-mediated transfection optimisation in HT1080 cells using the *HPRT* reporter gene

A. HT1080 cells were transfected with a plasmid encoding I-SceI using a range of commercially available transfection reagents then plated in varying numbers and their survival quantified as a percentage of total cells seeded per 10 cm² dish following treatment with 6-TG. Cell survival calculated as a mean of three independent biological repeats seeding 500, 2000 or 5000 cells using the stated transfection reagents and conditions. Error bars: SEM (n=3). Fugene ratio = Fugene HD reagent (µl):DNA mass (µg); Lipofectamine value = µl of Lipofectamine 2000 reagent.

individual conditions, suggesting that the efficiency of gene targeting is increased with efficient delivery of DNA into cells.

Whilst both Lipofectamine 2000 4µl (11.1% mean surviving colonies, standard error (SE): 5.84) and Xfect 1µg (12.1%, SE: 7.16) produced higher mean cell survival than FuGENE HD 4 µl reagent:1 µg DNA (7.5%; SE: 1.58), the latter was chosen to be used in CRISPR-Cas9 studies due to the smaller amount of variability (standard error) in I-SceI-mediated gene targeting efficiency.

3.2.4 I-SceI protein transfection

In addition to plasmid-mediated delivery of I-SceI, attempts were made to transfect I-SceI protein (NEB) into HT1080 cells using Xfect (~10^6 cells; 2 µg I-SceI) and Amaxa nucleofection (10^6 cells; 10 µg I-SceI; Kit T, programme L-005). Both methods proved highly toxic to cells with very little (Xfect) or no (Amaxa; data not shown) colony formation following 6-TG treatment. As with GFP protein transfection it was assumed that some component of the protein storage buffer was toxic to cells.

At this point having failed to transfect HT1080 cells with two different proteins it was decided that protein transfection would not be a viable option for gene targeting experiments and as such future experiments focused on gene targeting using plasmid-based methods.

3.2.5 Erroneous repair of Cas9-induced DSBs within *HPRT* causes 6-TG sensitivity

Having established that the *HPRT* reporter system works well in our hands as a measure of gene targeting efficiency and that optimum transfection efficiency for the HT1080 cell line can be achieved using FuGENE HD (4 µl reagent/1 µg DNA), it was important to identify the most effective way of delivering into HT1080 *HP-I-RT* cells the two components necessary for CRISPR-Cas9 gene targeting: a guide RNA (gRNA) and the Cas9 endonuclease. A gRNA was designed to target the region within exon 6 of *HPRT* containing the sequence mutated to the I-SceI recognition sequence (ACTGCCTGACTAGGGATAAC - Figure 3-3A; bottom). Cells were either co-transfected

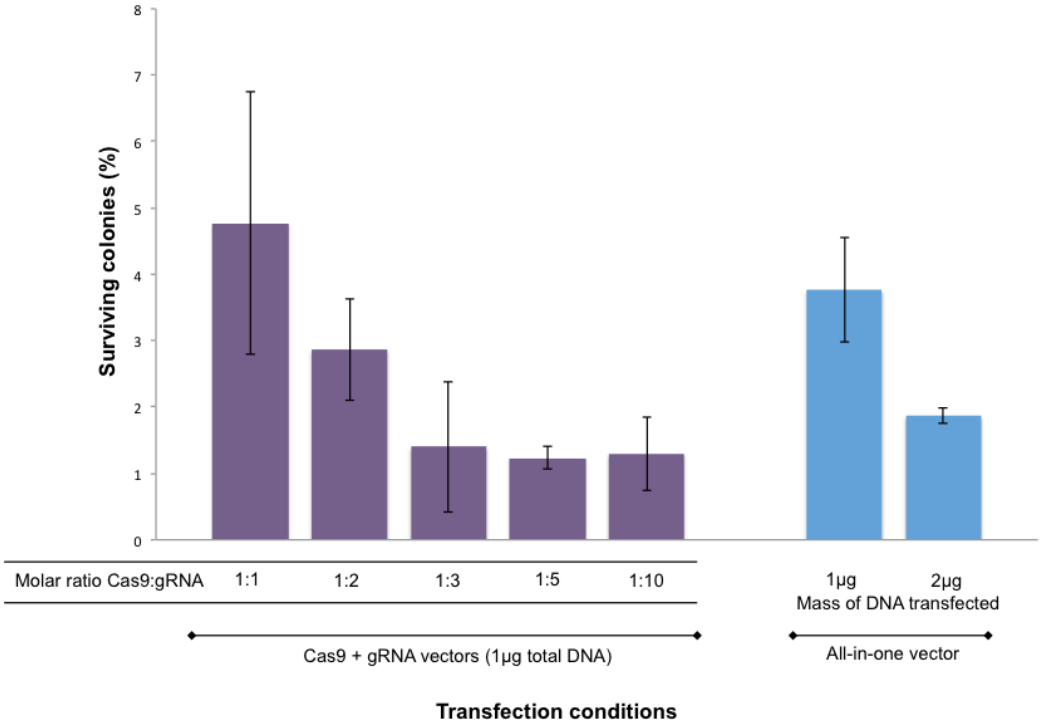


Figure 3-5 Optimisation of Cas9-mediated gene knockout in HT1080 cells using the *HPRT* reporter gene

HT1080 cells were transfected either with different molar ratios of Cas9:gRNA plasmid (1µg total DNA; purple bars) or an all-in-one Cas9+gRNA expression vector (1µg or 2µg DNA; blue bars) using a gRNA targeting exon 6 of *HPRT*. Following transfection, 1000 cells were seeded per 10 cm² dish and mean colony survival calculated from three biological replicates following growth in 6-TG for 10 days. Error bars: SEM (n=3).

with two separate plasmids, one expressing the gRNA and the other expressing human codon optimised Cas9 (both a gift from Dr. Andy Porter) in differing molar ratios of Cas9:gRNA (1:1; 1:2, 1:3, 1:5, 1:10) or with a single plasmid that expresses both gRNA and Cas9 (CRISPR nuclease vector, Geneart). ~5×10^5 HT1080 *HP-I-RT* cells were transfected with 1 μg total DNA using FuGENE HD (4 μl reagent/1 μg DNA).

24 hours after transfection 10^3 cells were seeded in 10 cm dishes. After a further 4 days, growth medium was supplemented with 6-TG (15 μg/ml). After 10 days growth in 6-TG, surviving colonies were stained with methylene blue and counted, with targeting efficiency scored from the number of colonies that grew as a percentage of the total number of cells plated. Mean survival was calculated from three independent experiments. Perhaps unsurprisingly, among cells transfected with two separate plasmids, transfection with equimolar quantities of separate gRNA and Cas9 plasmids led to the highest mean colony formation (4.77% ± standard error of the mean (SEM): 1.98) with lower targeting efficiency seen for molar excesses of gRNA vector (1:2: 2.87% ± SEM: 0.77; 1:3: 1.40% ± SEM: 0.98; 1:5: 1.23% ± SEM: 0.17; 1:10: 1.30% ± SEM: 0.56). When transfected with the CRISPR nuclease vector containing both gRNA and Cas9, targeting efficiency was higher when 1μg DNA was used (3.77% ± SEM: 0.79) than when transfected with 2μg DNA (1.87% ± SEM: 0.12) (Figure 3-5). This difference can presumably be accounted for by the increased toxicity associated with increased mass of transfected DNA (Sun et al. 2013).

In this section I have shown that I can successfully transfect and target the *HPRT* gene in human HT1080 cells using the I-SceI restriction endonuclease and CRISPR-Cas9. For gene targeting with CRISPR-Cas9 transfection of one single plasmid expressing both the guide RNA and Cas9 results in a greater gene targeting efficiency than expressing each component from separate plasmids, presumably due to the absence of requirement for two different plasmids to enter a cell for gene targeting to occur.

3.4 RNase H2A is essential in human TK6 cells

In parallel to the previous work a gRNA was designed using the online software CRISPRdirect (Naito et al. 2015) to knock out *RNASEH2A* in human lymphoblastoid TK6 cells using a strategy developed in the laboratory of Professor Shunichi Takeda at Kyoto University, Japan. TK6 cells have the advantage of being p53 positive and having a stable diploid karyotype (Lalle et al. 1995; Crompton et al. 2002). Furthermore, TK6 cells grow in suspension and therefore do away with the requirement for trypsinisation and have a relatively fast doubling time of ~13 hours making them easy to work with and allowing for faster workflow. In this system, gene knockout is achieved through the replacement of the first exon of a gene with a LoxP-flanked antibiotic resistance gene under the control of its own promoter. It is hoped that removal of the first coding exon is sufficient to prevent gene expression; it is however important to check that no splice isoforms exist for the gene that exclude exon 1. It is possible that in addition to targeting the first exon, knockout of gene function can be achieved by replacing an exon containing codons for amino acids key to catalytic function. In the case of *RNASEH2A*, the first exon is an excellent candidate for targeting since it not only contains the start codon but also contains the codon for glycine 37, an amino acid that plays an important role in catalysis (Crow, Leitch, et al. 2006). Furthermore the ensembl database (Zerbino et al. 2018) lists no splice isoforms for human *RNASEH2A*.

To attempt to knock out *RNASEH2A*, 10^6 TK6 cells were co-transfected using Neon electroporation (Invitrogen; 1350 V, 10 ms, ×3 pulse, 10µl Neon tip) with two plasmids, one to co-express both Cas9 and a gRNA (pX330; 200 ng) and one to be used as a template for homology directed repair following a Cas9-induced DSB (800 ng). The gRNA was designed such that Cas9 was targeted within *RNASEH2A* exon 1. The repair template was designed with ~1 kb arms homologous to regions within ~100 bp either side of the gRNA sequence, such that upon repair, the entire first exon would be excluded. Furthermore, the design of the repair template excludes the gRNA sequence to prevent cutting of the plasmid by Cas9 (Figure 3-6). The use of a plasmid to be used as a template for homology directed repair and resultant incorporation of an antibiotic resistance gene offers the advantage of screening by antibiotic selection rather than the

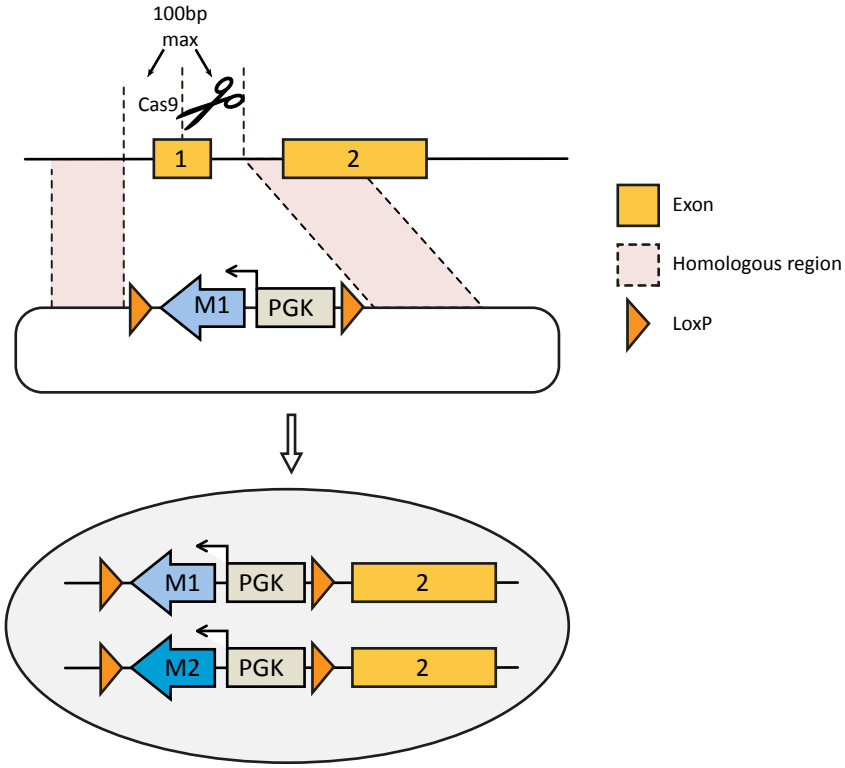


Figure 3-6 *RNASEH2A* knockout strategy
A Cas9-induced double strand break in exon 1 is repaired by homology-directed repair such that exon 1 is replaced by a cassette containing an antibiotic resistance gene (M1/M2) driven by a PGK promoter, flanked by two LoxP sites allowing subsequent marker removal via expression of Cre recombinase.

potentially more time-consuming screening required when gene knockout is achieved by inaccurate repair of a Cas9-induced DSB by non-homologous end joining.

Cells can be transfected with either one repair template with a single antibiotic resistance gene or two repair templates containing different antibiotic resistance genes. Selection with two different antibiotics increases the chances that clones picked up after selection are homozygously targeted, although since it is thought that the efficiency of targeting one single allele of a gene in TK6 cells using this system is ~1:10^3, the use of two different selection markers reduces the efficiency of targeting to ~1:10^6. Furthermore, whilst selection with only one antibiotic increases the chances of picking up heterozygously targeted clones, it does not preclude the possibility of picking up clones targeted homozygously with the same antibiotic resistance gene present at both alleles.

Candidate knockout cells were generated in the Takeda laboratory using the parameters stated above. We were provided with one candidate clone that had been selected using one single antibiotic resistance marker and screened by PCR to verify the presence of the marker at the correct locus (data not shown). The PCR was carried out using a forward primer that annealed within the antibiotic resistance gene and a reverse primer that annealed outside the sequence used for the right homology arm. Once we had received the cells, this clone was screened by immunoblotting using an antibody against RNase H2A (Figure 3-7) but the presence of a band at ~32 kDa suggests that this clone expresses RNase H2A and is therefore assumed to only be targeted heterozygously. The similarity in protein level between wild type and the clone that we presumed to have lost a single copy of *RNASEH2A* can be explained in one of three ways: 1) the plasmid used for homology directed repair of the Cas9-induced DSB randomly integrated somewhere in the genome conferring G418 resistance to cells and the PCR used to screen for the presence of the knockin construct at the correct locus somehow produced an amplicon of the correct size; 2) the knockout is indeed heterozygous and the intensity of the band in the knockout candidate clone is lower than wild type but the Western blot is overexposed or 3) the knockout is heterozygous but protein levels remain the same due to transcriptional upregulation of the remaining copy of

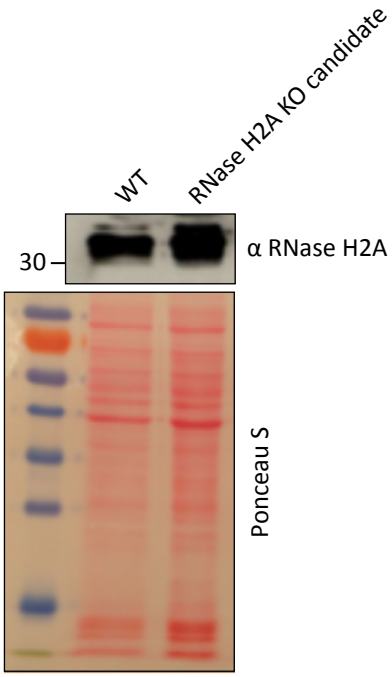


Figure 3-7 RNase H2A knockout candidate TK6 cells are only targeted heterozygously

Immunoblot of wild type and RNASEH2A knockout candidate TK6 cells probed with an anti-RNase H2A antibody shows that the candidate knockout cells still express RNase H2A protein (~32 kDa) and as such are presumed to be only targeted heterozygously.

RNASEH2A. Whatever the reason for the band intensity seen at the size corresponding to RNase H2A in the Western blot, it was clear that these cells would be no use to us.

Having discovered that the cells provided to us by the Takeda laboratory were not homozygously knocked out for RNase H2A, another two attempts were made to knock the gene out using one or two repair templates to attempt to pick up heterozygous or homozygous knockouts respectively. 10^6 TK6 cells were transfected as previously described with either 200 ng pX330 and 800 ng $G418^R$ repair template or 200 ng pX330 and 400 ng each of $G418^R$ and $hygromycin^R$ repair templates for selection of heterozygous or homozygous knockout respectively. Two days after transfection, cells were selected in 1 mg/ml G418 (heterozygous targeting) or 1 mg/ml G418 and 0.625 mg/ml hygromycin (homozygous targeting). In both instances, individual colonies started to grow but after approximately 5 or 6 rounds of mitosis, no further cell division was seen and the cells ultimately died, suggesting that RNase H2A function is indeed essential for cell viability.

It is possible that the failure to generate cells in which RNase H2A was knocked out may have been due to low transfection efficiency resulting in only a small proportion of transfected cells taking up the required plasmid DNA. Indeed concomitant transfection of a plasmid expressing GFP to act as a control for transfection showed very low transfection efficiency for all four transfections described in the previous section when using the same transfection reagent and parameters as used in the Takeda laboratory (Neon; 1350 V, 10 ms, ×3 pulse, 100µl Neon tip). To attempt to increase transfection efficiency a transfection optimisation was carried out using a range of parameters defined by the manufacturer. Cells were transfected with a plasmid expressing GFP and transfection efficiency scored as a percentage of GFP-positive cells 24 hours post-transfection. Transfection parameters resulting in >20% GFP-positive cells were further qualitatively scored for cell health based on morphology and survival post transfection (Table 3-1). The highest transfection efficiency (29.1% GFP-positive) was obtained using a pulse voltage of 1400 V, pulse width of 30 ms and a pulse number of 1, however due to the decreased toxicity to cells it was decided to use conditions of 1500 V, 20 ms, × 1

Pulse Voltage	Pulse Width	Pulse No.	% GFP Positive
0	0	0	0.0
850	30	2	10.1
1500	10	3	10.9
1400	10	3	11.0
1100	40	1	11.5
1600	10	3	12.0
950	30	2	12.4
1150	30	2	13.5
1050	30	2	14.7
1000	40	1	14.9
1200	40	1	14.9
1100	30	1	15.8
1100	20	2	17.1
1300	20	2	18.3
1300	30	1	18.5
1300	10	3	19.4
1200	20	2	22.0
1700	20	1	22.3
1600	20	1	24.0
1200	30	1	24.0
1400	20	1	26.4
1500	20	1	27.2
1400	20	2	29.0
1400	30	1	29.1

Table 3-1. TK6 Neon transfection optimisation.
TK6 cells (10^5/condition) were transfected with a plasmid expressing GFP using Neon electroporation (10 µl tip) under the stated transfection parameters. The percentage of GFP-positive cells 24 hours after transfection was used as a measure of transfection efficiency. Transfection parameters causing >20% efficiency were scored qualitatively based upon cell morphology to take into account the effect of the electroporation conditions on cell health 24 hours after transfection. Green indicates an abundance of cells displaying normal morphology and low numbers of dead cells. Amber indicates that the electroporation conditions proved moderately harmful to cells, whilst red indicates that the transfection conditon caused high levels of cell sickness and death.

pulse for further TK6 transfections despite the slightly lower transfection efficiency (27.2% GFP-positive).

A further unsuccessful attempt was made to knockout *RNASEH2A* either heterozygously or homozygously using the optimised electroporation parameters, however no cells grew after 10 days in selective media (G418 only or G418 and hygromycin for heterozygous and homozygous targeting respectively).

The results of our attempts to knockout RNase H2A in human TK6 cells and since attempts to generate RNase H2 knockout mice by independently targeting either the B or C accessory subunits show that RNase H2 deficiency causes embryonic lethality and is therefore not compatible with life in mammals (Reijns et al. 2012; Hiller et al. 2012b) and that there are no known loss-of-function pathogenic RNase H2 mutations (Reijns et al. 2012; Crow, Leitch, et al. 2006) we assume that *RNASEH2A* is an essential gene in human cells.

3.5 DISCUSSION

In this chapter I have created a protocol for gene knockout using CRISPR-Cas9. Whilst gene knockout can be achieved through transfection of a single plasmid expressing both Cas9 and gRNA, relying on gene knockout through inaccurate repair of Cas9-induced DSBs by NHEJ, this system does have limitations in terms of screening for successfully targeted clones. Current strategies to screen for NHEJ-mediated gene knockout can be time consuming requiring optimisation and can be sequence-limited (Hua et al. 2017) and do not take account of the possibility that repair may exclude one or more codons but that functional protein may still be expressed. Gene knockout by replacing an entire exon with an antibiotic resistance marker as we did with *RNASEH2A* offers the advantage of antibiotic selection but has the disadvantage of not being seamless since even after marker removal by expression of Cre recombinase, the LoxP sequence remains in the genome.

In this chapter I have also shown that different transfection reagents show differing levels of transfection efficiency and furthermore that the same transfection reagent and conditions used with the same cell line may vary in efficiency in different laboratories. This is probably particularly true in the case of electroporation platforms which may not necessarily treat the cells with exactly the parameters that are set. It is therefore advisable that before genome editing is attempted in an unfamiliar cell line that a transfection optimisation is carried out using a range of transfection reagents and conditions.

I have shown that whilst it may be possible to knockout one copy of *RNASEH2A* in TK6 cells, heterozygous knockout is insufficient to significantly reduce overall cellular protein levels. It would have been advantageous to try to knockout *RNASEH2A* in other cell lines to see if the presumed upregulation seen in heterozygous knockout cells was specific to TK6 cells, but the existing evidence for the requirement for RNase H2A for cell viability meant that further research in this area would not have proved fruitful.

In conclusion, I have shown that homozygous knockout of RNASEH2A in the human TK6 cell line is not possible and since we presume this to be the case across most, if not all human cell lines, it will be necessary to regulate expression of the RNase H2 complex. Work carried out to attempt to regulate RNase H2 can be seen in results chapter 2.

Chapter 4

Regulating RNase H2 in human cells

4.1 Introduction

Due to the essential nature of RNase H2A in TK6 cells it was necessary to develop a system to induce the rapid shutoff of RNase H2 function. Several mechanisms have been used to downregulate production of specific proteins by either transcriptional shutoff of the gene of interest, by degrading the messenger RNA (mRNA) encoding the protein or by targeting the protein itself post-translationally.

In this chapter I will discuss three mechanisms of downregulating expression of the different subunits of human RNase H2: auxin-inducible degron (AID), a novel double degron tag composed of a mini AID degron (mAID) tag fused with a small molecule-assisted shutoff (SMASh) tag and a SMASh-tagged RNase H2 subunit under control of tetracycline-regulated transcriptional expression.

Finally, I will show that RNase H2C can be strongly downregulated via a C-terminal mAID-SMASh double degron tag.

4.2 Assessing the suitability of all three RNase H2 subunits for tagging in *Schizosaccharomyces pombe*

Tagging proteins can often be problematic due to the possibility of negatively affecting protein folding or of disrupting a protein domain required for the stability of a multi-subunit protein complex. To identify the suitability of each of the three RNase H2 subunits for tagging, we carried out an initial study in *S. pombe*. We hypothesized that if we were able to separately tag the C-termini of each of the three RNase H2 subunits with GFP in *S. pombe* without affecting the ribonuclease activity of the holoenzyme,

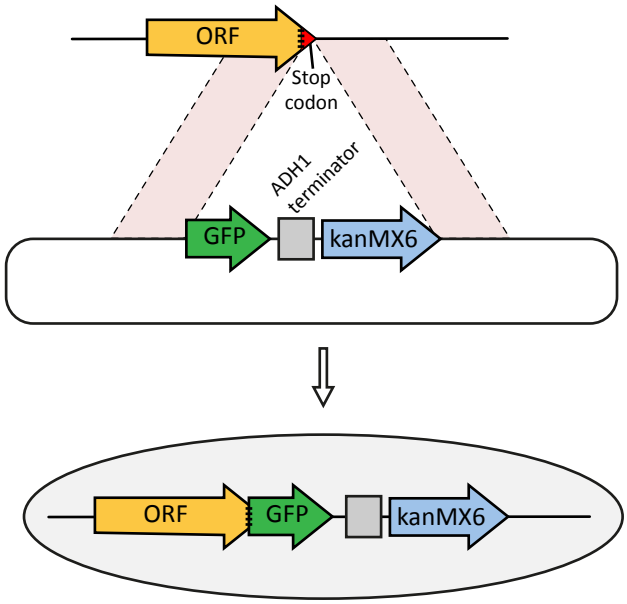


Figure 4-1 C-terminal GFP tagging strategy in *S. pombe*
A GFP tag is added to the C-terminus of RNase H2 subunits via recombination with a plasmid which excludes the endogenous stop codon. The plasmid also confers kanamycin resistance (kanMX6) allowing for selection of positively targeted colonies.

then there was a good chance that the presence of C-terminal tags on the human RNase H2 subunits would not affect catalytic activity of the enzyme. Similarly, it was presumed that any decrease in catalytic activity would likely have been due to the destabilizing effect of the tag on trimer formation.

4.2.1 Tagging each of the three subunits of RNase H2 with a C terminal GFP

In *S. pombe*, the three subunits of RNase H2, are encoded by the genes *rnh201*, *rnh202* and *rnh203* coding for RNase H2A, RNase H2B and RNase H2C respectively. A strain of *S. pombe* containing C-terminally GFP tagged *rnh201* already existed. To introduce C-terminal GFP tags to *rnh202* and *rnh203*, a cassette containing GFP and a kanamycin resistance gene (KanMX6) was amplified by polymerase chain reaction from a plasmid (Bahler et al. 1998) using forward and reverse primers containing ~100 bp extensions containing sequences homologous to the regions to be targeted. The homologous regions were designed such that upon integration of the cassette at the C-terminus of each of the genes the endogenous stop codon would be excluded (Figure 4-1). 10^8 cells were transformed with 10 μg purified PCR product using lithium acetate and DMSO and selected on kanamycin plates.

To screen for the presence of the GFP tag on *rnh202* and *rnh203*, gDNA was extracted from 5 and 8 colonies respectively and used as a template for PCR using a forward primer that annealed within the KanMX6 fragment and a reverse primer that annealed outside the right homology arm. 4 of the 5 *rnh202*-GFP candidates and 8 out of 8 *rnh203*-GFP candidates screened were found to have the GFP tag integrated at the correct locus (Figure 4-2 A). To further verify the presence of the GFP tag, 2 colonies of each strain (*rnh201*-GFP, *rnh202*-GFP & *rnh203*-GFP) were analysed by immunoblotting using an antibody against GFP (Figure 4-2 B). All colonies screened were found to contain C-terminal GFP tags on each of the three RNase H2 subunits.

In order to assess the catalytic activity of the RNase H2 holoenzyme containing a C-terminal GFP tag on each of the three subunits, we used a system previously established in the Carr laboratory (unpublished) that assesses the ability of RNase H2 to remove ribonucleotides misincorporated into gDNA at higher than wild type levels by a mutant

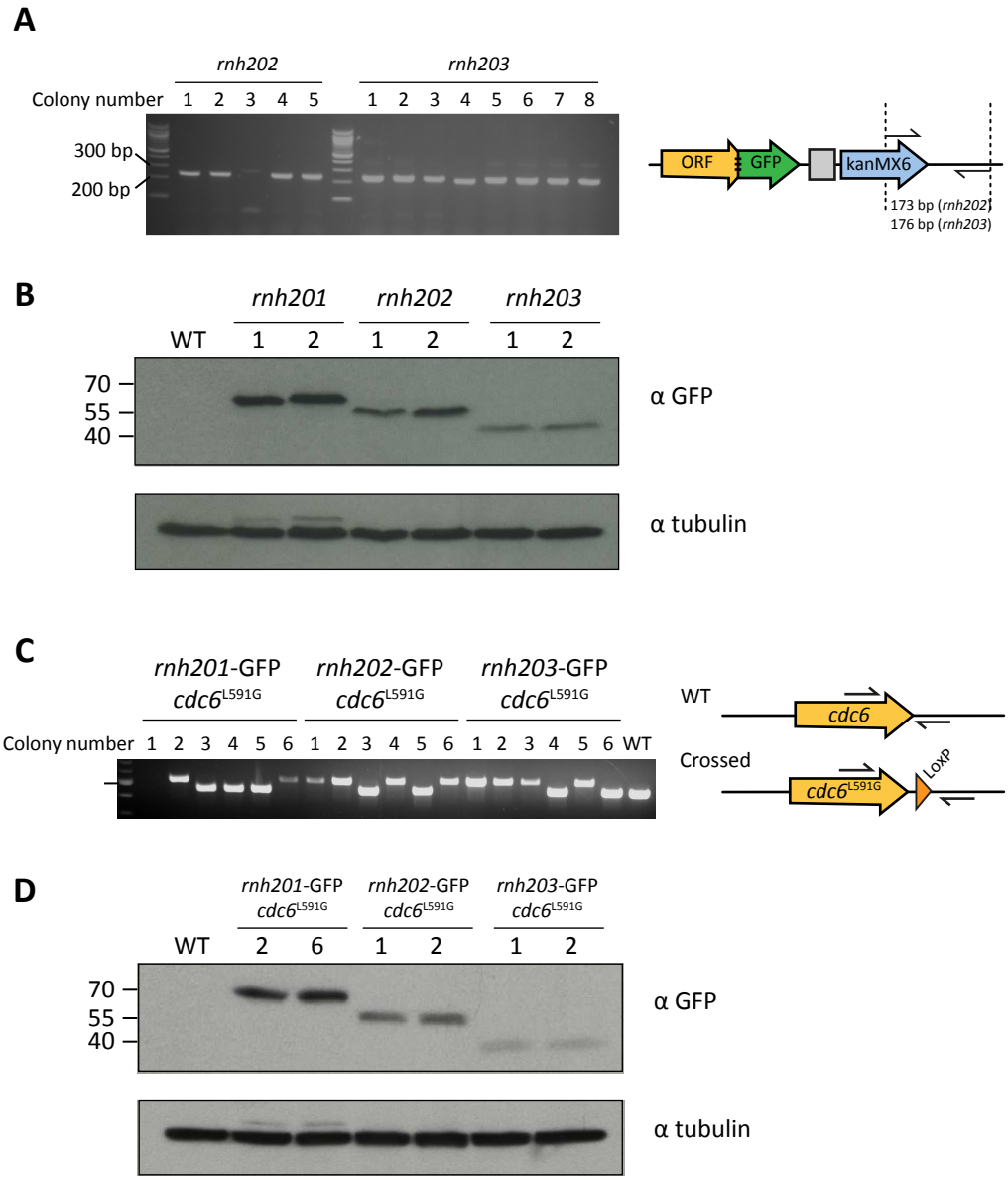


Figure 4-2 The C-termini of the three subunits of *S. pombe* RNase H2 can all be tagged with GFP in a *cdc6*^{L591G} background

A. Following transformation with the C-terminal GFP tagging plasmid and selection on kanamycin, colonies were screened by PCR to check for the presence of the kanMX6 marker at the correct locus using a forward primer that anneals within, and a reverse primer that anneals outside kanMX6. B. Immunoblot using an anti-GFP antibody shows that all three subunits express a C-terminal GFP tag. C. After crossing with strains containing mutant *cdc6*^{L591G}, resultant colonies were screened by PCR for the presence of *cdc6*^{L591G} by PCR, using a forward primer that anneals within, and a reverse primer that anneals downstream of the gene. The presence of a LoxP site initially used to integrate the mutant *cdc6* causes a band shift and allows for selection of positive colonies. D. Colonies that contained *cdc6*^{L591G} were screened by immunoblotting using an anti-GFP antibody to ensure the presence of the GFP-tagged RNase H2 subunit.

DNA polymerase δ containing an L591G mutation at the catalytic site. To introduce the mutant polymerase δ, encoded by the gene *cdc6*, strains of *S. pombe* containing the mutant polymerase were crossed with strains containing each of the C-terminally tagged RNase H2 subunits. To screen for the presence of the mutant polymerase δ in the crossed strains, random spore analysis was carried out by PCR using a forward primer that annealed within the *cdc6* gene and a reverse primer that annealed downstream of a LoxP site that had been previously used to introduce the mutant *cdc6* gene (Figure 4-2 C). The presence of the LoxP site downstream of the mutant *cdc6*^{L591G} gene causes an increase in amplicon size in comparison to amplicons from gDNA containing wild type *cdc6*. 2 out of 6 *rnh201*-GFP colonies screened were found to contain *cdc6*^{L591G} (colonies 2 & 6) and 4 out of 6 colonies screened for both *rnh202*-GFP and *rnh203*-GFP were found to contain *cdc6*^{L591G} (colonies 1, 2, 4 & 6 and colonies 1, 2, 3 & 5 respectively).

To verify that these colonies containing *cdc6*^{L591G} still contained the desired GFP-tagged RNase H2 subunits, 2 colonies of each were screened again by immunoblotting using an antibody against GFP and all colonies screened were found to still have GFP-tagged *rnh201*, *rnh202* or *rnh203* (Figure 4-2 D).

4.2.2 Assessing the effect of C-terminal GFP tags on each of the three subunits on catalytic activity of the RNase H2 holoenzyme

The presence of the 2'-OH on the ribose moiety of a ribonucleotide within genomic DNA renders the sugar phosphate backbone at that point five times more susceptible to alkaline hydrolysis than bonds between adjacent deoxyribonucleotides (Clausen et al. 2015). As such, the level of RNase H2 activity can be inferred from the extent of smearing of genomic DNA due to the presence of unrepaired genomic ribonucleotides when subjected to alkaline hydrolysis and agarose gel electrophoresis.

To assess the extent to which catalytic activity of the RNase H2 trimer is affected when one of the subunits contains a C-terminal GFP tag we compared the extent of smearing of gDNA following alkaline hydrolysis and agarose gel electrophoresis from strains containing C-terminal GFP tags on each of the three RNase H2 subunits in combination with the *cdc6*^{L591G} allele with gDNA from a wild type strain of *S. pombe* and strains

containing the *cdc6*^{L591G} allele with untagged RNase H2 subunits, strains containing the *cdc6*^{L591G} allele with one of the RNase H2 subunits knocked out and with strains with wild type *cdc6* with GFP-tagged RNase H2 subunits.

Since smearing of gDNA when run on agarose gels following alkaline hydrolysis could appear to be greater if the gDNA had been damaged during the purification process, we subjected untreated gDNA from each of the strains to agarose gel electrophoresis to assess the integrity of extracted gDNA (Figure 4-3 A-C). With the exception of gDNA extracted from strain 1392 (*cdc6*^{WT} *rnh203*-GFP; Figure 4-3 C), gDNA from all strains used for this study appeared undamaged by the purification process. Since strain 1392 would not be expected to have excessive amounts of genomic ribonucleotides due to the presence of wild type polymerase δ, the data from this strain were less informative to us than that provided by the strain containing the *cdc6*^{L591G} allele in which *rnh203* had been knocked out and taking into consideration a separate trial that had shown RNase H2 catalytic activity in this strain to be unaffected by GFP-tagged *rnh203* (data not shown), we decided that we would be able to infer the information we needed despite the poor gDNA integrity of this strain.

The gel images in Figure 4-3 A-C were used to calculate the concentration of extracted gDNA by comparing the signal intensity of each band with a known mass of DNA in the DNA size marker using ImageJ (Schneider et al. 2012), such that equal masses of gDNA could be subjected to alkaline hydrolysis. To assess the extent of unrepaired genomic ribonucleotides, gDNA from each of the strains was either treated with 0.3M KOH at 55°C for 1 hour or left untreated and subjected to 2% agarose gel electrophoresis (Figure 4-3 D-F). To ensure that signal from alkaline-treated samples would be visible on the gel in the event of high levels of alkaline hydrolysis, three times more gDNA was alkali-treated and run on agarose gels than was run for the untreated samples. Furthermore, this was an important control since alkaline hydrolysis can also occur stochastically between adjacent deoxyribonucleotides such that a certain level of smearing is expected even with gDNA from wild type cells.

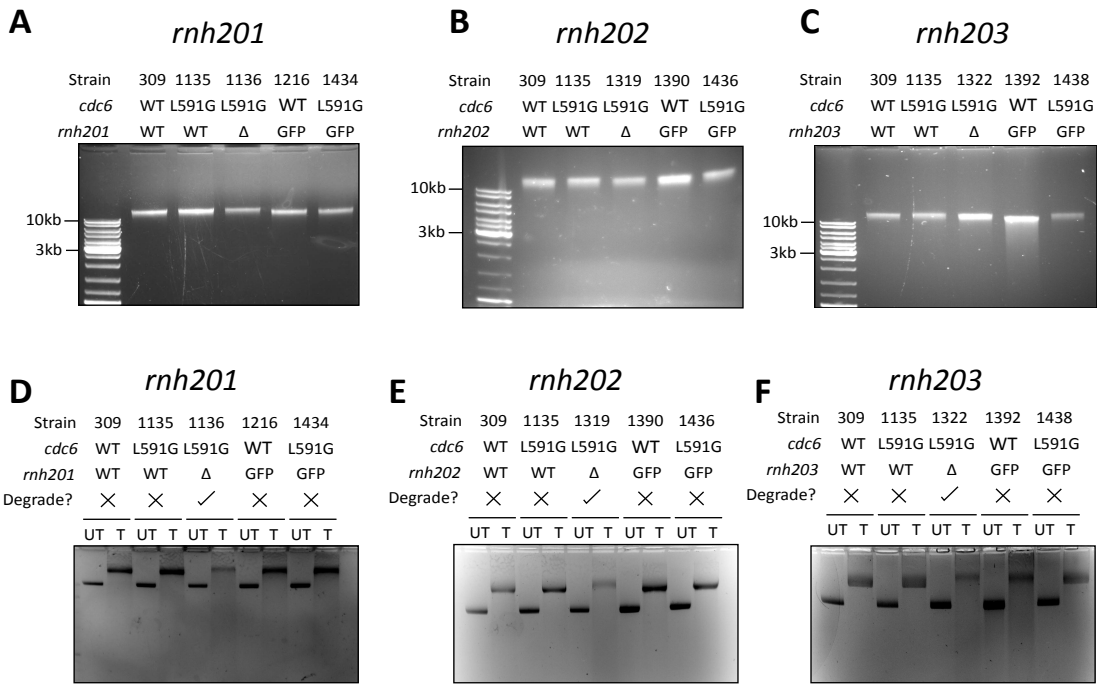


Figure 4-3 The presence of a C-terminal GFP tag on each of the three subunits of RNase H2 in *S. pombe* has no discernable effect on enzyme activity

A-C. Genomic DNA (gDNA) from cells with the stated genotypes was run on agarose gels to infer the integrity of the gDNA, with greater smearing indicating greater shearing of the gDNA during the extraction process. D-F. gDNA from the indicated strains was either left untreated (UT) or treated with alkali (T) and run on agarose gels. Decreased RNase H2 function can be inferred from increased gDNA smearing following treatment with alkali due to elevated ribonucleotide levels in gDNA resulting in increased hydrolysis of gDNA. GFP-tagged *rnh201* (D), *rnh202* (E) and *rnh203* (F) in combination with ribo-incorporating *cdc6*^{L591G} show similar levels of degradation to wild type suggesting that the RNase H2 holoenzyme retains full function.

In wild type cells, the low levels of ribonucleotides incorporated by polymerase δ are rapidly repaired by RNase H2 and as such signal intensity between bands for untreated and treated are equal. Cells containing the *cdc6*^{L591G} allele are expected to incorporate higher than wild type levels of ribonucleotides but in the presence of wild type RNase H2 these ribonucleotides are rapidly repaired. Accordingly, the signal intensity for untreated and treated gDNA from this strain is equal on all three gels (Figure 4-3 D: *cdc6*^{L591G} *rnh201*^{WT}; Figure 4-3 E: *cdc6*^{L591G} *rnh202*^{WT}; Figure 4-3 F: *cdc6*^{L591G} *rnh203*^{WT}). In the absence of any of the three RNase H2 subunits, catalytic activity of the RNase H2 is reduced such that in strains additionally harboring the *cdc6*^{L591G} allele, a large level of gDNA smearing is seen following alkaline hydrolysis (Figure 4-3 D: *cdc6*^{L591G} *rnh201*Δ; Figure 4-3 E: *cdc6*^{L591G} *rnh202*Δ; Figure 4-3 F: *cdc6*^{L591G} *rnh203*Δ). The presence of C-terminal GFP tags on each of the three RNase H2 subunits in an otherwise wild type background has no effect on smearing (Figure 4-3 D: *cdc6*^{WT} *rnh201*-GFP; Figure 4-3 E: *cdc6*^{WT} *rnh202*-GFP; Figure 4-3 F: *cdc6*^{WT} *rnh203*-GFP). There is no discernible difference in signal intensity between untreated and treated gDNA from strains harboring a GFP tag on each of the three RNase H2 subunits in combination with the *cdc6*^{L591G} allele, indicating that the presence of the GFP tag has little or no impact on catalytic activity (Figure 4-3 D: *cdc6*^{L591G} *rnh201*-GFP; Figure 4-3 E: *cdc6*^{L591G} *rnh202*-GFP; Figure 4-3 F: *cdc6*^{L591G} *rnh203*-GFP). As such, it was decided that theoretically all three RNase H2 subunits could be tagged at the C-terminus without affecting catalysis.

4.3 Regulation of human RNase H2A with an auxin inducible degron tag.

As described in section 1.4.3 human RNase H2 is a heterotrimer consisting of a catalytic subunit, RNase H2A and two accessory subunits RNase H2B and RNase H2C. Since the sustained presence of misincorporated ribonucleotides (rNTPs) within genomic DNA is an essential principle of Polymerase Usage sequencing (PU-seq), we decided in the first instance to attempt to tag human RNase H2A with a degron tag which would allow a temporary knockdown of RNase H2A and a subsequent decrease in RNase H2 catalytic activity. Furthermore, temporal ablation of RNase H2 activity offers the

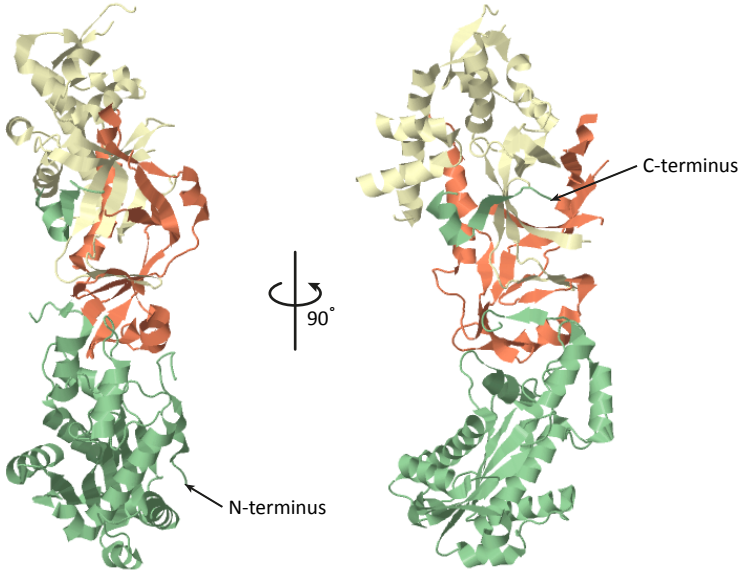


Figure 4-4 Crystal structure of the human RNase H2 holoenzyme
Shown is the overall structure of the human RNase H2 complex coloured by subunits (green, A; yellow, B; orange, C) with the N- and C-termini of RNase H2A indicated, imaged using JSmol (Hanson *et al.*, 2013). Structure from PDB ID: 3PUF (Figiel. *et al.*, 2011).

advantage that upon release from knockdown, cellular levels of RNase H2 activity would return to normal thereby allowing removal of misincorporated rNTPs. In contrast complete knockout of RNase H2 function, as used in *S. pombe* PU-seq, does not allow for the repair of genomic rNTPs between experiments, increasing the chances of ribonucleotide-derived DNA damage and increasing the background ribonucleotide signal. It is advantageous to be able to remove genomic rNTPs since the sustained presence of genomic rNTPs over time can cause widespread genome instability in yeast, mice and human cells (Nick McElhinny, Kumar, et al. 2010; Hiller et al. 2012b; O'Connell et al. 2015; Pizzi et al. 2015; Reijns et al. 2011).

In silico analysis of the crystal structure of the RNase H2 heterotrimer revealed that RNase H2A has a free N-terminus (Figure 4-4) whilst the C-terminus appeared less accessible due to its requirement in stabilising the formation of the holoenzyme through its interactions with the RNase H2B-RNase H2C heterodimer. Furthermore this interaction mediated by the C-terminus of RNase H2A is essential for enzymatic function (Reijns et al. 2011). However, due to our discovery that a C-terminal GFP tag on RNase H2A has no discernible effect on RNase H2 catalysis in *S. pombe*, and due to the increased difficulty in tagging the N-terminus of proteins, it was decided as a starting point to attempt to tag the C-terminus of RNase H2A with an auxin inducible degron (AID) tag (Nishimura et al. 2009). This decision was also taken since a strategy for tagging the C-termini of a range of different proteins with AID had already been established in the Takeda laboratory where much of this preliminary work was carried out.

Auxins are a widespread family of plant hormones, the most common of which is indole-3-acetic acid (IAA) which is responsible for the regulation of many genes involved in plant development (Teale et al. 2006). In plants, auxin family hormones cause the degradation of a family of transcriptional repressors known as Aux/IAAs by mediating an interaction between the Aux/IAAs and the F-box transport inhibitor response 1 (TIR1) subunit of the SCF^{TIR1} E3 ubiquitin ligase complex. SCF^{TIR1} then recruits an E2 ubiquitin conjugating enzyme which polyubiquitinates the SCF^{TIR1}-associated Aux/IAA, marking it for proteasomal degradation (Figure 4-5) (Abel & Theologis 1996; Teale et al. 2006; Tan et al. 2007). This system has been reconstituted in yeast and human cells to induce

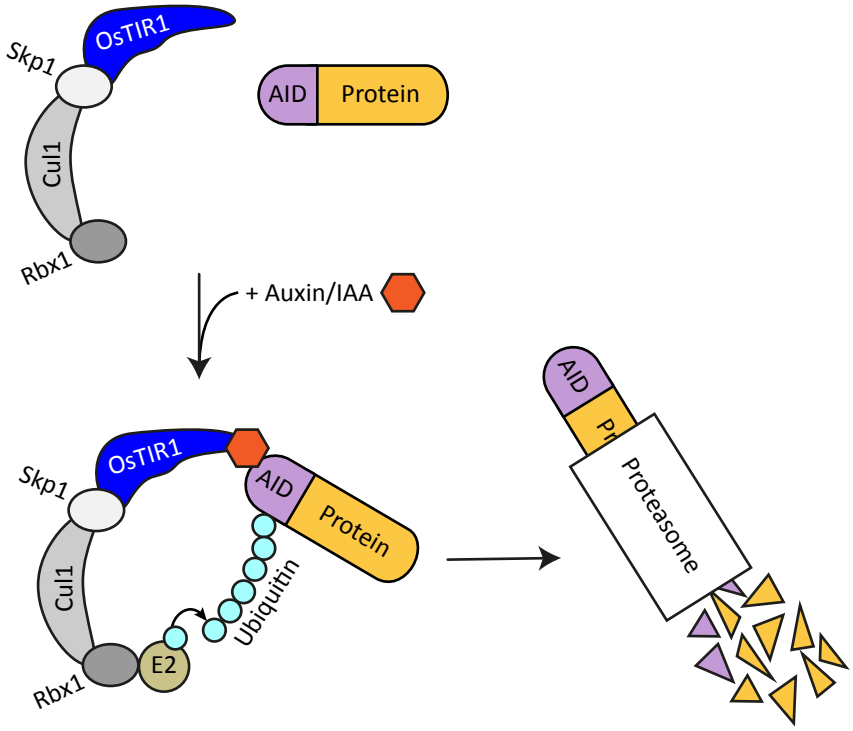


Figure 4-5 Schematic illustration of the AID system
Auxin and auxin-family hormones such as indole-3-acetic acid (IAA) promote the interaction between the AID degron tag (AID) of the target protein and OsTIR1. The OsTIR1-SCF complex acts as an E3 ubiquitin ligase to recruit an E2 ubiquitin ligase which polyubiquitinates the degron tag, marking the target protein for proteasomal degradation.

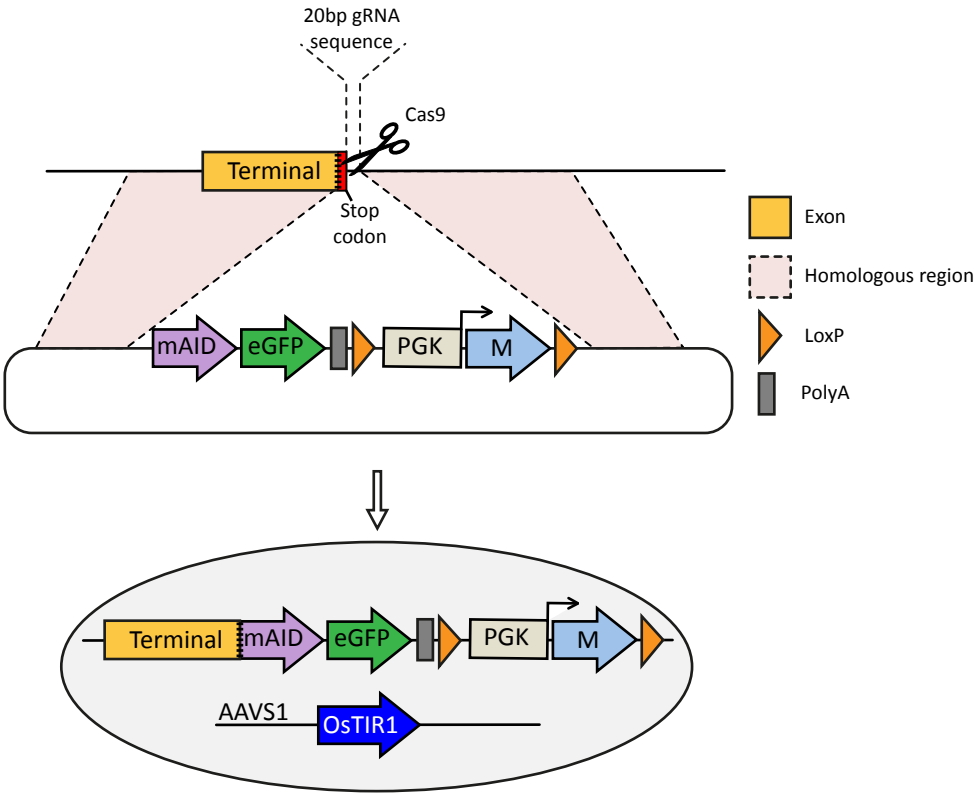


Figure 4-6 *RNASEH2A* C-terminal AID tagging strategy
A Cas9-induced double strand break is targeted as close as possible after the stop codon of the terminal coding exon of the gene. The break is repaired by homology directed repair using a repair template designed such that both the endogenous stop codon and the 20bp gRNA sequence are excluded upon repair. OsTIR1 is ectopically expressed constitutively from the AAVS1 locus.

degradation of specific proteins by using *Arabidopsis thaliana* IAA17 as a degron, fused with the target protein in the presence of IAA and *Oryza sativa* TIR1 (OsTIR1) (Nishimura et al. 2009; Natsume et al. 2016).

We decided to tag the C-terminus of RNase H2A with a mini AID degron tag fused to eGFP in human TK6 cells. To target the tag to the C-terminus of RNase H2A, 10^6 TK6 cells expressing OsTIR1 from the AAVS1 locus were co-transfected using Neon electroporation (Invitrogen; 1350 V, 10 ms, ×3 pulse, 10µl Neon tip) with two plasmids, one to express both Cas9 and a gRNA (pX330) and the second, a repair template. For each transfection, cells were transfected with 200 ng pX330 and 800 ng repair template. The gRNA was designed to target a Cas9-induced DSB to as close as possible to the stop codon of *RNASEH2A*. The repair template was designed with homology arms that would exclude the 20bp gRNA sequence upon repair and so that upon repair, the C terminus of RNase H2A would be fused to a mini AID degron tag and eGFP. The repair construct also contains an antibiotic resistance marker ($G418^R$) under the control of a PGK promoter (Figure 4-6). Two days after transfection, cells were selected in 1 mg/ml G418 for 10 days to isolate potentially successfully targeted clones. Clones that survived selection were further screened for the presence of the tagging cassette at the correct locus by PCR using a forward primer that anneals within the G418 resistance gene and a reverse primer that anneals outside the right homology arm (data not shown). This work was started in the Takeda laboratory during my collaboration, after which we were provided with a candidate clone to carry out further screening.

The candidate clone was further screened by immunoblotting using two different anti-RNase H2A antibodies to check for the presence of the mAID-eGFP fusion tag either before or after growth in 1mg/ml IAA (= 571 µM). Using Bethyl antibody A304-149A (Figure 4-7, top panel) bands corresponding to endogenous, untagged RNase H2A (~32 kDa) and bands at a molecular weight corresponding to RNase H2A tagged with mAID-eGFP (~55 kDa) can be seen both before and after growth in IAA, suggesting that this clone was only tagged heterozygously. However, when probed with Bethyl antibody A304-150A (Figure 4-7, bottom panel) bands can only be seen corresponding to endogenous untagged RNase H2A and not for a tagged version of the protein.

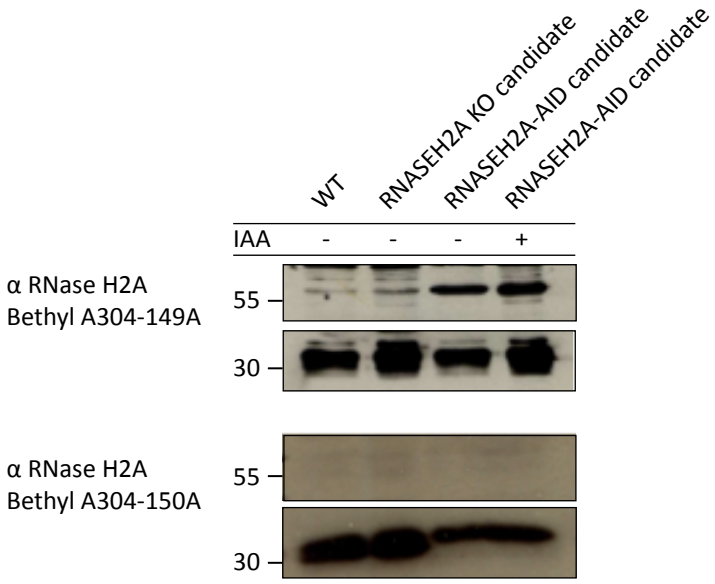


Figure 4-7 C-terminally AID-tagged RNase H2A candidate TK6 cells are either not successfully targeted or are only targeted heterozygously and the AID degron is non-functional.

Immunoblot of wild type, RNase H2A knockout candidate and C-terminally AID-tagged RNase H2A candidate (-/+ IAA) TK6 cells probed with two different anti-RNase H2A antibodies. Both antibodies show that the AID-tagged candidates still express endogenous, untagged RNase H2A (~32 kDa). Antibody Bethyl A304-149A (top panel) shows that whilst at least one copy of *RNASEH2A* may be tagged (~55 kDa), there is no decrease in signal seen after 24 hours growth in IAA. The absence of a band at 55 kDa when probing with antibody Bethly A304-150A (bottom panel) suggests that *RNASEH2A* may not have been successfully targeted.

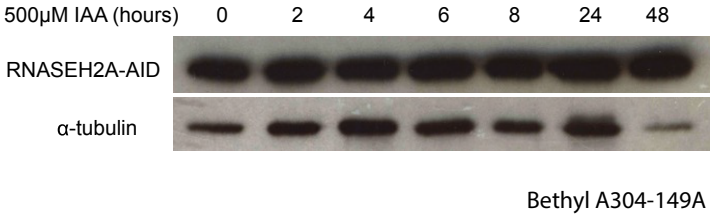


Figure 4-8 The RNase H2A C-terminal AID degron tag in TK6 cells is non-functional.
Immunoblot of C-terminal mAID-tagged candidate TK6 cells probed with an anti-RNase H2A antibody. Whole cell extract was taken either before addition of, or after the stated amounts of time of growth in 500 μM IAA. The lack of decrease in signal over time suggests that the mAID tag does not cause significant degradtion of RNase H2A

Concurrent with our screening, the Takeda laboratory carried out a further screen for the presence of the mAID tag and to check the efficiency of protein degradation following growth in IAA. C-terminal RNase H2A mAID-tagged TK6 cell extracts were immunoblotted with an anti-RNase H2A antibody (Bethyl A304-149A) after growth in 500 μM IAA at a range of time points (Figure 4-8). The signal intensity of the band that is presumed to correspond to mAID-tagged RNase H2A shows no decrease in time even after growth in IAA for 48 hours. Previous work in the Takeda laboratory (e.g. Hoa et al. 2015) has found that after 48 hours, other proteins tagged using the same construct used to tag RNase H2A with mAID are significantly degraded which would suggest that the mAID tag on RNase H2A was not functional.

The results of these Western blots suggest that either RNase H2A is untagged and the bands seen at ~55 kDa using Bethyl antibody A304-149A (Figure 4-7, top panel) are non-specific or that RNase H2A is tagged but not degraded by growth in auxin. The presence of faint bands at the same molecular weight as tagged RNase H2A in wild type cell extract using Bethyl antibody A304-149A (Figure 4-7, top panel) suggest that this band could be non-specific, however, the cells provided to us by the Takeda laboratory did express eGFP suggesting that the mAID tagging construct had been successfully incorporated at the correct locus. No decrease in GFP-positive cells could be seen after growth in 1 mg/ml IAA for 24 hours (data not shown) suggesting that the mAID-eGFP was expressed but non-functional in TK6 cells. It is possible that the presence of the mAID-eGFP tag masks the epitope used for raising Bethyl antibody A304-150A (Figure 4-7, bottom panel), which could account for the absence of a band at ~55 kDa.

4.4 Tagging the C-terminus of RNase H2A with a novel double degron tag.

Concurrent with attempts to tag the C-terminus of RNase H2A with the mini AID degron (mAID), we attempted to tag the RNase H2A C-terminus with a novel double degron tag composed of mAID fused with a newly published self-cleaving **S**mall **M**olecule **A**ssisted **Sh**utoff degron tag (SMASh) (Chung et al. 2015) (Figure 4-9). The

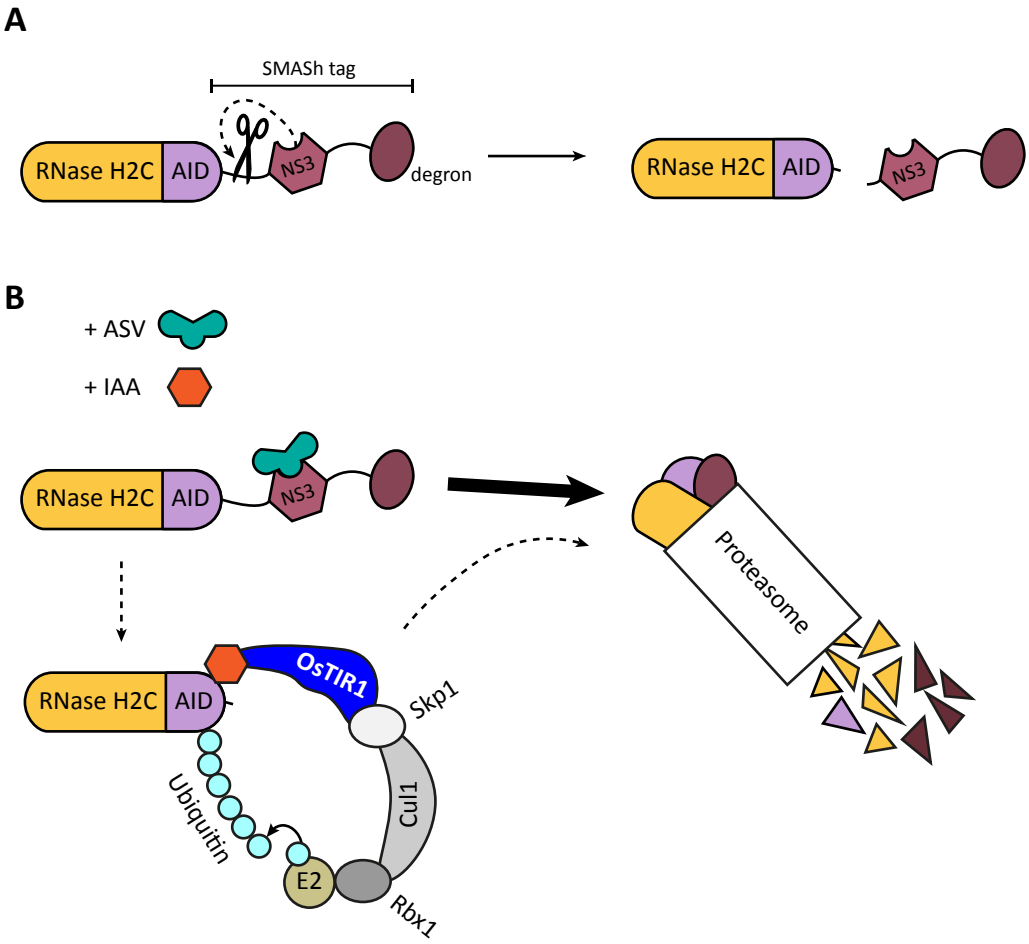


Figure 4-9 Schematic illustration of the AID-SMASh system
A. The SMASh tag consists of a HCV NS3 protease joined to a degron tag which is fused with the target protein via the recognition sequence for NS3. Under normal conditions, the SMASh tag self-cleaves and is degraded by the proteasome allowing stable expression of the target protein. B. Addition of the NS3 inhibitor asunaprevir (ASV) results in the proteasomal degradation of the majority of newly synthesised protein (bold arrow). Addition of IAA allows degradation via the AID degron system (dashed arrows) in the event that SMASh self-cleaves from newly synthesised protein before interaction with ASV, and to degrade any residual protein from which SMASh had self-cleaved before ASV addition.

SMASh tag is a self-cleaving degron allowing tunable and reversible protein degradation. It consists of the hepatitis C virus (HCV) NS3 protease fused with a degron derived from the HCV NS4A protein, joined to the target protein via the NS3 protease cleavage sequence. The NS3 protease can be inhibited by a number of non-toxic, cell permeable drugs including asunaprevir (ASV) (Lamarre et al. 2003; McPhee et al. 2012; Talwani et al. 2013; Jiang et al. 2014). In the absence of ASV, the protease cleaves itself and the degron off the target protein allowing stable expression of the protein. In the presence of ASV, the protease activity of NS3 is inhibited, preventing cleavage of the degron and targeting the protein for degradation via the proteasomal pathway.

Theoretically we could increase the efficiency of degradation by fusing the SMASh tag with the mAID tag. We hypothesized that upon addition of ASV, the majority of newly synthesized proteins would be targeted for proteasomal degradation and any protein that had already lost the SMASh degron would be degraded via the mAID tag (Figure 4-9). The rationale behind this approach offers three distinct advantages over the tagging system described in section 4.3 used to attempt to tag RNase H2A with a mAID-eGFP fusion. Firstly, the AID system depends upon the interaction of a second component SCF^{TIR1} mediated by auxin/IAA for degradation whereas the SMASh tag is a single-component system. As a result, it is presumed that the SMASh tag offers a more robust method of protein degradation than mAID. Secondly, post-translational degradation via the SMASh tag can only target proteins synthesized after the addition of ASV allowing the persistence of forms of the protein from which the SMASh tag has already self-cleaved. The presence of the mAID tag on these residual proteins therefore allows for their degradation. Thirdly, since the C-terminus of RNase H2A is required for stabilizing the formation of the holoenzyme by interacting with the RNase H2B-RNase H2C dimer (Reijns et al. 2011), we assumed that the smaller the tag, the less likely it would be to disrupt trimer formation. It is possible that the presence of the eGFP component of the tag that we attempted to introduce in section 4.3, which was unnecessary for protein degradation, disrupted formation of the trimer sufficiently that homozygous tagging would have been lethal. By removing eGFP from this construct, it was hoped that the mAID tag alone was sufficiently small to allow for trimer formation.

Attempts to tag the C-terminus of RNase H2A with an AID-SMASh double degron were initially carried out in a human TK6 cell line constitutively expressing OsTIR1 from the AAVS1 locus. For attempts to tag RNase H2A with the AID-SMASh double degron we used the same gRNA sequence and repair template homology arms as used for tagging with the mAID degron described in section 4.3. The repair template was designed such that upon repair the AID-SMASh degron would be fused to the terminal exon of RNASEH2A via a short peptide linker (GSGAGA). The repair template was also designed to incorporate a LoxP-flanked antibiotic resistance marker under control of a PGK promoter allowing for selection and subsequent marker removal (Figure 4-10). 10^6 TK6 cells were co-transfected with 200 ng of pX330 and 800 ng repair template (either 800 ng of $G418^R$ or $hygromycin^R$ repair template or 400 ng of each) using Neon electroporation (Invitrogen; 1350 V, 10 ms, ×3 pulse, 10µl Neon tip). 48 hours after transfection, cells were selected in either 1 mg/ml G418 or 625 µg/ml hygromycin to select for heterozygously targeted clones or both 1 mg/ml G418 and 625 µg/ml hygromycin to select for homozygously targeted clones. In all three instances, no clones grew following selection.

To check that the failure to generate AID-SMASh-tagged TK6 cells was not cell line specific and working under the assumption that this failure may have been due to the poor transfection efficiency with TK6 cells, attempts were made to fuse the AID-SMASh tag to RNase H2A in human retinal pigment epithelium cells (RPE1), a wild type cell line immortalized with hTERT which offer the advantage of a very stable karyotype and a modal chromosome number of 46 in over 90% of cells. The RPE1 cells used in this study express OsTIR1 conditionally from the ROSA26 locus which would allow for protein degradation via the mAID degron. Initially, we only sought to isolate heterozygously targeted clones since we still assumed that presence of the mAID tag could prove lethal if present homozygously. 10^6 RPE1 cells were transfected with 0.5 µg of pX330 and 1.5 µg $G418^R$ repair template using Neon electroporation (Invitrogen; 10 µl tip, 1350 V, 20 ms, 2 pulses). 48 hours after transfection, cells were selected in 1 mg/ml G418 for 2 weeks during which some colonies started to grow very slowly but after several rounds of mitosis all cells eventually died. A subsequent second attempt to introduce the AID-SMASh tag to the C-terminus of RNase H2A in RPE1 cells was made using a repair

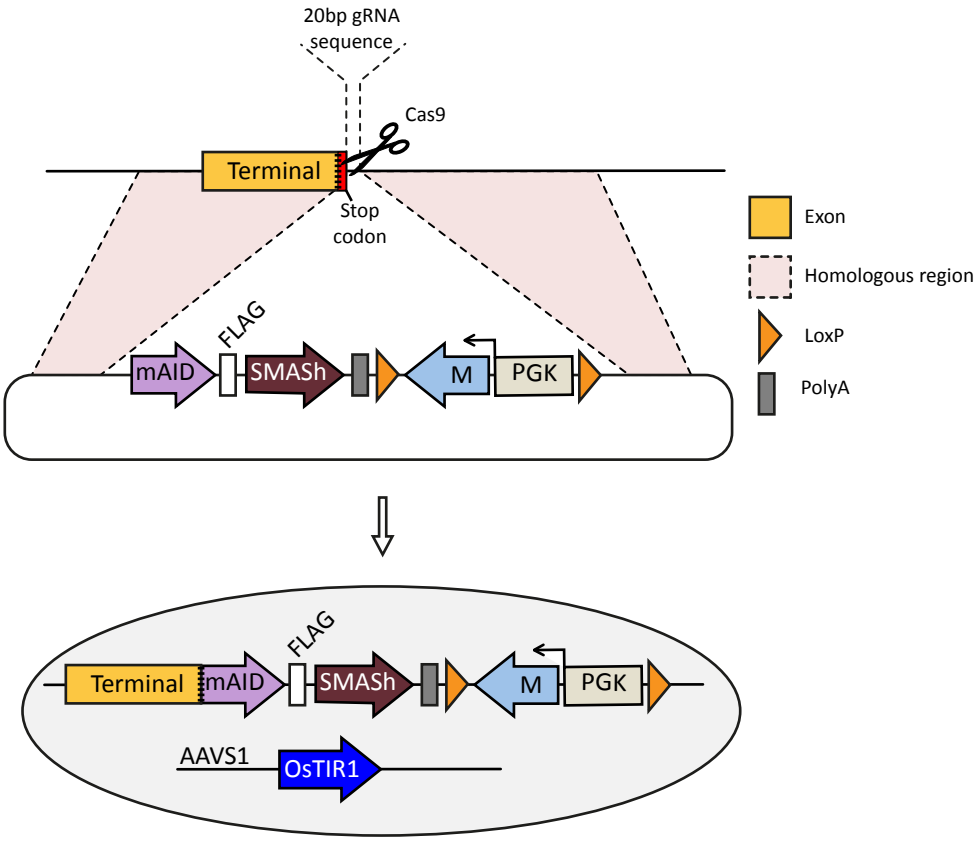


Figure 4-10 *RNASEH2A* C-terminal AID-SMASh tagging strategy
A Cas9-induced double strand break is targeted as close as possible after the stop codon of the terminal coding exon of the gene. The break is repaired by homology directed repair using a repair template designed such that both the endogenous stop codon and the 20bp gRNA sequence are excluded upon repair. Integration of a LoxP-flanked antibiotic resistance gene allows for selection and subsequent marker removal. OsTIR1 is constitutively expressed ectopically from the AAVS1 locus.

template containing a gene conferring resistance to blasticidin in case the previous slow growth phenotype was anything to do with selection in G418. Cells were transfected using the same conditions and reagents as before but with 1.5 μg of blasticidin^R repair template instead of the G418^R template and 48 hours after transfection were selected in 10 μg/ml blasticidin for 2 weeks. As with selection using G418, some clones grew during selection but ultimately died within 2 weeks.

Presuming that at least some of the clones that started to grow following tagging with a C-terminal AID-SMASh tag on RNase H2A were only tagged heterozygously, we assumed that the presence of a C-terminal tag on RNase H2A would be lethal, even if only present at one allele. As such we decided it may be necessary to change our approach and target one of the two RNase H2 accessory subunits, RNase H2B and RNase H2C.

4.5 Regulating RNase H2B and RNase H2C with a C-terminal double degron tag in human RPE1 cells

Having discovered that the presence of a C-terminal GFP tag on both RNase H2B and RNase H2C in *S. pombe* had an undetectable effect on catalytic activity of the holoenzyme, we set out to tag the C-termini of both accessory subunits with a double degron tag in human RPE1 cells. Since deletion of a large part of the C-terminus of RNase H2B (residues 234-312) causes only a very slight decrease in RNase H2 catalytic activity and does not affect trimer formation *in vitro* (Figiel et al. 2011), it was expected that disruption of this terminus with a degron tag may have little effect on complex stability and activity. In contrast, the C-terminus of RNase H2C is essential for trimer formation (Reijns et al. 2011) so we expected that tagging this subunit may be harder due to the destabilizing effect of the degron tag on trimer formation.

We decided to attempt to tag the C-terminus of RNase H2B or RNase H2C with a double degron tag consisting of a mini AID fused with a SMASh tag which functions in the same way as described in section 4.4 and Figure 4-9. This work was carried out in human RPE1 cells that conditionally express Os-TIR1 from the ROSA26 locus allowing for protein

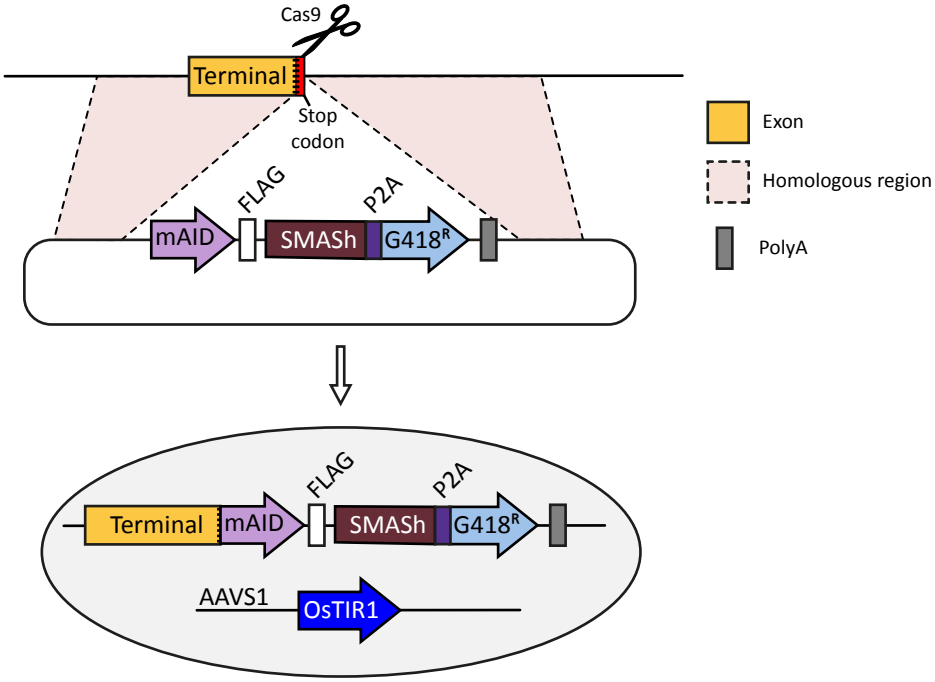


Figure 4-11 ***RNASEH2A*** **C-terminal AID-SMASh tagging alternative strategy**
A Cas9-induced double strand break is targeted as close as possible to the stop codon of the terminal coding exon of the gene. The break is repaired by homology directed repair using a repair template designed such that the original PAM sequence is mutated, preventing cutting of the repair template by Cas9, and the endogenous stop codon is excluded upon repair. A P2A peptide between the SMASh tag and the antibiotic resistance marker ($G418^R$) allows for selection. OsTIR1 is constitutively expressed ectopically from the AAVS1 locus.

degradation via the mAID degron. The knockin construct used to attempt to tag the C-termini of RNase H2B and RNase H2C differed from that used to attempt to tag RNase H2A in TK6 cells in that it encoded a longer linker between the protein and the tag (VGASAASAAGAGGASGSAAGGRVGSGAGA compared to GSGAGA). This longer linker had been previously shown to work for tagging other proteins with the mAID-SMASh double degron and it was hoped that its greater length would cause increased flexibility resulting in better folding of the tagged RNase H2B and RNase H2C and increased interaction with protein partners. In addition antibiotic resistance was conferred by in-frame coupling of a $G418^R$ selection marker using the T2A ribosomal skipping motif (J. H. Kim et al. 2011) rather than via integration of an antibiotic resistance gene with its own promoter. This system offers the advantage that cells should only develop antibiotic resistance following integration of the tagging construct at the correct locus. A further advantage of using this tagging construct was that it had been developed by the Hochegger laboratory and had been used to successfully degrade several proteins (unpublished data) so we knew that the mAID-SMASh tag in this form was functional.

To tag the C-termini of RNASEH2B and RNASEH2C with the mAID-SMASh double degron tag, 10^6 human RPE1 cells were co-transfected with 0.5 μg of pX330 expressing a gRNA designed to target Cas9 to as close as possible to the stop codon of either RNASEH2B or RNASEH2C and 1.5 μg of a repair plasmid containing ~500 bp homology arms (Figure 4-11) using Neon electroporation (Invitrogen; 10 μl tip, 1350 V, 20 ms, 2 pulses). 48 hours after transfection, transfected cells were transferred into media containing 1 mg/ml G418 for 21 days to select single successfully targeted clones. For tagging of RNase H2B, more than 20 clones grew and 18 were expanded for further screening. For tagging of RNase H2C, 7 clones grew and were expanded for further screening.

4.5.1 Screening C-terminal mAID-SMASh tagged RNase H2B candidates for correct integration of the tagging construct

Following selection 18 C-terminal mAID-SMASh-tagged RNase H2B candidate clones were grown up and whole cell extracts were immunoblotted using an antibody against RNase H2B (Invitrogen MA523523) to check for the presence of the tag. This antibody is raised against a sequence within the centre region of RNase H2B so the

presence of a C-terminal tag should not prevent epitope recognition. Successfully tagged RNase H2B should result in a band at ~50 kDa which would represent the combined size of the wild type protein (~40 kDa) plus the ~10 kDa size shift caused by the presence of the mAID tag. No signal should be visible for the fully mAID-SMASh-tagged version of the protein (~75 kDa) as the SMASh portion of the double degron self-cleaves upon translation.

A band corresponding to wild type RNase H2B (Figure 4-12; top panel) can be seen in all 18 of the candidate clones and in wild type whole cell extracts however no band was present at the size corresponding to mAID-tagged RNase H2B. Whilst high exposure causes the appearance of bands that could correspond to mAID-tagged RNase H2B in clones 12 and 13 (Figure 4-12; bottom panel) these bands are significantly weaker than those created by wild type RNase H2B. It is not possible to determine if this is because the bands are a result of non-specific interactions of the antibody, or if the protein is being made, but due to poor folding is being degraded. Thus, even if these clones were heterozygously targeted they would be unsuitable for our purposes as wild type RNase H2B would still be present in cells resulting in a certain level of wild type RNase H2 activity. Assuming that all 18 clones screened were false positives and therefore untagged (or only tagged heterozygously and the protein highly unstable) we decided to focus further work on the C-terminally mAID-SMASh-tagged RNase H2C clones.

4.5.2 Screening C-terminal mAID-SMASh tagged RNase H2C candidates for correct integration of the tagging construct and efficient protein degradation

The 7 C-terminal mAID-SMASh-tagged RNase H2C clones that grew after selection were expanded and their whole cell extracts immunoblotted with an antibody against RNase H2C (Atlas Antibodies HPA065375) (Figure 4-13). Clones 1 and 3 only had a signal corresponding to wild type untagged RNase H2C and were assumed to be false positives. Clone 4 had bands corresponding to both wild type (~20 kDa) and tagged RNase H2C (~30 kDa) and was assumed to be targeted heterozygously. Clones 2, 5, 6 and 7 only had bands corresponding to tagged RNase H2C and had lost the wild type untagged signal and were therefore assumed to be targeted homozygously with the C-terminal double degron tag.

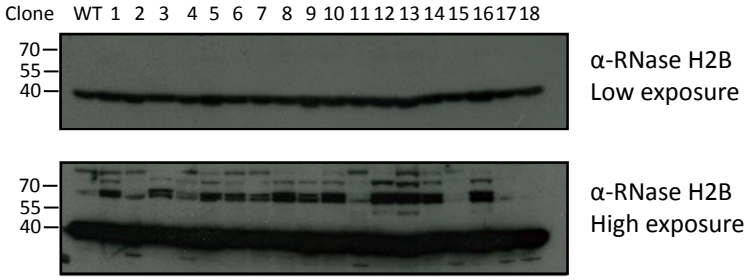


Figure 4-12 *RNASEH2B* is not able to be C-terminally AID-SMASh tagged

RNASEH2B C-terminal AID-SMASh candidate clones were screened by immunoblotting using an antibody against RNase H2B. Signal can be seen for endogenous wild type RNase H2B (~40 kDa), whilst no signal is detected at ~50 kDa which would represent RNase H2B tagged with AID only (the SMASh tag having self-cleaved) therefore all clones were assumed to be negative.

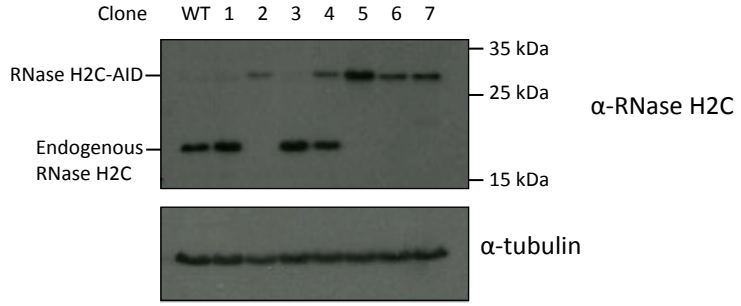


Figure 4-13 Tagging RNase H2C with a C-terminal AID-SMASh double degron tag

RNASEH2C C-terminal AID-SMASh candidate clones were screened by immunoblotting using an antibody against RNase H2C. Clones 2, 5, 6 and 7 only have a signal at a size corresponding to RNase H2C tagged with AID (the SMASh tag having already self-cleaved) and are assumed to be tagged homozygously. Clone 4 has signal for both untagged and tagged RNase H2C and is assumed to be targeted heterozygously. Clones 1 and 3 only have a signal for untagged RNase H2C and are assumed to be untargeted.

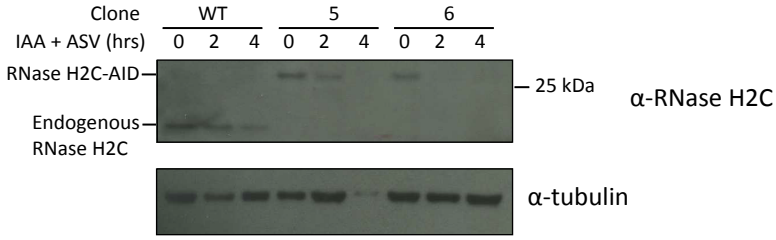


Figure 4-14 A C-terminal AID-SMASh double degron tag causes rapid depletion of RNase H2C in human RPE1 cells

Wild type RPE-1 cells and two C-terminally mAID-SMASh tag RPE1 clones (5 & 6) were treated with 1 mg/ml IAA and 3 µM ASV, following 2 hours growth in the presence of 1 µM doxycycline to drive expression of Os-TIR1 and immunoblotted using an antibody against RNase H2C. Depletion of RNase H2C can be seen after 2 hours with comprehensive protein degradation seen after only 4 hours.

To check for the ability of the double degron system to rapidly deplete tagged RNase H2C, two clones (5 and 6) were selected for further screening. These clones were selected as exemplars expressing close to wild type levels of tagged RNase H2C (clone 5) and expressing reduced levels of the tagged RNase H2C relative to wild type (clone 6). RPE1 cells expressing OsTIR-1 from the ROSA26 locus under control of tetracycline-controlled transcriptional activation that were otherwise wild type and clones 5 and 6 were pre-treated with 1 μM doxycycline for 2 hours to drive expression of OsTIR1 then grown in media containing 1 μM doxycycline, 1 mg/ml IAA and 3 μM ASV; samples were taken before, or 2 and 4 hours after, addition of IAA and ASV. Protein levels were determined by immunoblotting with an antibody against RNase H2C (Figure 4-14). Marked RNase H2C depletion can be seen after 2 hours growth in IAA and ASV in both clones 5 and 6 with undetectable levels of the protein after just 4 hours. The low level of total protein loaded at the 4-hour time point for clone 5 prevents definitive conclusions being drawn about RNase H2C depletion but it is assumed that RNase H2C levels decrease in the same manner as seen for clone 6. Interestingly, levels of endogenous RNase H2C in wild type cells decrease as time in IAA-and ASV-containing growth medium increases however this phenomenon has been reported in the case of other proteins in the case of persistent OsTIR1 expression (Natsume et al. 2016).

4.6 Two-tiered regulation of RNase H2C at the transcriptional and protein level

In parallel to the work carried out in section 4.5 we developed a system whereby expression of RNase H2C could be regulated by a two-tier system targeting transcription and protein stability. The transcriptional regulation system of the Tn10 bacterial operon has been repurposed for use in eukaryotic systems to switch transcription on in the presence (Tet-on) or absence (Tet-off) of Tetracycline (Tc) and its more stable derivative doxycycline (dox) (Gossen & Bujard 1992; Gossen et al. 1995). The Tet-on system was developed by fusing the transcription activation domain of herpes simplex virus VP16 protein to a mutant reverse Tet Repressor protein (rTetR) to form the reverse

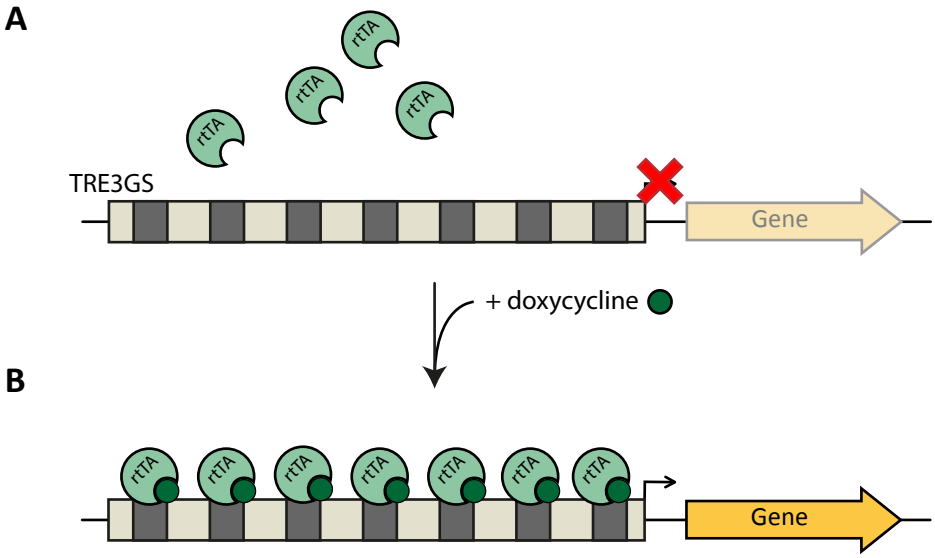


Figure 4-15 Schematic of the Tet-on method of tetracycline-controlled transcriptional activation

A. In the absence of doxycycline, the reverse tetracycline-controlled transactivator (rtTA) is unable to bind to the TRE3GS promoter and transcription is shut off. B. A conformational change in rtTA brought about by interaction with doxycycline allows binding of rtTA to tet operator regions (dark grey) within the TRE3GS promoter driving gene expression.

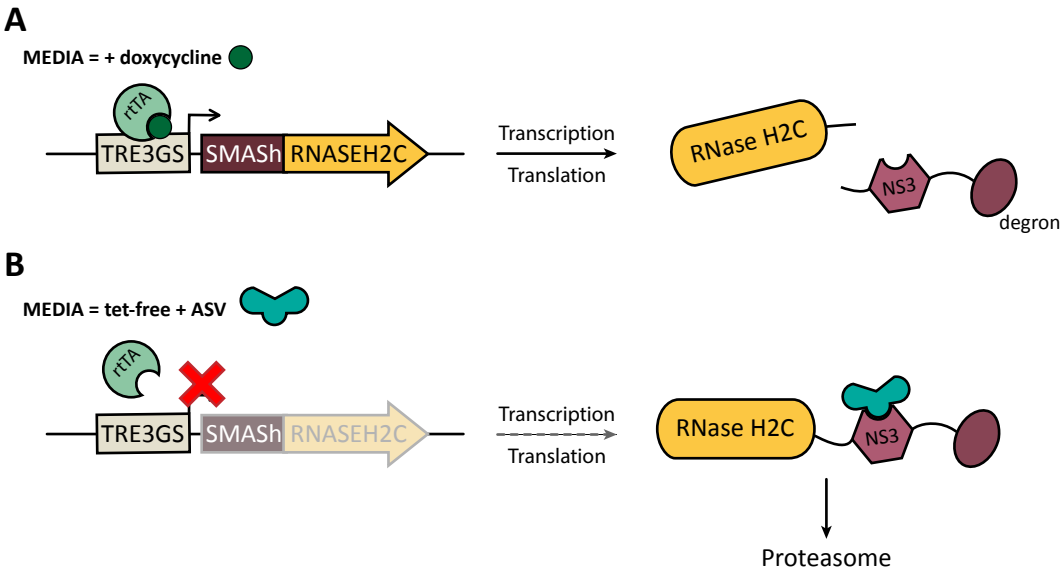


Figure 4-16 Schematic illustration of the two-tiered mechanism of regulating ectopic expression of RNase H2C

A. Transcription of C-terminally SMASh-tagged *RNASEH2C* can be driven by the interaction of the reverse tetracycline-controlled transactivator (rtTA) with the TRE3GS promoter in the presence of doxycycline. In the absence of asunaprevir (ASV), the SMASh tag self-cleaves allowing stable protein expression. B. In the absence of doxycycline or tetracycline, transcription is turned off. In the event of leakiness in the system (dashed arrow), any synthesised RNase H2C is targeted for proteasomal degradation via the SMASh degron in the presence of ASV.

tetracycline-controlled transcriptional activator (rtTA) that binds the tet operator and drives transcription in the presence of Tc/dox (Figure 4-15).

We hoped that we could significantly downregulate transcription of RNase H2C by placing it under the control of the TRE3GS promoter (Heinz et al. 2011) and expressing it ectopically at the ROSA26 locus. Despite the tight control of gene transcription offered by the TRE3GS promoter (Loew et al. 2010; Heinz et al. 2011) some leakiness in the system is expected. In the event that transcriptional leakiness allowed expression of some protein, we added an N-terminal SMASh tag (Chung et al. 2015) to RNase H2C which would allow for protein degradation via the proteasomal pathway (Figure 4-16).

This work was carried out in HT1080 *HP-I-RT* cells that contain an I-SceI restriction enzyme recognition sequence within exon 6 of the *HPRT* gene as described in section 3.3.1. It was hoped that using a cell line containing a site for inducible double strand break (DSB) formation may prove useful further down the line if we wished to investigate DNA replication dynamics in the context of a specific DSB. To integrate the this two-tiered expression system at the ROSA26 locus, 10^6 HT1080 *HP-I-RT* cells were transfected using Amaxa nucleofection (Kit T, programme L-005) with 0.4 μg pX330 encoding a gRNA targeting Cas9 to a region within the ROSA 26 locus and 1.6 μg of repair template. The repair template was designed such that repair of the Cas9-induced DSB would incorporate a cassette containing three different components: 1) a PGK promoter driving constitutive expression of the rtTA (Tet-on protein); 2) an N-terminally FLAG-SMASh-tagged, human codon optimised, Cas9-resistant *RNASEH2C* cDNA and 3) an SV40 promoter driving constitutive expression of a zeocin resistance gene (Figure 4-17).

48 hours after transfection, cells were transferred to growth medium containing 400 μg/ml zeocin for 14 days to isolate single successfully-targeted clones. To minimise the risk of zeocin-mediated DNA damage, cells were selected in a concentration of zeocin at the lower end of the concentration required to kill cells in a kill curve experiment (kill curve data not shown). To show that clones that survived selection in zeocin did so due to integration of the zeocin resistance gene (Zeo^R) and not due to the low concentration of zeocin used for a selection, an initial PCR screen was used to verify presence of Zeo^R

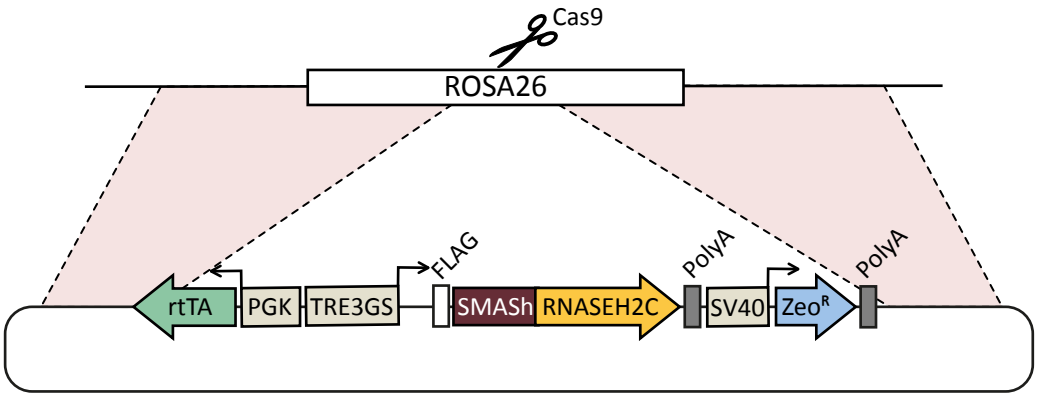


Figure 4-17 Two-tiered regulation of RNase H2C knockin strategy
A cas9-induced double strand break is targeted within the ROSA26 locus. Repair of the break by homology directed repair introduces a cassette that constitutively expresses the reverse tetracycline-controlled transactivator (rtTA) and confers zeocin resistance (Zeo^R) to successfullly targeted cells. Additionally the cassette contains human codon optimised, cas9-resistant RNASEH2C cDNA tagged N-terminally with both flag and SMASh tags, under control of the TRE3GS promoter.

using forward and reverse primers that both annealed within Zeo^R (Figure 4-18 A). All 13 of the clones screened contained Zeo^R and were therefore assumed to have integrated the repair template. A further PCR to verify integration of the repair template at the correct locus was carried out using a forward primer that anneals within Zeo^R and a reverse primer that anneals outside the region used for the right homology arm. Despite several attempts to optimise the PCR using different combinations of primer sequences and a variety of thermocycling parameters, no conclusive data was obtained from this experiment (data not shown). Since all of the components required for this two-tiered system were contained within the knockin construct and were not dependent on regulatory sequences within or around the ROSA26 locus, it was decided that even random integration within the genome would be suitable for our purposes so no further work was carried out to ascertain the location of cassette integration.

To check for expression of N-terminally SMASh-tagged RNase H2C, 10 candidate clones were grown for 24 hours in medium containing 1 μM doxycycline and 3 μM ASV with the final 4 hours either with or without 10 μM MG132 proteasome inhibitor and immunoblotted using an anti-FLAG antibody which recognises the FLAG epitope placed upstream of the SMASh tag in this construct (Figure 4-18 B). In this system, doxycycline drives overexpression of the SMASh-tagged RNase H2C whilst ASV prevents SMASh tag self-cleavage. To allow stable expression of the SMASh-tagged RNase H2C and detection of the tagged protein via the FLAG tag, protein degradation was inhibited by use of MG132. SMASh-tagged RNase H2C (~55 kDa) can be seen in clones 1, 2, 3, 6, 7, 9 and 12 with an increase in signal for each seen 4 hours after addition of MG132 showing that degradation occurs via the proteasomal pathway. It is assumed that the differences in expression level of the tagged RNase H2C can be accounted for by either heterozygous (low expression) or homozygous (high expression) integration of the knockin construct or by homozygous (low expression) or multiple random integration of the knockin construct (high expression).

A further immunoblot using an antibody against RNase H2C (Atlas Antibodies HPA065375) was carried out on the four clones that express the highest levels of SMASh-tagged RNase H2C as seen in Figure 4-18 B (clones 1, 2, 6 and 9; Figure 4-18 C) using the

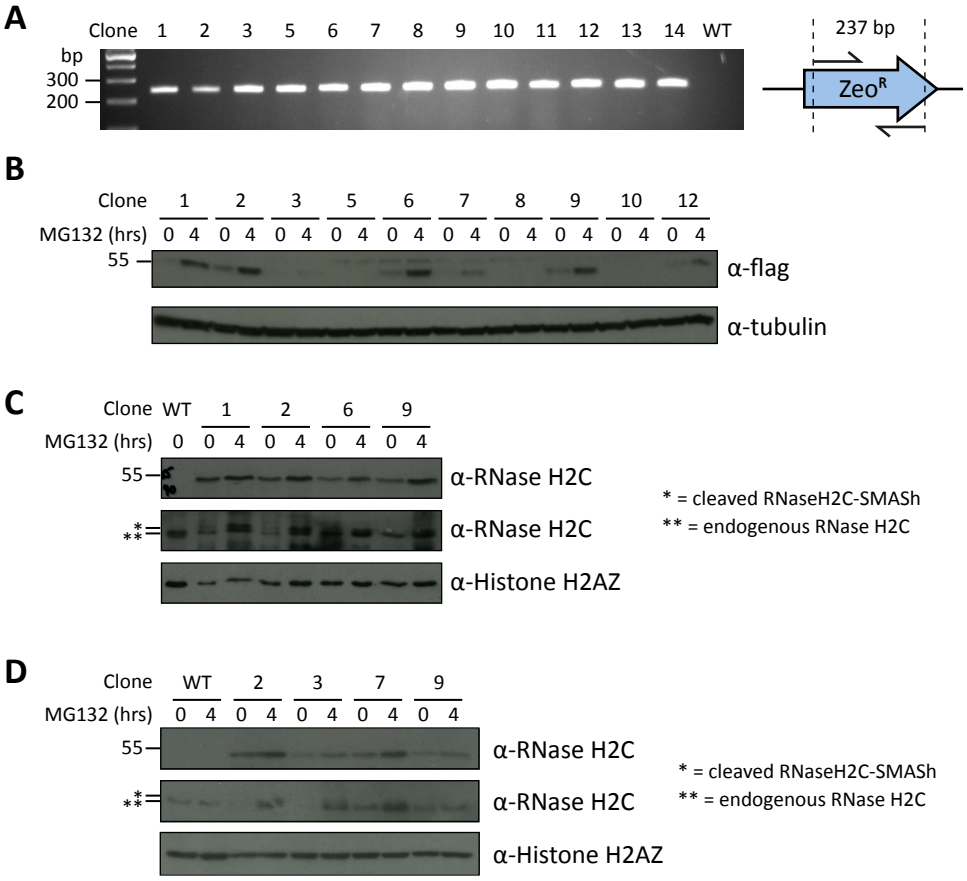


Figure 4-18 N-terminally SMASh-tagged RNase H2C can be ectopically expressed from the ROSA26 locus in human HT1080 cells and is degraded via the proteasomal pathway.
A. Clones were screened by PCR using primers that anneal within Zeo^R to verify the presence of the zeocin resistance gene (Zeo^R) in the genome of targeted cells. B. To check for the presence of the SMASh tag and to verify that degradation occurs via the proteasome, clones were grown in media supplemented with 1 μM doxycycline to drive expression of ectopically expressed SMASh-tagged RNase H2C and screened by immunoblot using an anti-flag antibody either before, or 4 hours after, addition of the proteasome inhibitor MG132 (25 μM). Clones 1-3, 6, 7, 9 & 12 were successfully targeted and the increased signal in the presence of MG132 indicates proteasomal degradation. C & D. Clones expressing higher levels of tagged RNase H2C as determined from figure B (C; clones 1, 2, 6, 9. D; clones 2, 9) and those expressing lower levels (D; clones 3, 7) were treated as in B and analysed by immunoblotting using an antibody against RNase H2C. The self-cleaving of the SMASh-tag leaves a short peptide on the N-terminus of RNase H2C resulting in a slight upshift in molecular weight in relation to endogneous RNase H2C, seen in the middle panels of C and D.

same drug-treatment conditions as described in the previous section. As before, all 4 candidate clones have bands at ~55 kDa corresponding to full length, uncleaved SMASh-tagged RNase H2C in addition to a band corresponding to endogenous RNase H2C (~20 kDa). Cleavage of the SMASh tag leaves a short peptide at the N-terminus of ectopically expressed RNase H2C consisting of the linker and a part of the protease recognition sequence. This peptide results in a slight upshift in molecular weight which can be seen for all 4 candidate clones, but in particular for clone 1 at the 4-hour time point. Together these results indicate that doxycycline drives transcription of ectopically expressed SMASh-tagged RNase H2C and that the SMASh-tag is functional in terms of self-cleavage.

In the event that it would be advantageous to have lower overall cellular levels of the ectopically expressed SMASh-tagged RNase H2C, a further immunoblot was carried out using the same methods as described for Figure 4-18 C to check that clones that expressed low levels of SMASh-tagged RNase H2C when probed with an anti-FLAG antibody (clones 3 & 7) showed similar low expression when probed with an anti-RNase H2C antibody (Figure 4-18 D). Clone 3 exhibited lower expression of SMASh-tagged RNase H2C than a previously identified high expressor (clone 2); by contrast clone 3 showed higher expression than another previously identified high expressor (clone 9). Furthermore, clone 7 showed expression levels of SMASh-tagged RNase H2C similar to clone 2. Despite the inconsistencies, the data from this immunoblot do however suggest that clone 3 would be a suitable candidate for a clone expressing low levels of SMASh-tagged RNase H2C.

To assess the extent to which the tetracycline-controlled transcriptional activation system could downregulate expression of SMASh-tagged RNase H2C, candidate clones were cultured for 24 hours either in growth medium supplemented with Tet-free serum, regular serum or regular serum with 1 μM doxycycline and immunoblotted using an anti-RNase H2C antibody (Atlas Antibodies HPA065375) (Figure 4-19 A). For all four clones screened (1, 2, 6 & 9), growth in doxycycline causes overexpression of SMASh-tagged RNase H2C as seen by the presence of a strong band at a molecular weight slightly above that of endogenous RNase H2C. The absence or near-absence of a band corresponding

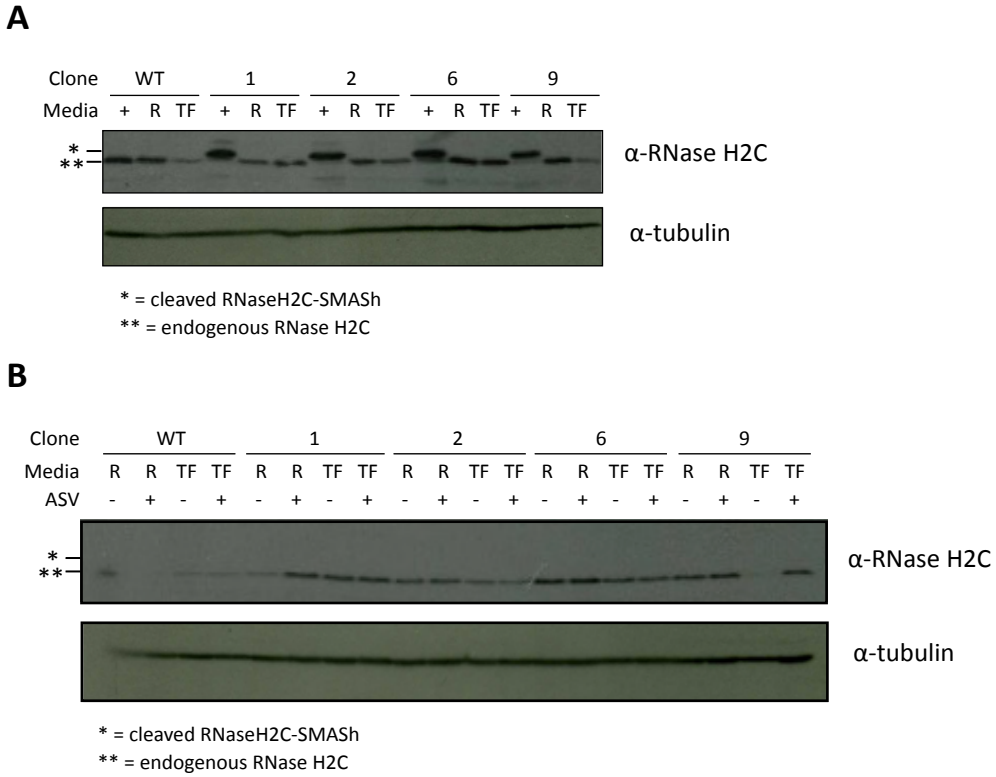


Figure 4-19 Ectopically expressed RNase H2C can be downregulated in a two-tiered manner by tetracycline-controlled transcriptional acitvation and SMASh tag-mediated degradation.
A. To assess the ability of the Tet-on system to shut off transcription of *RNASEH2C*, cells were grown for 24 hours in DMEM supplemented with 1 μM doxycycline (+), DMEM with regular FCS (R) or DMEM with tetracycline-free FCS (TF) and immunoblotted using an antibody against RNase H2C. Growth in doxycycline resulted in overexpression of SMASh-tagged RNase H2C whilst cells grown in regular or tet-free media have expression of RNase H2C shut off at the transcriptional level. B. To assess the ability of the two-tiered system to knock down expression of RNase H2C, cells were grown in regular media (R) or tet-free media either with or without Asunaprevir (ASV) and immunoblotted with an anti-RNase H2C antibody. Downregulating RNase H2A at the transcriptional and protein levels results in a strong knockdown of the protein.

to endogenous RNase H2C for all four clones when grown in doxycycline suggests that expression of endogenous RNase H2C is transcriptionally downregulated. For all four clones growth in medium containing regular serum, which may contain trace amounts of tetracycline resulting in some transcriptional leakiness, and medium containing tet-free serum expression of SMASh-tagged RNase H2C was undetectable suggesting very tight transcriptional control.

To check whether growth in ASV in combination with transcriptional downregulation would have an additive effect on protein expression in comparison to transcriptional downregulation only, cells were grown in the presence or absence of 3 μM ASV in medium supplemented with either regular serum or tet-free serum and immunoblotted using an antibody against RNase H2C (Atlas Antibodies HPA065375) (Figure 4-19 B). Due to the tight downregulation of SMASh-tagged RNase H2C in the absence of doxycycline seen in Figure 4-19 B, it is unsurprising that for all 4 clones screened, no signal is seen at a size corresponding to ectopically expressed RNase H2C from which the SMASh tag has self-cleaved. It is however assumed that any SMASh-tagged RNase H2C synthesized after addition of ASV is directly degraded and would have no additive effect on the signal seen just above the band corresponding to endogenous RNase H2C.

4.7 DISCUSSION

In this chapter I have shown that the C-terminus of RNase H2A is not a valid target for homozygous tagging with a range of degron tags, despite the homologous protein in *S. pombe* tolerating a C-terminal GFP tag. Through failure to C-terminally tag endogenous RNase H2B with an mAID-SMASh-tag in RPE1 cells as described in section 4.5.1, I have shown that it is likely not possible to introduce large tags to the C-terminus of RNase H2B. Furthermore, unsuccessful attempts to generate HT1080 *HP-I-RT* cells expressing N-terminally SMASh-tagged RNase H2B from the ROSA26 locus using the same methods described in 4.5.2 (data not shown) suggest that the N-terminus of RNase H2B may also be unsuitable for tagging. In the latter case, some colonies grew following selection but died after a few cell cycles, despite the persistence of endogenous RNase H2B. It was thought in this case that expression of the tagged version of RNase H2B was

sufficiently high in comparison to expression of endogenous RNase H2B that the majority of RNase H2B-RNase H2C heterodimers were formed using the tagged version of RNase H2B and that the presence of the tag prevented association of the RNaseH2B-RNase H2C dimer with RNase H2A. Furthermore, it was assumed that the damage resulting from persistent genomic ribonucleotides caused by the failure to form stable RNase H2 trimers was tolerated for a certain time before proving lethal.

I have developed two different systems to downregulate expression of RNase H2C in two different human cell lines. Despite the efficacy of the N-terminally SMASh-tagged RNase H2C under the control of tetracycline-controlled transcriptional activation generated in HT1080 cells, further work in these cells was halted due to time constraints and the requirement to knock out endogenous RNase H2C before any studies on holoenzyme activity could be carried out. Further screening to assess the impact of RNase H2C depletion via the double degron tag described in section 4.5 is carried out and described in chapter 5.

Chapter 5

Developing an assay to quantify ribonucleotide incorporation in human genomic DNA

5.1 Introduction

Key to the methodology of Polymerase Usage Sequencing (PU-seq) is the persistence of ribonucleotides misincorporated into genomic DNA by DNA polymerases containing active site mutations that decrease their fidelity. For PU-seq in *S. pombe*, the catalytic subunit (RNase H2A) of RNase H2, the major enzyme responsible for the removal of misincorporated ribonucleotides, is knocked out. As described in chapter 4 RNase H2A knockout is not possible in most human cells so we developed a system in which an accessory subunit of RNase H2, RNase H2C, can be conditionally knocked down via a C-terminal mAID-SMASh double degron tag. To assess the impact on RNase H2 catalysis of mAID-SMASh-mediated RNase H2C depletion, we needed to develop an assay that would allow us to examine the decrease in ribonucleotide removal from genomic DNA. Moreover, it was hoped that this assay would be useful to us when screening candidate ribo-incorporating mutations in the two major replicative polymerases.

Work in this chapter therefore focuses on the development of such an assay as well as recapitulating two previously published methods to assess RNase H2 activity *in vitro* and to infer the extent of ribonucleotide incorporation in genomic DNA.

5.2 Quantifying ribonucleotide incorporation by alkaline hydrolysis and alkaline agarose gel electrophoresis

The current method used in the Carr laboratory to assess the extent of ribonucleotide (rNTP) incorporation in strains of *S. pombe* that are used in Polymerase Usage sequencing experiments involves treating 5 μg genomic DNA extracted from each strain with 0.3 M KOH or NaOH at 55°C for 2 hours. The alkali-treated gDNA was then resolved on a 1% alkaline agarose gel. The presence of the 2'-OH on the ribose moiety of a misincorporated rNTP embedded in gDNA causes the sugar phosphate backbone at this point to be 5 times more susceptible to alkaline hydrolysis (Clausen et al. 2015); as such, the greater the amount of rNTPs incorporated within gDNA, the greater the extent of smearing is seen on agarose gels due to the increased hydrolysis of the DNA backbone. This increase in smearing of alkali-treated gDNA can be seen when comparing strains of *S. pombe* containing increasingly more effective methods of RNase H2A knockdown that result in decreasing levels of cellular RNase H2 activity (Figure 5-1 A).

We suspect that the increased complexity of human genomic DNA in comparison to that of *S. pombe* may make this method unsuitable for assessing genomic ribonucleotide incorporation. In order to test this theory, we were able to use genomic DNA from two well characterized Aicardi-Goutières Syndrome (AGS) patient cell lines containing hypomorphic mutations in either RNase $H2A^{G37S}$ or RNase $H2B^{A177T}$ (a kind gift from Professor Andrew Jackson). 1.6 μg gDNA from each of the two patient cell lines and from a wild type lymphoblast cell line (a kind gift from Professor Mark O'Driscoll) were treated with 0.3M KOH at 55°C for 2 hours and subjected to 1% agarose gel electrophoresis alongside untreated gDNA from each of the three cell lines (Figure 5-1 B). Since the reduction in RNase H2 activity in cells containing the G37S mutation is more severe than in cells containing the A177T mutation, and both have lower activity than wild type RNase H2 (Pizzi et al. 2015) we would expect to see a greater level of smearing of gDNA following alkaline hydrolysis and gel electrophoresis from wild type

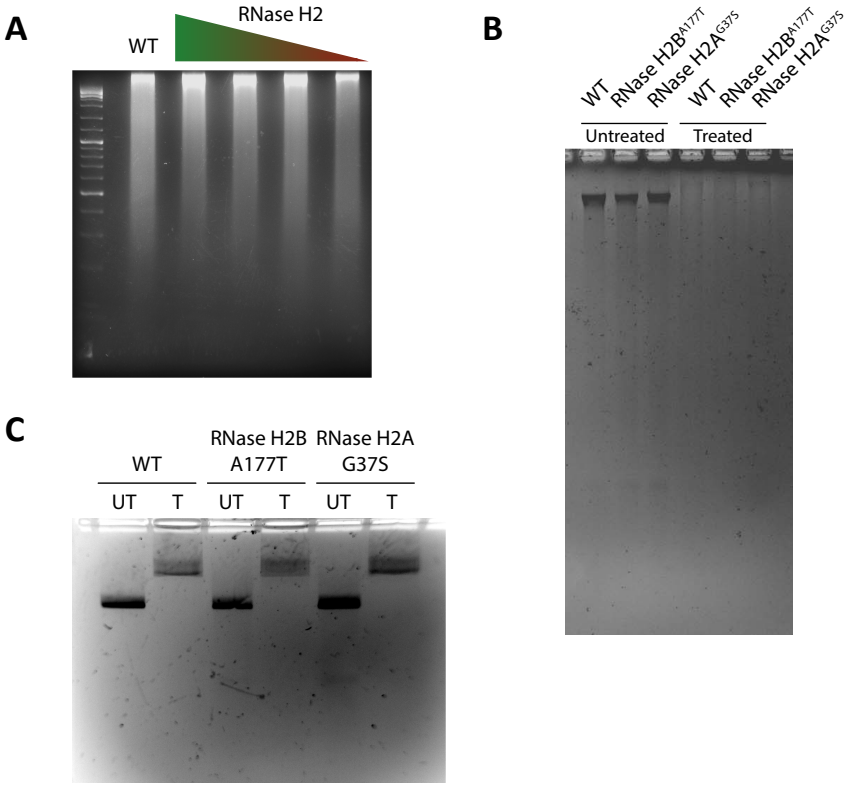


Figure 5-1 Alkaline hydrolysis and gel elctrophoresis is an unsuitable method to assess human genomic ribonucleotide incorporation

A. Alkali-treated gDNA from strains of *S. pombe* containing increasingly efficient mechanisms to downregulate expression of RNase H2A. Lower cellular RNase H2 levels result in greater smearing of gDNA due to increased alkaline hydrolysis at sites of genomic ribonucleotides. Image from Catherine Doust MSc dissertation. B. Genomic DNA from human cell lines with the indicated genotypes was left untreated or treated with 0.3M KOH at 55˚C for 2 hours and subjected to 1% agarose gel electrophoresis. The resultant DNA signal following alkaline hydrolysis was too diffuse to determine differences between treated samples. C. Genomic DNA was treated as in B and subjected to 2% agarose gel electrophoresis. No discernable difference in signal intensity could be seen between treated samples from cell lines with different levels of RNase H2 activity and wild type cells.

to A177T and from A177T to G37S, however the extent of hydrolysis in all three samples was sufficiently high that no information could be determined from this assay.

To increase the signal intensity for alkali-treated gDNA following gel electrophoresis, 1.2 μg gDNA from the three cell lines was treated with 0.3 M KOH at 55°C for 2 hours as before, but resolved on a 2% agarose gel and run alongside 0.4 μg untreated gDNA from each of the cell lines (Figure 5-1 C) in a similar assay as described for *S. pombe* in section 4.2.2. Whilst the bands for each of the alkali-treated samples appear more diffuse than their corresponding untreated samples, there is no discernable difference between the three alkali-treated samples, suggesting that this method of assessing genomic ribonucleotide incorporation is unsuitable for human gDNA.

5.3 Developing a genomic ribonucleotide incorporation assay using an *Arabidopsis thaliana* tRNA ligase

5.3.1 Arabidopsis thaliana tRNA Ligase

As described in section 5.2 the current method used by the Carr laboratory to assess the extent of unrepaired ribonucleotides in genomic DNA in *S. pombe* involves extracting gDNA from the required strains, treating it with alkali and subjecting it to agarose gel electrophoresis. As shown in section 5.2, this method is unsuitable for use with human gDNA and furthermore this method does not discriminate between breaks caused by alkaline hydrolysis at sites of genomic ribonucleotides and stochastic DNA breaks (Figure 5-2 A).

In developing an assay sensitive enough to detect differences in ribonucleotide incorporation between different human cell lines and with the ability to discriminate between fragments of DNA with ribonucleotides at the 3' termini and those fragments arising as a result of stochastic breaks between neighbouring deoxyribonucleotides following alkaline hydrolysis, we decided to exploit the distinctive ligation mechanism of an *Arabidopsis thaliana* tRNA ligase (AtRNL). Alkaline hydrolysis at the site of a

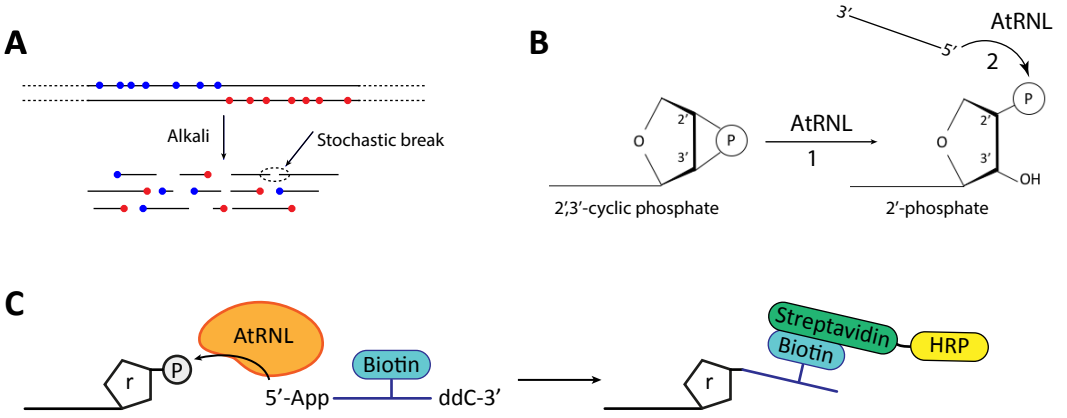


Figure 5-2 ***Arabidopsis thaliana*** **tRNA ligase mechanism and assay principle**
A. Following alkaline hydrolysis, genomic DNA from polymerase usage sequencing cells breaks at sites of ribonucleotides (rNTPs; red and blue dots) leaving single stranded DNA fragments terminating with an rNTP moiety at the 3' end. A further source of DNA breaks arises through stochastic breaks between adjacent deoxyribonucleotides. B. Upon alkaline hydrolysis, rNTP termini form either 2',3'-cyclic phosphate or 2'-phosphate moieties. *Arabidopsis thaliana* tRNA ligase (AtRNL) performs two key functions to specifically select for DNA fragments terminating with these rNTP moieties: it first converts any cyclic phosphate groups into 2'-phosphates and it then ligates the 5' termini of other nucleotides to these ends . C. Ligating an internally biotinylated oligonucleotide pre-adenylated at the 5' end and blocked at the 3' end with dideoxycytidine (ddC) allows for probing with a streptavidin-HRP conjugate.

genomic ribonucleotide leaves either 2'-phosphate or 2',3'-cyclic phosphate termini. AtRNL possesses two catalytic activities that allow it to select only those fragments terminating in a ribonucleotide: firstly, it converts 2',3'-cyclic phosphate groups to 2'-phosphates and it then ligates these 2'-phosphate termini to the 5' end of DNA or RNA (Schutz et al. 2010; Koh et al. 2015) (Figure 5-2 B).

5.3.2 Quantifying ribonucleotide incorporation using biotin

Our initial intention was to use a commercially available Quant*Tag biotin quantitation kit (Vector Laboratories) to quantify how many moles of biotinylated oligomer ligated to ribo-terminating ends of alkali-treated gDNA were present in a sample. Strains of *S. pombe* with *rnh201*, the gene encoding the catalytic subunit of RNase H2, knocked out are estimated to incorporate at least one ribonucleotide per 3000 nucleotides polymerized. 1 μg of gDNA from *rnh201*Δ would therefore contain ~3×10^{11} ribonucleotides meaning that the Quant*Tag kit and subsequent assays would have to be sensitive to quantities of biotin in the range of 1-10 pmoles. To test the sensitivity of the Quant*Tag kit and in order to plot a standard curve from which future quantities of biotin could be inferred a range of quantities from 0.01-10 nmol of biotinylated oligomer were prepared as per the manufacturer's instructions and subjected to spectrophotometry. Whilst quantities of biotin in excess of 0.05 nmol provided absorbance readings >0.001, quantities below this amount were below the range of sensitivity of the spectrophotometer (Table 5-1) and as such it was decided to proceed with experiments using streptavidin-alkaline phosphatase (AP) to quantify biotin.

5.3.3 The principle of the ribonucleotide incorporation assay using AtRNL

Following alkaline hydrolysis of genomic DNA from PU-seq cell lines, we hoped that we would be able to use AtRNL to ligate internally biotinylated DNA oligomers to ribo-terminating gDNA fragments and visualize the resultant ligation products using a streptavidin-HRP or streptavidin-AP conjugate (Figure 5-2 C). This process would require six steps: 1) alkaline hydrolysis of genomic DNA; 2) ligation of biotinylated oligonucleotide to 2'-phosphate groups; 3) removal of unligated oligomers; 4) transfer

Biotin amount (nmol)	Absorbance @535 nm
10	0.207
5	0.099
2.5	0.049
1	0.019
0.5	0.009
0.05	0.001
0.025	NR
0.01	NR

Table 5-1 Quant*Tag spectrophotometer absorbance values
The stated amounts of biotinylated 18mer oligonucleotide were prepared using the Quant*Tag biotin quantitation kit and analysed by spectrophotometry. Amounts of biotin below 0.05 nmol were undetectable by this assay. NR = no reading.

of ligated DNA fragments to a membrane; 5) hybridization of streptavidin-HRP/AP conjugate to biotinylated DNA fragments; 6) chemiluminescent visualization and quantification of HRP/AP to determine molar quantity of biotin in the sample.

5.3.4 Optimizing biotin quantification using streptavidin-alkaline phosphatase

To assess the sensitivity of detecting and differentiating between quantities of biotin in the predicted range (~1-10 pmoles) and to ascertain the lowest possible levels of detection in the event of low ligation efficiency, the biotinylated 18mer was diluted in Nucleospin Gel and PCR-Clean-up elution buffer NE (Macherey Nagel) to obtain stocks with varying concentrations ranging from 5×10^{-5} pmoles/μl to 50 pmoles/μl. 1 μl of each stock was blotted onto Genescreen nylon hybridization transfer membrane (Perkin Elmer) and allowed to dry then UV crosslinked at $1200\times100\mu J/cm^2$. The membrane was blocked using Tropix I-BLOCK (Invitrogen), probed with streptavidin-AP and visualized using CDP-Star chemiluminescent substrate (GE Healthcare) in an ImageQuant LAS 4000 imager (GE Healthcare) (Figure 5-3). This method of biotin quantification allows detection of signal within the predicted range of required sensitivity 1-10 pmoles) as well as detecting amounts of biotin as low as 0.05 fmoles.

Having established the sensitivity of biotin quantification using streptavidin-AP, we used slot blotting to apply 50 μl of biotinylated 18mer to the membrane to reflect the requirement for analyzing larger volumes of samples following removal of the unligated 18mer oligonucleotide following the AtRNL ligation step (Figure 5-4 A). This method of blotting allowed detection of 0.05 pmoles of biotinylated 18mer after 1 minute of exposure by ImageQuant LAS 4000 but required longer exposure times (>5 minutes) to see amounts of biotinylated 18mer below 0.005 pmoles.

The intensity of the biotin-streptavidin-AP signal can be assigned a numerical value based on pixel saturation using ImageJ (Schneider et al. 2012), however this function suffers from the problem of pixel saturation of high intensity signals when trying to compare the intensity of a high concentration signal with that of a signal at the lower end of the range of sensitivity. To overcome this problem, a plugin for ImageJ was designed by Dr. Alex Herbert that stacks images taken at 1-minute time points and plots

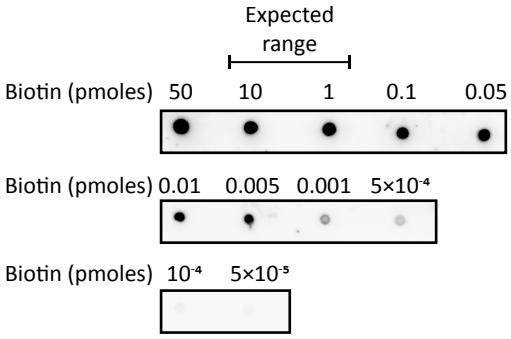


Figure 5-3 Visualising biotinylated DNA with streptavidin-alkaline phosphatase is highly sensitive

1 µl of the indicated amounts of an internally biotinylated 18 nucleotide oligomer were dot-blotted onto nylon transfer membrane, incubated with streptavidin-alkaline phosphatase and visualised using an ImageQuant LAS 4000. This method of detection was sensitive to quantities of biotinylated oligomer in the attomole range. Amounts of biotinylated oligomer in the expected range for the AtRNL assay were comfortably within the limits of detection of this method.

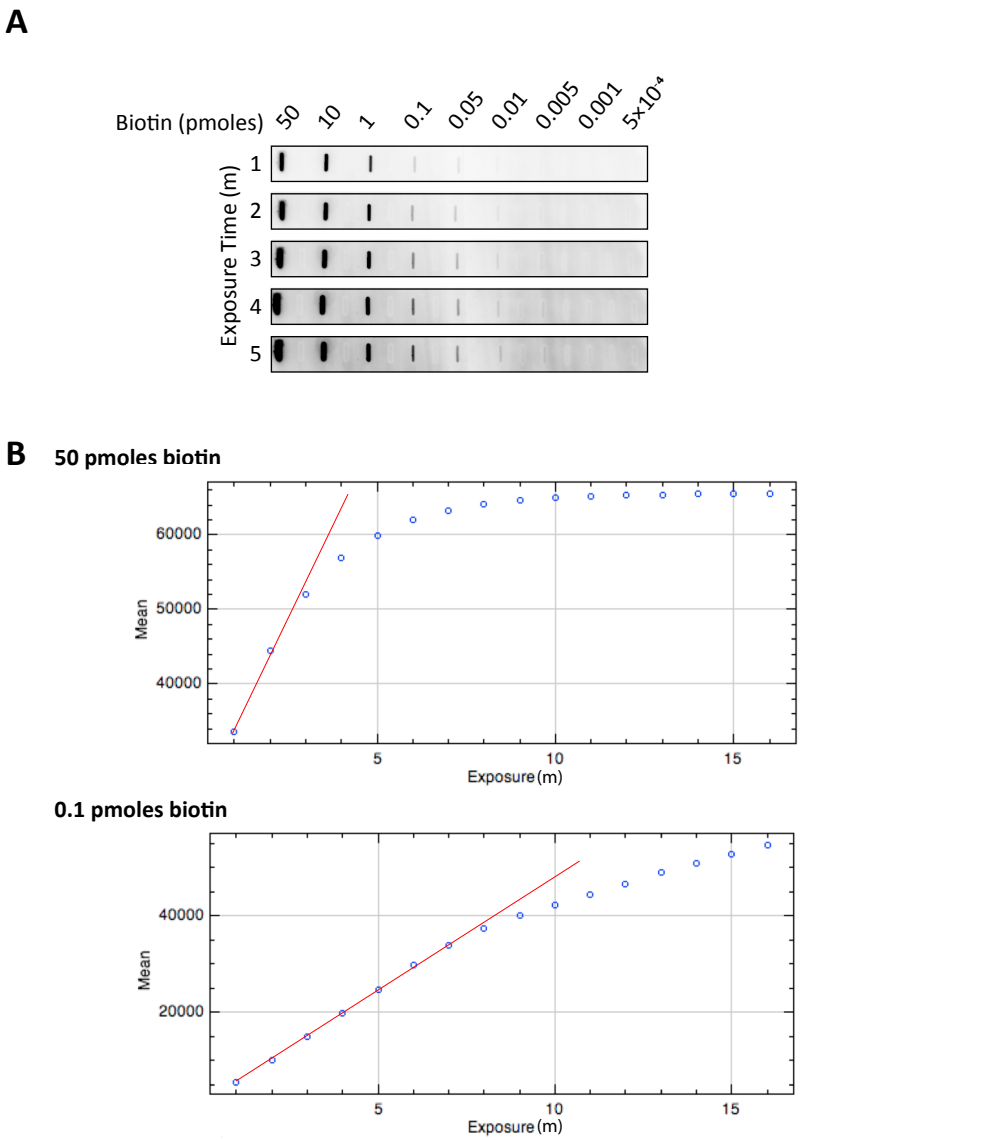


Figure 5-4 Quantifying slot-blotted biotinylated DNA with an image layering-based ImageJ plugin

A. 50 µl of the indicated amounts of an internally biotinylated 18 nucleotide oligomer were slot-blotted onto nylon transfer membrane, incubated with streptavidin-alkaline phosphatase and visualised using an ImageQuant LAS 4000 at one minute intervals for 16 minutes. B. An ImageJ plugin was designed to plot the signal intensity of each concentration of 18mer at each time point to assign numerical values to each sample based on the gradient of the resultant curve during the linear phase of signal intensity increase (red line). Shown are representative curves for 50 pmoles oligomer and 0.1 pmoles oligomer. The results indicate that this method of quantification would allow for sensitive quantification of biotin amounts over a range of concentrations.

the signal intensity of each sample over time. The gradient of the linear phase of the resultant curve for each sample is then used to assign a numerical value to each sample, where the concentration of biotin in a sample is proportional to the gradient of the curve. Representative curves for samples containing 50 pmoles and 0.1 pmoles of biotin can be seen in figure 5-4 B. This method proved sensitive for determining differing amounts of biotin in samples as seen by the steeper gradient of the exponential phase of the curve for 50 pmoles in comparison to that for 0.1 pmoles biotin. It was decided that this method of detection would be suitable for our purposes were we to continue with slot blotting of samples. However, due to problems associated with inefficient removal of unligated adaptor (described in section 5.3.5-5.3.7) and due to technical inconsistencies with the slot blotting method itself, we decided that we would continue using a Southern blotting method (described in section 5.3.7) having established a system to optimize AtRNL ligation efficiency and 18mer oligonucleotide removal.

5.3.5 Optimizing Urea-PAGE loading and 18mer removal

To determine the amount of DNA that should be loaded to visualize AtRNL ligation efficiency in an *in vitro* reaction described in section 5.3.6, we subjected differing amounts of a 70-nucleotide oligomer, ranging from 0.1 to 50 pmoles, to denaturing urea polyacrylamide gel electrophoresis (Figure 5-5 A). It was determined that loading 1 pmole of 70mer provided the strongest signal without seeing signal derived from degradation of the 70mer.

Having established the optimal amount of 70mer to load, we wanted to look at the ability of Nucleospin Gel and PCR-Clean-up columns (Macherey Nagel) to remove any unligated 18mer from ligation reactions. To do this, 10 pmoles of 70mer and 50 pmoles of biotinylated 18mer in 50 μl water were added to 100 μl of different dilutions of NTI binding buffer and subjected to Nucleospin Gel and PCR-Clean-up. 1/10 of the eluate volume of each sample and an untreated control were then subjected to urea PAGE (Figure 5-5 B). The greatest amount of the 18mer oligonucleotide was removed by dilution in 100% buffer NTI, however this condition led to the greatest loss of the 70mer oligonucleotide, whereas as the concentration of buffer NTI decreased, greater levels of 70mer were retained but the ability to remove the 18mer decreased. It was decided to

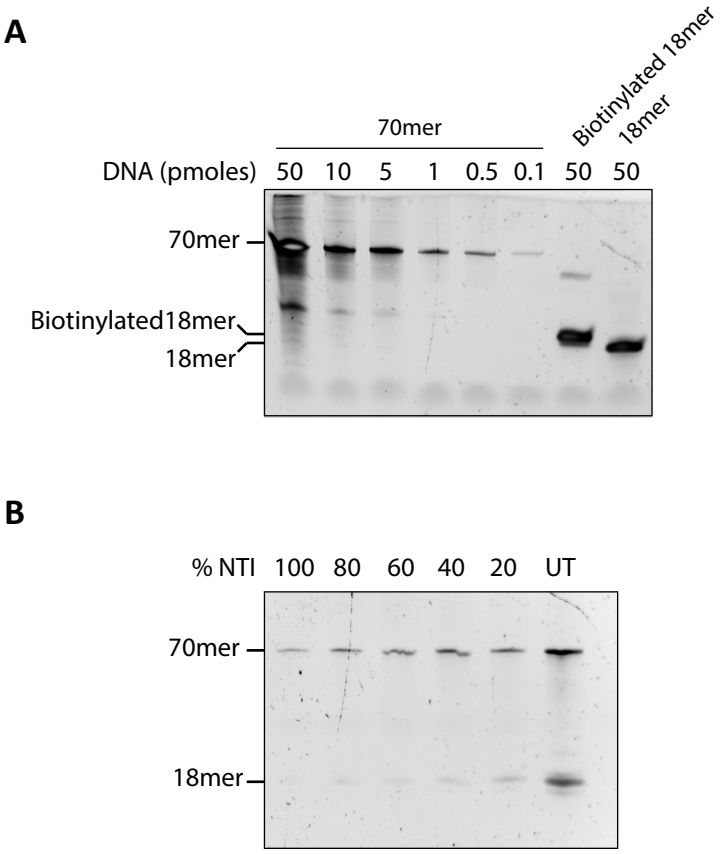


Figure 5-5 Optimising urea-PAGE DNA loading and 18mer removal
A. 5 µl of the indicated amounts of 70 nucleotide oligomer, internally biotinylated 18mer and unbiotinylated 18mer were subjected to urea-PAGE gel electrophoresis and stained with SYBR Green to determine the optimal quantity of DNA to load for future AtRNL ligation efficiency assay B. To assess the ability of Macherey-Nagel Gel and PCR cleanup columns to remove 18mers whilst retaining as much 70mer as possible, 10 pmoles of 70mer and 50 pmoles of biotinylated 18mer were suspended in 50 µl H_2O and diluted in 100 µl of the indicated dilutions of Macherey-Nagel buffer NTI then subjected to Macherey-Nagel Gel and PCR cleanup, eluting in 20 µL elution buffer. Following cleanup, 2 µL of each sample eluate and an untreated control (UT) were subjected to urea-PAGE gel electrophoresis. Dilution in undiluted buffer NTI prior to cleanup resulted in the greatest loss of 18mer.

use undiluted buffer NTI to clean-up ligation reactions since efficiency of unligated 18mer removal was a priority over retention of shorter DNA fragments, particularly since the rate of ribonucleotide incorporation in PU-seq strains of *S. pombe* would result in fragments with an average length of hundreds of nucleotides.

5.3.6 Optimizing AtRNL ligation efficiency

AtRNL was purified by Jonathan Wing, a colleague in the Carr laboratory using plasmid pET28a-At-TRL (Addgene plasmid # 32242), a kind gift from Professor Stan Fields (Schutz et al. 2010). An *in vitro* assay was designed to assess the catalytic activity of AtRNL on a substrate generated by alkaline hydrolysis of a 70-nucleotide oligomer containing an internal ribonucleotide towards the 3' end (rTOP), resulting in a 65-nucleotide substrate with a ribonucleotide terminus to which the 18 nucleotide internally biotinylated oligomer could be ligated. A parallel reaction was carried out on an oligonucleotide of the same length and same sequence but containing only deoxyribonucleotides (dTOP) (Figure 5-6 A). 100 pmoles of rTOP and dTOP were treated with 0.3 M NaOH at 55°C for 2 hours after which 10 pmoles of each substrate was used in a 30-minute room temperature ligation reaction with AtRNL and a 5-fold molar excess of biotinylated 18mer. Following ligation ~1 pmole of rTOP/dTOP ligation mixture and 5 pmoles of biotinylated 18mer were subjected to denaturing urea PAGE. All of the rTOP subjected to treatment with NaOH was hydrolyzed however only a small amount of the resultant substrate was ligated to the biotinylated 18mer indicating that whilst AtRNL was active, ligation efficiency was low.

We were unable to determine the concentration of AtRNL in the storage buffer since it co-purified with a large amount of a 70 kDa protein that we assumed to be a molecular chaperone. Since we assumed the concentration of AtRNL in the storage buffer to be low we hoped that by increasing the volume of AtRNL in the reaction, and by increasing incubation time we might increase ligation efficiency. Taking into account the negative effect of glycerol in the storage buffer on the ability of proteins to move freely in solution, we limited the total volume of AtRNL in the reaction to 20%. Increasing the percentage of AtRNL in the reaction mixture from 10% to 20% increased the amount of biotinylated 18mer ligated to alkali-treated rTOP and a further increase in ligation was

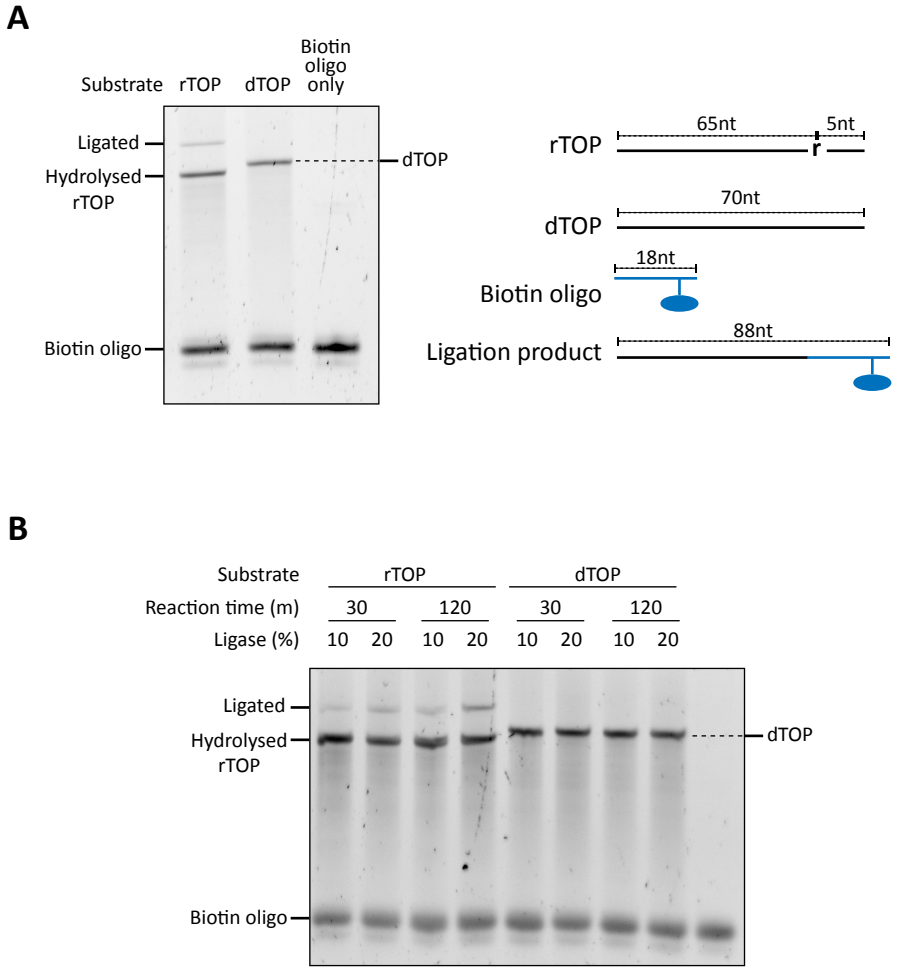


Figure 5-6 AtRNL shows specific activity on an alkali-treated ribonucleotide-containing 70mer

A. 100 pmoles of a ribonucleotide-containing 70mer (rTOP) and a DNA-only control (dTOP) were alkali-treated and used as substrate in a reaction with *Arabidopsis thaliana* tRNA ligase (AtRNL). AtRNL showed low efficiency specificity for substrate terminating in a 3'-ribonucleotide. B. Reactions were carried out as in A but using either 10% or 20% of the final volume of AtRNL and incubating for 30 or 120 minutes. The greatest ligation efficiency was seen using 20% AtRNL in a 120 minute reaction.

seen by incubating the reaction for 120 minutes over a 30-minute incubation (Figure 5-6 B).

5.3.7 AtRNL ligation assay with *S. pombe* gDNA and Southern blotting

Having established the best conditions for ligation and for removal of unligated 18mer, we set out to test the efficiency of our assay in the context of ligating the biotinylated 18mer onto a substrate derived from alkali-treated gDNA from a strain of *S. pombe* lacking RNase H2 activity due to deletion of *rnh201*. Following Nucleospin column cleanup, we resolved ligation reactions on agarose gels to help separate unligated 18mer by size from those higher molecular weight fragments resulting from ligation of the 18mer to ribo-terminating DNA fragments. It was hoped that this would overcome the problem caused by incomplete removal of unligated adapter that we would have encountered by slot blotting samples. With slot blotting, any unligated adaptor would bind to the membrane along with successfully ligated products, causing an unwanted additive effect on signal intensity.

500 ng of alkali treated *rnh201*Δ *S. pombe* gDNA was used as substrate in an AtRNL ligation reaction with 50 pmoles of biotinylated 18mer. After incubation at room temperature for 2 hours the ligation reaction was cleaned up using Nucleospin columns and samples resolved by agarose gel electrophoresis and Southern blotting. Samples were imaged in an ImageQuant LAS 4000 in 1-minute increments (Figure 5-7). Raw images taken at 1-, 4- and 7-minutes exposure time (Figure 5-7 i) show an abundance of signal at ~0.1 kb. Exposure for 4 and 7 minutes reveals the presence of fragments between ~0.3 and 1 kb, with fewer fragments >1kb. Visualizing the membrane using an ImageQuant function that colours saturated pixels in red shows that the signal produced by the high amount of the lowest molecular weight fragments may cause pixel saturation in areas of the membrane populated by higher molecular weight fragments (Figure 5-7 ii). In the event that this early saturation of pixels at molecular weights >0.3 kb was caused by the intensity of signal from the lowest molecular weight fragments, the membrane was cut at ~0.3 kb and re-exposed (Figure 5-7 iii) however the pixel saturation of fragments >0.3 kb after 7 minutes is comparable both before and after the

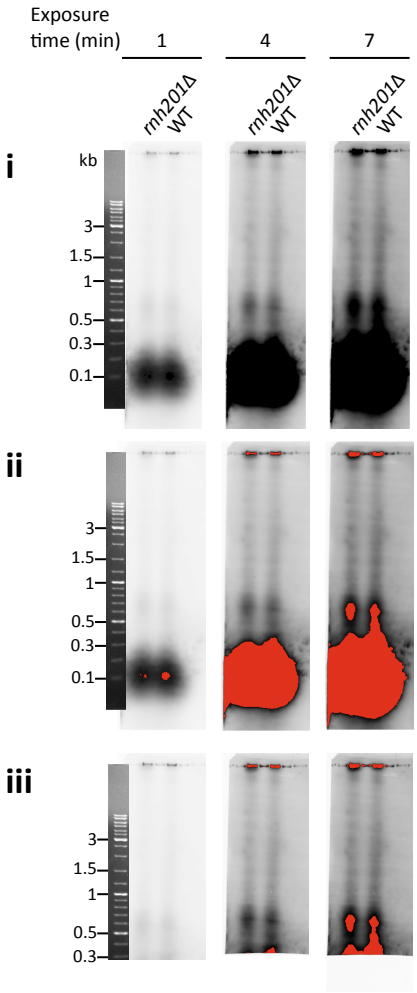


Figure 5-7 Macherey Nagel Nucleospin column cleanup
Genomic DNA from strains of *S. pombe* with the indicated genotypes were treated with 0.3 M NaOH at 55˚C for 2 hours. Hydrolysed DNA was ligated to a biotinylated 18mer oligonucleotide via 2' phosphate groups of the resultant ribonucleotide-terminating DNA fragments then run on an agarose gel followed by southern blotting and hybridisation with a streptavidin-AP probe. Blots were imaged with GE Imagequant LAS 4000 at 1 minute increments either as a raw image (i), images in which red pixels indicate pixel saturation (ii & iii). Blots in iii were cut just above the molecular weight that contained the highest signal (~0.1 kb; cut at ~0.3 kb). Selected timepoints only are shown. Whilst after longer exposures signal can be seen at a range of molecular weights >0.5 kb, the high intensity of signal at ~0.1 kb possibly indicates that Macherey Nagel Nucleospin columns did not completely remove unligated oligo.

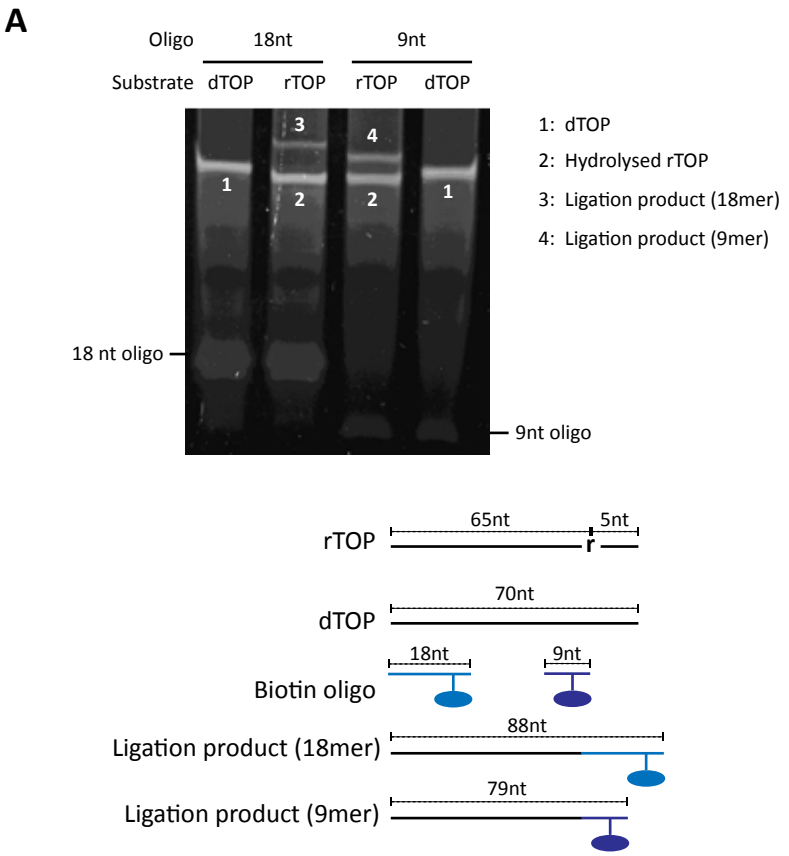
membrane was cut, suggesting that the high concentration of signal at ~0.1 kb does not affect signal intensity of fragments of 0.3 kb and greater.

This assay did however show that AtRNL is able to ligate short, biotinylated oligonucleotides onto ribo-terminating gDNA fragments. The majority of signal appears at low molecular weights, suggesting either that ribonucleotides are incorporated regularly with on average less than ~0.5 kb between them and that misincorporated ribonucleotides separated by more than 1 kb are rare, or that AtRNL ligates preferentially onto shorter fragments. Surprisingly, this assay showed very little difference in signal intensity between wild type and *rnh201*Δ gDNA.

5.3.8 Reducing background signal caused by biotinylated oligomer

It was assumed that the high amount of signal seen at ~0.1 kb in Figure 5-7 was caused either by unligated 18mer, or because AtRNL had a preference for ligating to short fragments of ribo-terminating gDNA. We decided to optimize the unligated oligomer removal step by shortening the length of the oligomer from 18 nucleotides to 9 nucleotides and by using Illustra MicroSpin G-50 columns (GE) which exclude fragments <20 bases using size exclusion rather than relying on matrix-binding efficiency as is the case with Nucleospin columns. We first tested the ability of AtRNL to ligate the biotinylated 9mer to alkali-treated rTOP using the same assay as described in section 5.3.6 and found that ligation was at least as efficient as ligation of the biotinylated 18mer (Figure 5-8 A). Furthermore, removal of the biotinylated 9mer from a mixed sample of 9mer and untreated rTOP using Microspin G-50 columns was highly efficient with very little 9mer visible after treatment, even at long exposure times (Figure 5-8 B).

We repeated the AtRNL ligation reaction described in section 5.3.7 with gDNA from a Pu-seq strain of *S. pombe* lacking *rnh201* and containing the ribo-incorporating Polε mutation M630F. A biotinylated 9mer oligonucleotide was used rather than the 18mer and sample clean-up was carried out using Microspin G-50 columns to remove unligated 9mer (Figure 5-9). To give an idea of how signal intensity may differ between samples containing low and high levels of genomic ribonucleotides, either 10 % or 90% respectively of the total ligation reaction was loaded onto the gel. Whilst there is a clear



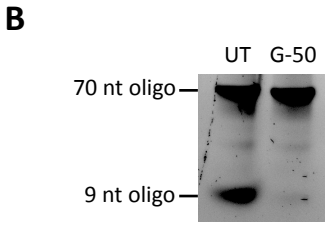


Figure 5-8 AtRNL ligates a 9 nucleotide biotinylated oligomer specifically onto ribonucleotide-terminating substrates

A. 100 pmoles of a ribonucleotide-containing 70mer (rTOP) and a DNA-only control (dTOP) were alkali-treated and used as substrate in a reaction with *Arabidopsis thaliana* tRNA ligase (AtRNL) and internally biotinylated 18mer or 9mer oligonucleotides. AtRNL showed specificity for ligating onto ribonucleotide terminating substrates only and ligated these ends to both 9mers and 18mers with similar efficiency. B. To assess the ability of Illustra G-50 columns to remove 9mers whilst retaining as much 70mer as possible, 10 pmoles of 70mer and 50 pmoles of biotinylated 9mer were suspended in 50 µl H_2O and subjected to G-50 column treatment. G-50 columns were found to be highly effective at retaining 9mers whilst eluting the majority of the 70mer input.

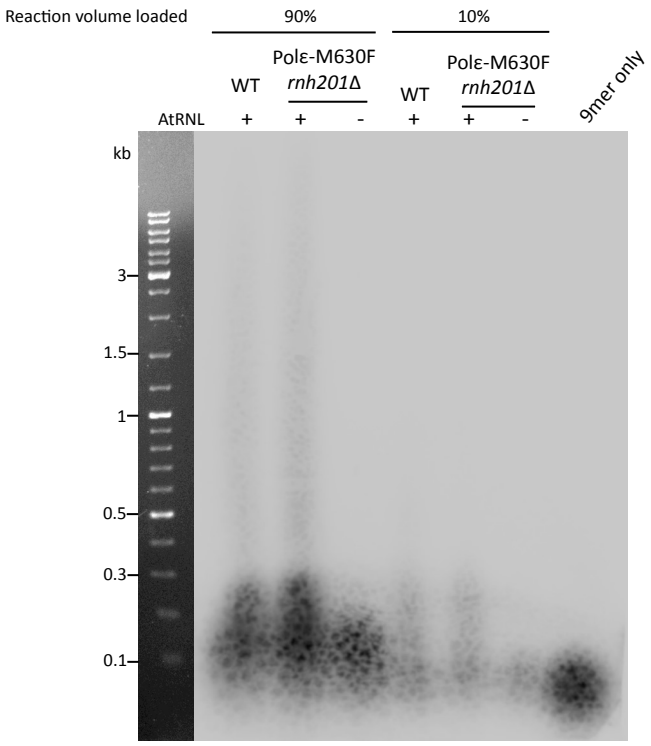


Figure 5-9 The AtRNL assay with Southern blotting lacks the desired sensitivity
Genomic DNA from strains of *S. pombe* with the indicated genotypes was alkali treated and used as a substrate in a reaction with AtRNL and biotinylated 9mer. Following incubation, samples were cleaned up using Illustra Nucleopsin G-50 columns and subjected to agarose gel electrophoresis and Soutern blotting. To assess the sensitivity of this assay to different levels of biotin, either 90% or 10% of the total reaction volume was loaded. Despite small differences in signal intensity between wild type and mutant genomic DNA, the difference was insufficient for reliable information to be obtained.

difference in signal intensity between gDNA treated with AtRNL and the no ligase control sample, there remains a significant amount of signal between 0.1 and 0.2 kb in the no ligase control sample, suggesting that in spite of the apparent robustness of 9mer removal exhibited in figure 5-8 B, this removal is incomplete. Furthermore, the difference in signal intensities between wild type DNA and Pu-seq DNA across the range of fragment sizes is not as marked as would have been expected and there still appeared to be a preference for ligation onto low molecular weight fragments.

In light of these concerns and due to the inactivity of AtRNL purified from a second batch (data not shown), we decided that despite the potential promise of this assay, two previously published assays would be sufficient for our purposes.

5.4 Using a FRET-based assay to assess RNase H2 activity

A key principle of the *S. pombe* Pu-seq system is an absence of RNase H2 function achieved by knocking out the RNase H2 catalytic subunit, RNase H2A. As discussed in chapter 4, RNase H2A knockout in many human cell lines is not possible, however we were able to develop two different methods of inducibly knocking down RNase H2C to levels undetectable by Western blot. Although levels of RNase H2C following degradation are undetectable by Western blot it is possible that a small amount of RNase H2C remains stably expressed. In this event it is possible that enough RNase H2 trimers form to efficiently undertake ribonucleotide excision repair. To ensure that knockdown of RNase H2C resulted in a decrease of RNase H2 catalytic activity we took advantage of a previously published fluorescence-based RNase H2 activity assay (White et al. 2013). The assay assesses the ability of RNase H2 in whole cell extract to cleave an oligonucleotide duplex containing an internal ribonucleotide in one strand. The strand containing the internal ribonucleotide contains a 3'-fluorescein fluorophore which is quenched by a 5'-dabcyl on the opposite strand. RNase H2-mediated hydrolysis of the ribonucleotide releases the fluorescein resulting in an increase in fluorescence signal. An oligonucleotide duplex substrate of the same sequence but lacking the internal ribonucleotide was used to assess the level of non-RNase H2-mediated fluorophore release (Figure 5-10 A).

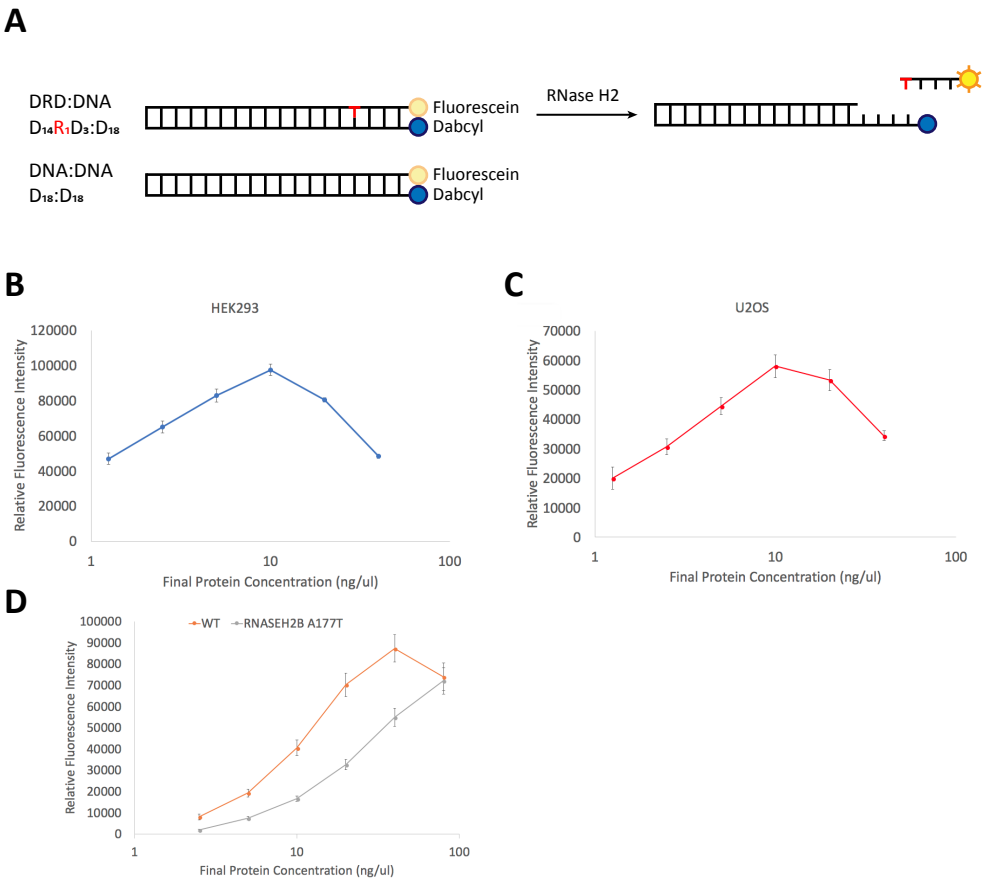


Figure 5-10 Design and implementation of a fluorescence-based assay to determine RNase H2 activity

A. Schematic representation of the RNase H2 activity assay substrate. RNase H2 activity on the ribonucleotide-containing duplex (DRD:DNA) causes release of the 3'-fluorescein fragment from the 5'-dabcyl quencher resulting in an increase in fluorescence intensity. A DNA-only substrate (DNA:DNA) is used as a control in some experiments to control for non-RNase H2 derived-substrate cleavage. B & C. The RNase H2 activity assay carried out using cell extract from HEK293 (B) and U2OS (C) over a range of protein concentrations. For both cell lines, maximum fluorescence intensity is seen using 10 ng/µl final protein concentration. D. RNase H2 activity assay using whole cell extract from wild type human lymphocytes and human cells containing a hypomorphic A177T mutation in RNase H2B shows the decrease in RNase H2 activity resulting from the RNase H2B mutation. Error bars: ± SDM from three independent biological replicates.

At the time of optimizing this assay, we had been trying to set up a system to downregulate RNase H2 using HEK-293 and U2OS cell lines. We first set out to determine the optimal final concentration of HEK-293 and U2OS whole cell extract to be used in each 100 μl reaction by carrying out an assay in which different concentrations of cell extract were used in a 90-minute reaction, after which an endpoint fluorescence readout was calculated after quenching the reaction with EDTA (Figure 5-10 B & C). This allowed us to determine a final protein concentration within the linear range of increase in fluorescence intensity that could be used in experiments looking at the increase in fluorescence over time. With whole cell extract from both HEK-293 (Figure 5-10 B) and U2OS (Figure 5-10 C) a linear increase in fluorescence was seen between 1.25 ng/μl and 10 ng/μl after which a decrease in fluorescence was observed. It is possible that incubation with 10 ng/μl whole cell extract for 90 minutes was sufficient for all ribonucleotide-containing substrate (DRD) to be cleaved, whilst cleavage of the DNA-only duplex proceeded at a slower rate. Since fluorescence values were calculated by subtracting the fluorescence readings from DNA-only substrate from those from DRD substrate, it is possible to account for the decrease in adjusted signal intensity due to the plateauing of DRD signal with a concurrent continued increase in fluorescence in samples containing DNA-only substrate with concentrations of whole cell extract over 10 ng/μl.

We repeated this assay using whole cell extract from wild type human lymphoblasts and from cells containing a hypomorphic A177T mutation in RNase H2B to assess the sensitivity of this assay in detecting differences in cellular RNase H2 activity (Figure 5-10 D). Pleasingly, differences in efficiency of substrate cleavage were seen between the two cell lines at final protein concentrations from 2.5 ng/μl up to 40 ng/μl. With wild type cell extract a decrease in adjusted fluorescence intensity was seen at 80 ng/μl, whilst fluorescence intensity for RNase H2B A177T continued to increase.

Since the rate of substrate cleavage by mutant RNase H2 would proceed at a slower rate than for wild type RNase H2, it would require a higher concentration of mutant RNase H2 than wild type RNase H2 for fluorescence intensity to plateau over 90 minutes. As

such, assuming that DNA-only substrate cleavage proceeds at a slower rate than DRD cleavage and at an equal rate in the presence of whole cell extract from either wild type cells or RNase H2 mutant cells, it is possible that after subtracting DNA-only fluorescence intensity from DRD fluorescence that a decrease in adjusted fluorescence intensity is seen. This observation further supports the theory that the decrease in adjusted fluorescence intensity seen above 10 ng/μl for HEK-293 and U2OS was caused by the adjustment calculation rather than by a true decrease in fluorescence intensity in samples containing the DRD substrate.

5.5 Assessing the efficiency of mAID-SMASh-mediated knockdown of RNase H2C in RPE1 cells using the FRET-based assay

To determine the extent to which RNase H2 activity in RPE1 whole cell extract was downregulated following degradation of RNase H2C via the mAID-SMASh double degron described in chapter 4, we carried out a kinetic time course assay to assess the rate of cleavage of the ribonucleotide-containing oligonucleotide duplex described in section 5.4. RPE1 cells conditionally expressing OsTIR1 in the presence of doxycycline but that were otherwise wild type, and RPE1 cells expressing mAID-SMASh-tagged RNase H2C were pre-treated with 1 μM doxycycline for 2 hours then grown in media containing 1 μM doxycycline, 1 mg/ml IAA and 3 μM ASV with whole cell extract taken before, or 2 and 24 hours after, addition of IAA and ASV. Due to time constraints we were unable to perform endpoint assays to determine the optimal concentration of RPE1 whole cell extract to be used in the time course assay so we used a final protein concentration that would ensure fluorescence signal saturation with wild type cell extract within the course of the experiment and additionally did not include the DNA-only duplex substrate to avoid problems of mean adjusted signal intensity decreasing beyond a certain level of substrate hydrolysis.

Due to time constraints it was not possible to carry out a comprehensive assay to determine a final protein concentration of RPE1 cell extract that would fall in the linear

range. Since the final protein concentration required varies between different cell lines (Figure 5-10) a quick assay was carried out using RPE1 cell extract to determine a suitable protein concentration to use in a time point assay (data not shown). From this assay it was determined that a final protein concentration of 100 ng/μl of RPE1 cell extract would be appropriate. For the time point assay, 100 ng/μl final protein concentration of whole cell extract was used in each 100 μl reaction and fluorescence intensity recorded at 10-minute intervals for 130 minutes (Figure 5-11). Treatment with doxycycline, IAA and ASV had no significant effect on overall RNase H2 activity in wild type cells, with the majority of substrate hydrolysed within 90 minutes, after which rates of hydrolysis proceeded at a markedly slower rate, presumably at least in part due to non-RNase H2-mediated fluorophore release. The rate of substrate hydrolysis in clones 5 (~43% of wild type activity) and 6 (~28%) is significantly lower than in wild type cells prior to degradation of RNase H2C suggesting that the presence of the C-terminal mAID-SMASh tag on RNase H2C causes a significant decrease in overall RNase H2 catalytic activity. Prior to addition of IAA and ASV, clone 5 exhibits a greater level of RNase H2 activity than clone 6; this is probably due to the higher levels of tagged RNase H2C expressed by clone 5 as seen by Western blot (Figure 4-13 and Figure 4-14) resulting in a greater number of RNase H2 trimers. A significant decrease in RNase H2 activity can be seen 2 hours after addition of IAA and ASV in both clones 5 and 6 (~22% and ~24% of wild type activity respectively). Following 24 hours growth in IAA and ASV, RNase H2 activity in both clone 5 and clone 6 is decreased to ~19% of wild type activity with no significant difference in enzymatic activity exhibited between the two clones. At least some of this residual fluorescence increase is presumed to be due to non-RNase H2-mediated fluorophore release, suggesting that activity levels of RNase H2 following degradation of RNase H2C are a maximum of 19% that of wild type RNase H2.

This assay provides a robust method to demonstrate that AID-SMASh-mediated degradation of RNase H2C in RPE1 cells causes a significant decrease in RNase H2 activity. To determine the true extent of the reduction in RNase H2 activity following knockdown of AID-SMASh-tagged RNase H2C in RPE1 cells, it would be advantageous to compare cell extract from these cells after treatment with Dox, IAA and ASV with cell

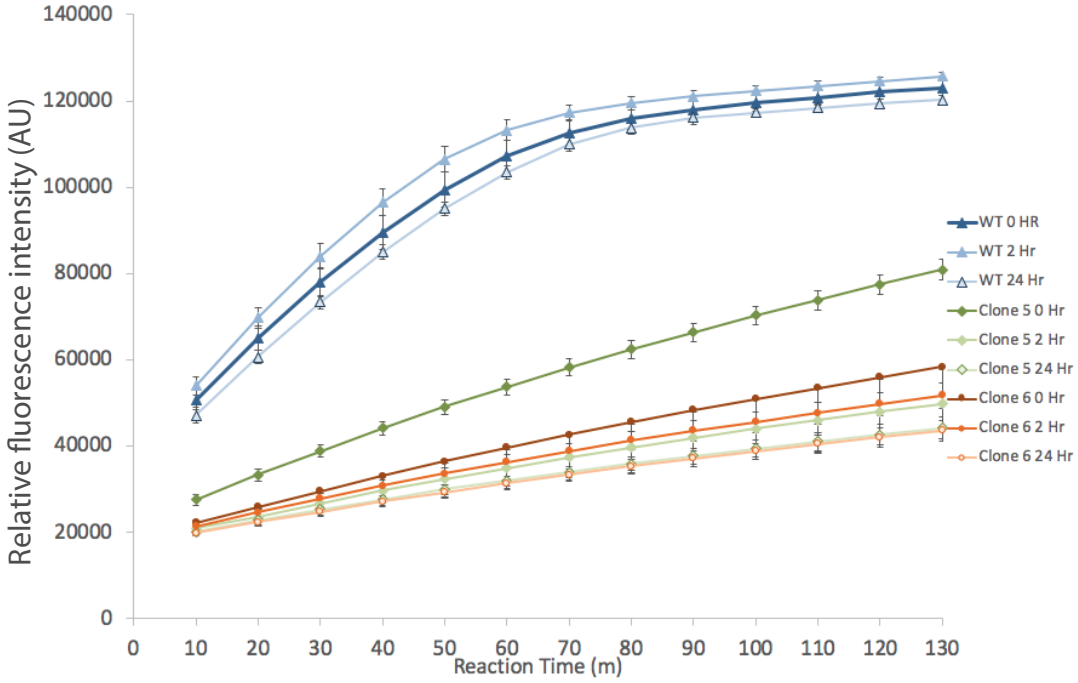


Figure 5-11 Fluorescence-based RNase H2 activity assay shows that depletion of RNase H2C causes a significant decrease in RNase H2 activity

Wild type RPE1 cells and two candidate RNase H2C mAID-SMASh double degron clones were pre-treated with 1 µM doxycycline for 2 hours then grown in media containing 1 µM doxycycline, 1 mg/ml IAA and 3 µM ASV for 24 hours with samples taken before addition of IAA and ASV, 2 and 24 hours after addition of IAA and ASV. 100 ng/µl of each sample was used in a kinetic assay time course with fluorescence intensity measured every 10 minutes. Knockdown of RNase H2C causes a marked decrease in RNase H2 activity after 24 hours. Error bars: ± SDM from three independent biological replicates.

extract from Aicardi-Goutières Syndrome patient cells and perhaps even with protein extracted from RNase H2A knockout *S. pombe*.

Whilst this assay is sensitive for determining differing levels of RNase H2 activity, it does not show that the decrease in activity results in an increase in ribonucleotide presence in the genome and would also therefore not be informative when analysing ribo-incorporating polymerase mutations. To this end, I also looked at using the alkaline comet assay to assess the extent of ribonucleotide incorporation.

5.6 Alkaline comet assay

Single cell gel electrophoresis, commonly called the comet assay, is a sensitive method to determine the extent of DNA damage at the single cell level. The presence of a ribonucleotide in genomic DNA renders the sugar phosphate backbone more susceptible to alkaline hydrolysis at that point. As such, to determine the extent of ribonucleotide presence in genomic DNA, the comet assay can be performed under alkaline conditions (Kind et al. 2018). To check that the alkaline comet assay would enable us to differentiate between cells containing different levels of genomic ribonucleotides, we subjected two Aicardi-Goutières Syndrome (AGS) patient cell lines to alkaline comet assay and compared them to wild type human TK6 lymphocytes. The AGS cell lines contained hypomorphic mutations in RNase H2B (A177T) or RNase H2A (G37S). The RNase $H2A^{G37S}$ mutation has a more severe impact on RNase H2 holoenzyme stability than RNase $H2BA^{177T}$ and RNase $H2A^{G37S}$ cells have been shown to contain a higher level of genomic ribonucleotides than RNase $H2BA^{177T}$ cells (Pizzi et al. 2015). These differences are further reflected in the slower growth rate of cells harbouring the RNase $H2A^{G37S}$ mutation over cells containing the RNase $H2BA^{177T}$ mutation. Consequently, we would expect the greatest mean comet tail length in cells harbouring the RNase $H2A^{G37S}$ mutation, with RNase $H2BA^{177T}$ cells providing a mean comet tail length intermediate between RNase $H2A^{G37S}$ and wild type cells.

For the alkaline comet assay ~3×10^4 cells of each genotype were embedded in 1.2% agarose on glass slides and lysed in a buffer containing 2.5M NaCl at pH10 for 1 hour

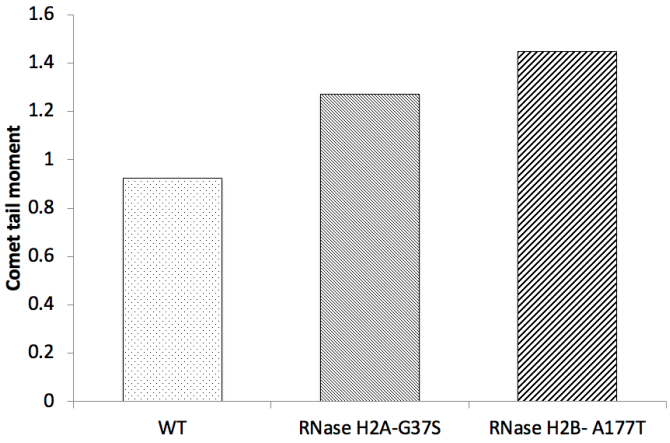


Figure 5-12 Alkaline comet assay to assess genomic ribonucleotide incorporation
Human lymphocytes with the stated genotypes were lysed in 2.5M NaCl at pH10 for 1 hour and subjected to alkaline electrophoresis. Mean comet tail moment was calculated from 50 cells. Decreased RNase H2 activity resulting from cells harbouring mutations in either RNase H2A or RNase H2B leads to increased comet tail moments in comparison to wild type cells.

and subjected to alkaline electrophoresis at 12V for 25 minutes. Migration of hydrolysed DNA towards the anode resulted in the formation of comet tails which were stained with SYBR Green and scored (Figure 5-12). Whilst the shortest comet tail lengths were seen for wild type cells, the comet tail length for was greater for cells containing the RNase H2BA^{177T} mutation that for cells containing the RNase H2A^{G37S} mutation. Since these are the results of one single biological non-blind replicate it is possible that the expected results would be seen with triplicate blind repeats. Furthermore, a greater difference in alkaline hydrolysis may be seen by lengthening the incubation in lysis buffer and running the samples in a more alkaline electrophoresis buffer. Similar studies use a 16-18 hour incubation in lysis buffer and a pH13 electrophoresis buffer in addition to synchronising cells by serum starvation (Gunther et al. 2015).

5.7 Discussion

In this chapter I have outlined the steps taken in attempting to develop a ribonucleotide incorporation assay using an *Arabidopsis thaliana* tRNA ligase (AtRNL). Whilst previously established methods of quantifying genomic ribonucleotide incorporation lack the specificity for recognizing DNA breaks caused only by the presence of an embedded ribonucleotide that the AtRNL assay would have offered, we encountered too many problems to make the assay worth pursuing.

In this chapter I also described an *in vitro* assay for analyzing the activity of RNase H2 holoenzyme in whole cell extract by looking at rates of hydrolysis of a ribonucleotide-containing oligonucleotide duplex substrate. I used this assay to demonstrate that overall RNase H2 catalysis is significantly decreased in the RNase H2C mAID-SMASh double degron RPE1 cell line following growth in doxycycline, IAA and ASV for 24 hours. Furthermore, I established that the catalytic activity of RNase H2 in the absence of IAA and ASV in the RNase H2C-mAID-SMASh candidate clone 5 is higher than for clone 6. As such, clone 5 is the best candidate to proceed with knocking in polymerase mutations since the ability to repair misincorporated ribonucleotide between PU-seq experiments will be greater in this clone than in clone 6. The importance of repairing genomic ribonucleotides between experiments is twofold: firstly, persistent genomic

ribonucleotide presence causes replication stress and genome instability (Pizzi et al. 2015) and secondly, the accumulation of genomic ribonucleotides over time in PU-seq cell lines could possible result in higher levels of background noise signal, decreasing the accuracy of experiments.

Finally, I have used a patient cell line harbouring hypomorphic mutant alleles of *RNASEH2A* (RNase H2A-G37S) and *RNASEH2B* (RNase H2B-A177T) to show that the alkaline comet assay offers the necessary sensitivity to detect different levels of genomic ribonucleotide presence. This will be important when screening the various polymerase mutants to assess the extent to which they misincorporate ribonucleotides.

Further work for the work outlined in this chapter therefore would be to knockin different mutations in polymerases delta and epsilon in the RNase H2-regulatable RPE1 cell line described in this chapter and in chapter 4 and to subject the resultant cell lines to the alkaline comet assay. Progress in screening the potential polymerase mutations is described in chapter 6.

Chapter 6

Investigating ribonucleotide-incorporating replicative DNA polymerase mutations

6.1 Introduction

One of the key features of the *S. pombe* polymerase usage sequencing (Pu-seq) system is the excess incorporation of ribonucleotides upon DNA replication by mutant alleles of the two major replicative polymerases, polymerases delta and epsilon. Decreased ribonucleotide discrimination by these mutant polymerases is achieved through mutating a residue adjacent to a steric gate tyrosine at the active sites (Polδ-L591G and Polε-M630F) which results in a decreased clash with the 2'-OH group of the incoming ribonucleotide (Nick McElhinny, Kumar, et al. 2010; Miyabe et al. 2011; Daigaku et al. 2015; Williams et al. 2016).

Whilst ribo-incorporating mutants of Polδ and Polε have been described and characterized in both *S. pombe* and *S. cerevisiae*, there is little evidence of such ribo-incorporating alleles in human cells. In selecting mutations that we hoped would result in increased ribonucleotide incorporation in human cells, we used sequence alignments to identify the conserved leucine and methionine in Polδ (Leu606) and Polε (Met630) respectively (Figure 6-1). In trying to identify the required polymerase mutations it is important to try a range of different substitutions. This is because in trying to generate strains of *S. pombe* containing the equivalent ribo-incorporating substitution used in *S. cerevisiae* (S. pombe: M630G; S. cerevisiae: M644G), the Carr laboratory discovered that this mutation was lethal in *S. pombe*. Since the equivalent substitution has different effects on cell viability between two closely related yeast species, it is likely that the same phenomenon will be observed when trying to generate equivalent mutations in human cells.

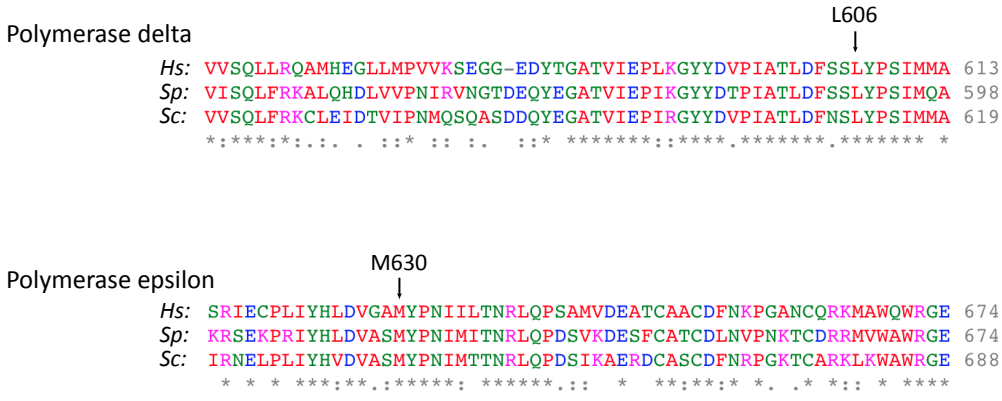


Figure 6-1 Identification of human polymerase residues to be targeted for mutagenesis

Sequence alignments of human, *S. pombe* and *S. cerevisiae* polymerase delta and polymerase epsilon catalytic subunits showing the conserved leucine (polymerase delta) and methionine (polymerase epsilon) to be targeted for mutagenesis.

For the human Polδ ribo-incorporating mutation, we took advantage of two Polδ mutations, L606M and L606G, that have been shown to incorporate higher than wild type levels of ribonucleotides in *in vitro* studies (Schmitt et al. 2010; Clausen et al. 2013). In addition, we decided to investigate a Polδ mutation, L606K, which does not incorporate elevated ribonucleotides *in vitro* (Schmitt et al. 2010), but does cause a mutator phenotype in heterozygously targeted mice, a phenomenon seen also with the L606G allele (Venkatesan et al. 2007).

Since there are no known mutations in human Polε that result in increased ribonucleotide incorporation either in cells or *in vitro*, five candidate mutations were chosen that it was hoped would decrease the steric gate effect of the adjacent tyrosine: M630F, M630L, M630V, M630A and M630G, in increasing order of likelihood of ribo-incorporation based upon the permissiveness of their side chains for allowing access of the ribonucleotide 2'-hydroxyl group.

In this chapter I will outline the progress made towards identifying mutations in polymerases delta and epsilon that will result in higher than wild type levels of ribonucleotide incorporation which can be combined with the RNase H2C degradation system described in chapter 4 to establish the polymerase usage sequencing system in human cells.

6.2 Generating human cell lines expressing mutant polymerase delta

6.2.1 Human TK6 cells can express Polδ-L606M heterozygously

To generate TK6 cells expressing mutant Polδ we designed a system whereby a mutant allele of *POLD1*, the gene encoding Polδ, would be expressed from cDNA integrated at the endogenous locus. To do this, a guide RNA was designed to target Cas9 to within 100bp of the *POLD1* start codon. The resultant double strand break was repaired using a repair template that upon repair would incorporate the mutant *POLD1*

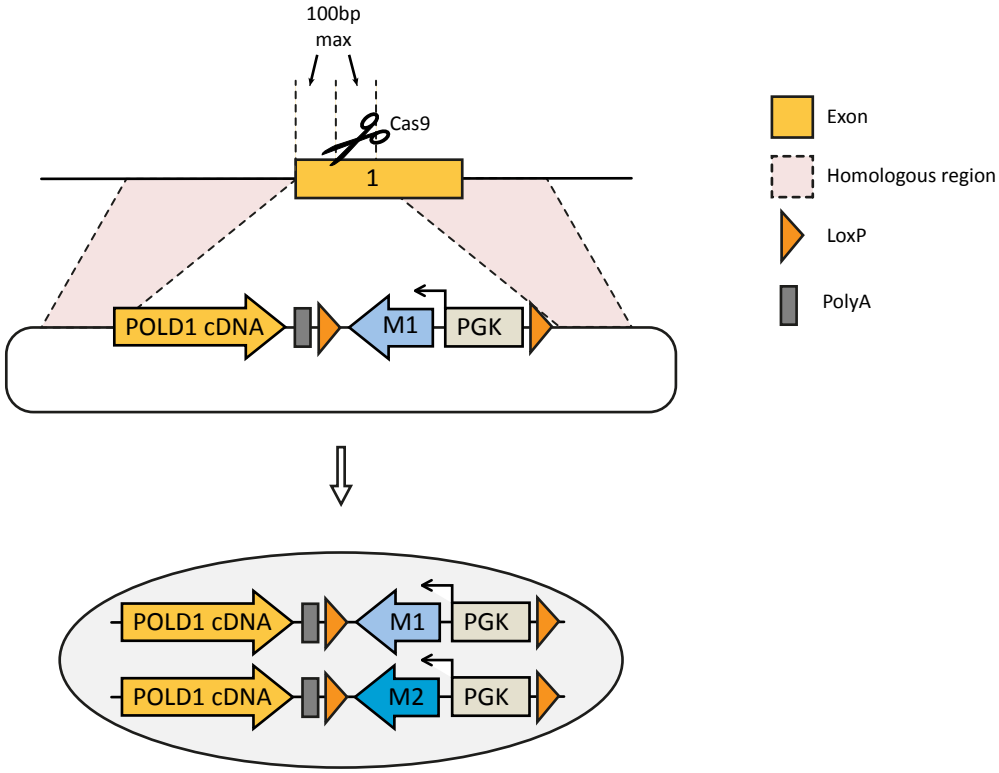


Figure 6-2 Polymerase delta mutation strategy
Cas9 is targeted to within 100bp of the *POLD1* start codon. The resultant break is repaired using a plasmid template that upon repair results in the introduction of *POLD1* cDNA and a LoxP-flanked antibiotic resistance marker under the control of a PGK promoter allowing for selection and subsequent marker removal. The use of repair templates containing different antibiotic resistance genes allows for selection of homozygously targeted clones.

cDNA followed by a PolyA signal and a LoxP-flanked antibiotic resistance marker under the control of its own promoter (Figure 6-2). In addition to attempting to knock in mutant *POLD1* cDNA we also targeted cells with a knockin construct that contained wild type *POLD1* cDNA. This was done as a control in the event that we failed to generate any positive clones after selection to allow us to know whether it was the presence of the mutation or the design of the system itself that was preventing clone growth. Initial work was carried out in the Takeda laboratory and sought only to investigate the Polδ-L606M mutation.

10^6 TK6 cells were co-transfected using Neon electroporation (Invitrogen; 1350 V, 10 ms, ×3 pulse, 10µl Neon tip) with two plasmids, one to express both Cas9 and a gRNA (pX330) and a repair template. For each transfection, cells were transfected with 200 ng pX330 and 800 ng histidinol^R repair template (WT or L606M). 48 hours after transfection, cells were selected in 500 µg/ml histidinol for 10 days. Following selection, we isolated 14 candidate clones for the L606M mutation and 8 for the wild type construct. To check for integration of the repair template at the correct locus, we screened each clone by PCR using a forward primer that annealed within the antibiotic resistance gene and a reverse primer that annealed outside the right homology arm (Figure 6-3 A). 5 of the L606M clones and 2 of the wild type clones had bands at the correct size. To check that the L606M candidate clones were expressing the mutant cDNA, each clone was subjected to RT-PCR and sequencing to check for the presence of mutant mRNA within the cells. Following sequencing, 4 of the 5 clones were found to express the mutant allele heterozygously, with no homozygously targeted clones isolated (Figure 6-3 B).

Subsequent attempts to generate homozygously targeted Polδ-L606M TK6 cells were unsuccessful and due to the problems associated with low transfection efficiency of TK6 cells described in chapter 3 it was decided to continue investigations into ribo-incorporating mutations in polymerases delta and epsilon in the human HT1080 cell line.

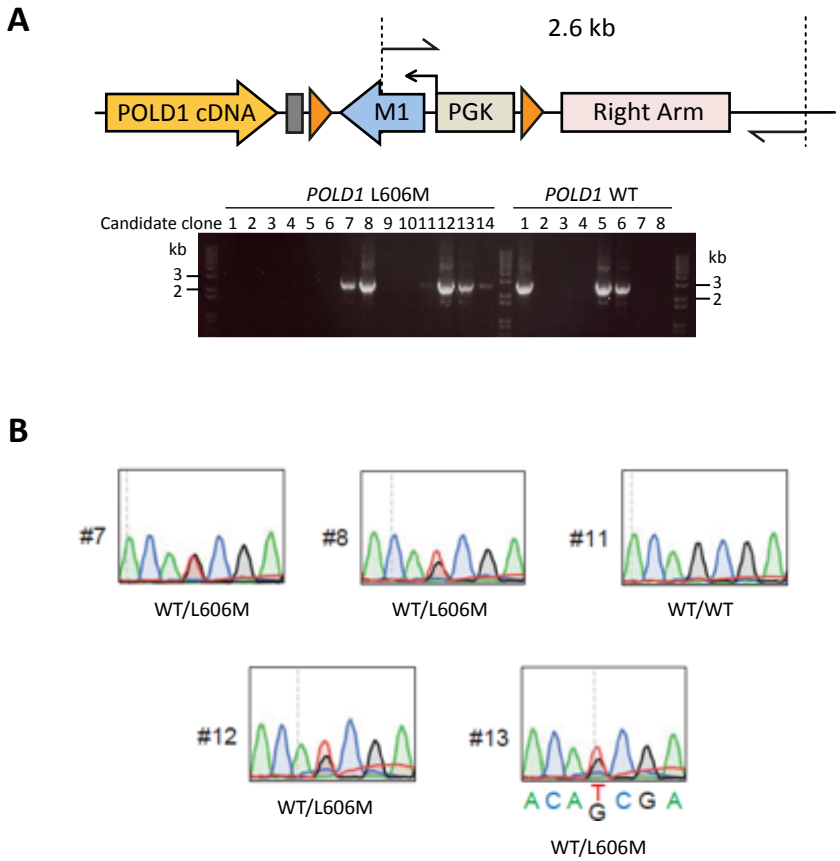


Figure 6-3 Polδ-L606M screening identifies heterozygously targeted TK6 cells
A. Following targeting with a Polδ-L606M or wild type Polδ knockin construct and selection, antibiotic-resistant clones were screened by PCR to check for integration of the knockin construct at the correct locus. B. Polδ-L606M clones that were correctly targeted were subjected to RT-PCR and sequencing. 4 of the 5 clones screened were found to be heterozygously targeted whilst one was untargeted. Sequencing traces show the sequence of the complementary strand: CAG = reverse complement of CTG (Leu); CAT = reverse complement of ATG (Met).

6.2.2 Human HT1080 cells can express Polδ-L606K heterozygously

To generate cells expressing Polδ-L606K, 10^6 HT1080 cells were transfected with 0.5 μg pX330 and 1.5 μg puromycin^R repair template using Amaxa nucleofection (Lonza; kit T, programme L-005). 48 hours after transfection, cells were selected in 0.4 μg/ml puromycin for two weeks after which 14 clones grew. Two of these 14 clones grew noticeably slower than the other 12 and eventually died following expansion into a larger volume of non-selective medium. Attempts to screen the remaining 12 clones for integration of the repair template at the correct locus were unsuccessful so a subsequent PCR screen was carried out to check for the presence of the puromycin resistance gene using forward and reverse primers that both annealed within the marker gene (Figure 6-4 A). All 12 clones screened were found to have the repair template integrated somewhere in their genomes. To check that the repair templates in these clones had integrated at the correct locus and that the cells were able to express Polδ-L606K, they were subjected to RT-PCR and sequencing (Figure 6-4 B). As with Polδ-L606M in TK6 cells, no homozygous mutants were found, but four clones (clones 1, 5, 6 & 12) were found to express Polδ-L606K heterozygously.

6.2.3 Failure to generate Polδ-L606G and Polδ-L606M in the HT1080 cell line

In parallel with work to generate the Polδ-L606K HT1080 cell line, attempts were also made to generate heterozygous and homozygous Polδ-L606G and Polδ-L606M mutant cell lines in addition to a cell line expressing Polδ-L606K homozygously. 10^6 HT1080 cells were transfected with 0.5 μg pX330 and 1.5 μg puromycin^R repair template or 0.5 μg pX330 and 0.75 μg each of repair templates conferring resistance to both blasticidin and puromycin to isolate heterozygously- or homozygously-targeted clones respectively. 48 hours after transfection, cells were selected in 0.4 μg/ml puromycin to select for heterozygously targeted clones or 0.4 μg/ml puromycin and 5 μg/ml blasticidin to select for homozygously targeted clones. In all instances, no clones grew after selection. A summary of attempts to generate HT1080 cell lines expressing mutant Polδ can be seen in Table 6-1.

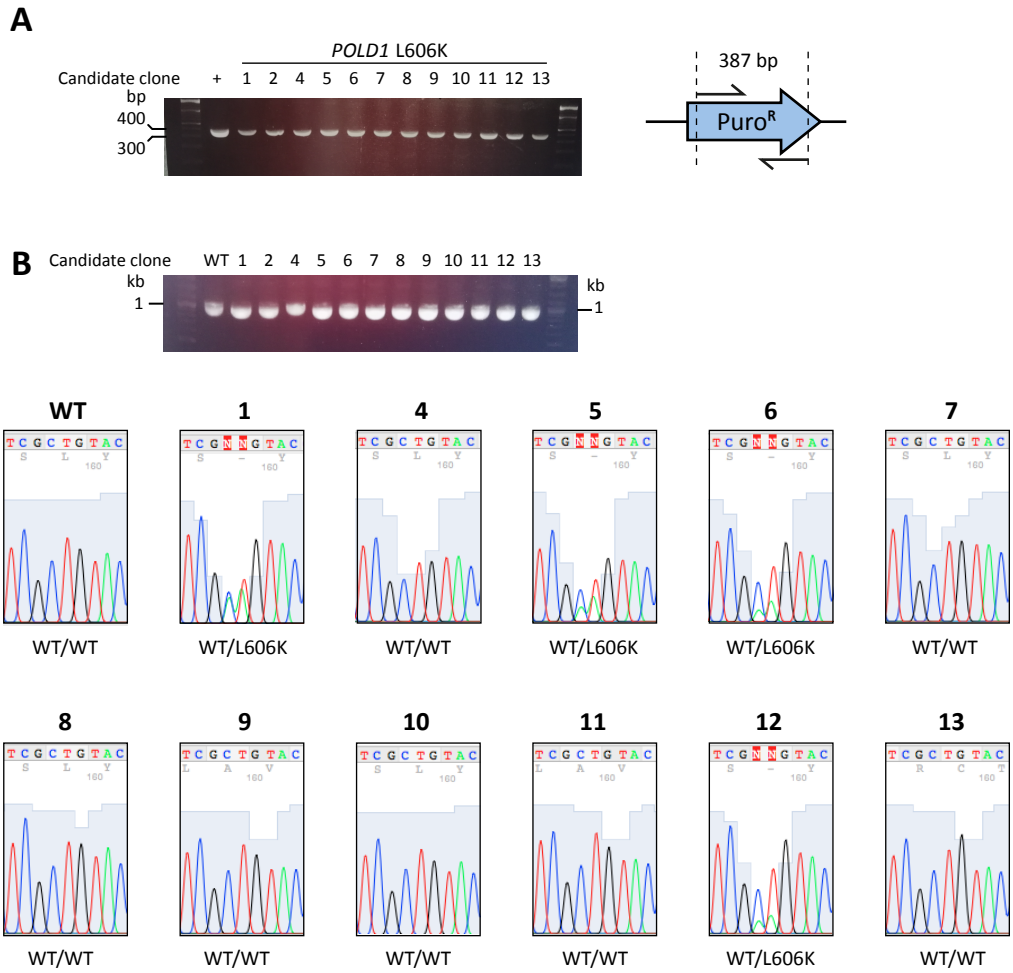


Figure 6-4 Polδ-L606K screening identifies heterozygously targeted HT1080 cells
A. Following targeting with a Polδ-L606K knockin construct and selection, antibiotic-resistant clones were screened by PCR to check for the presence of the puromycin resistance marker. The knockin construct was used as a positive control (+). B. Polδ-L606K clones that contained the puromycin resitance marker were subjected to RT-PCR and sequencing to check for integration of the knockin construct at the correct locus and to check that the construct was expressed. 4 of the 11 clones screened were found to be heterozygously targeted. Sequencing traces show the sequence of the coding strand: CTG (Leu); NNG = mixed signal: CTG (Leu) & AAG (Lys).

Mutation	Puromycin^R	Blasticidin^R Puromycin^R
L606G	Attempts: 2	Attempts: 1
L606K	Attempts: 1	Attempts: 1
L606M	Attempts: 2	Attempts: 1

Table 6-1 Summary of attempts to generate mutant Polδ HT1080 cell lines
Summary of attempts to generate heterozygously-targeted (Puromycin^R) and homozygously-targeted (Blasticidin^R-Puromycin^R) Polδ mutant HT1080 cell lines. White = succesfully targeted clones generated; blue = unsuccessful in HT1080 but succesfully targeted TK6 clones generated using histidinol^R; red = unsuccessful.

6.3 Generating human cell lines expressing mutant polymerase epsilon

Due to the large size of the cDNA of *POLE*, the gene encoding the catalytic subunit of human Polε (6861 nucleotides) we decided against trying to knock in mutation-containing *POLE* cDNA at the endogenous locus due to the size of the plasmid that would be required. Instead it was decided to directly mutate the codon coding for Met630 which resides in exon 17 of *POLE*. To introduce the mutations a guide RNA was designed to target Cas9 to an intronic region just upstream of exon 17. The resultant double strand break would be repaired by a plasmid repair template containing the desired mutation within the right homology arm. Successful repair of the break would also incorporate a LoxP-flanked antibiotic resistance gene under the control of its own promoter into the intron between exons 16 and 17 allowing for selection of successfully targeted cells (Figure 6-5). As a control to check whether the design of our knockin construct would be tolerated, a repair template containing no mutation in the right homology arm was also designed.

For each mutation (M630G, M630A, M630V, M630L & M630F) or wild type control knockin (heterozygous only), 10^6 HT1080 cells were transfected using Amaxa nucleofection (Lonza; kit T, programme L-005) with 0.5 μg pX330 and 1.5 μg blasticidin^R repair template or 0.5 μg pX330 and 0.75 μg each of blasticidin^R and puromycin^R repair templates to isolate heterozygously- and homozygously-targeted clones respectively. 48 hours after transfection cells were selected in 5 μg/ml blasticidin (heterozygous) or 5 μg/ml blasticidin and 0.4 μg/ml puromycin (homozygous).

We tried to generate heterozygous M630G, M630A, M630V, M630L and M630F Polε mutations twice and homozygous mutations once but in all cases no clones grew following selection. A summary of attempts to generate HT1080 cell lines expressing mutant Polε can be seen in Table 6-2. Following targeting with the wild type control repair template, several colonies grew, although many of them exhibited a significant slow growth phenotype. 13 clones were screened by PCR to check that the repair template had integrated at the desired locus using a forward primer that annealed

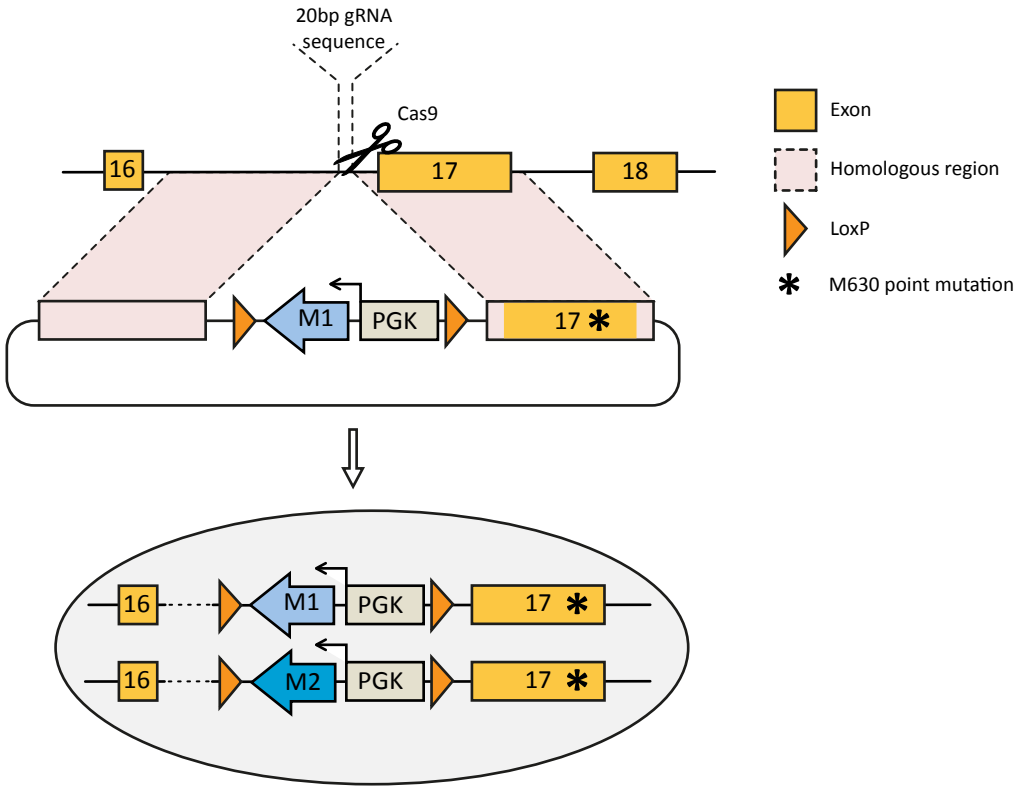


Figure 6-5 Polymerase epsilon mutation strategy
Cas9 is targeted to within the intron upstream of the *POLE* exon containing the codon to be targeted for mutagenesis. The resultant break is repaired using a plasmid template that upon repair results in the introduction of the desired point mutations. In addition a LoxP-flanked antibiotic resistance marker under the control of a PGK promoter is incorporated into the intronic region allowing for selection. Repair of the Cas9-induced DSB also excludes the original 20bp gRNA targeting sequence. The use of repair templates containing different antibiotic resistance genes allows for selection of homozygously targeted clones.

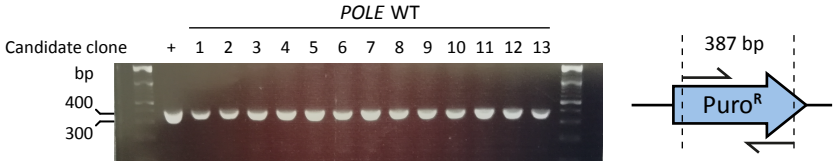


Figure 6-6 Screening for incorporation of the Polε-WT repair template
Following targeting with a Polε-WT knockin construct and selection, antibiotic-resistant clones were screened by PCR to check for the presence of the puromycin resistance marker. The knockin construct was used as a positive control (+). All 13 clones screened were found to contain the puromycin resistance gene somewhere in their genomes.

Mutation	Blasticidin^{R}	Puromycin^{R}	Blasticidin^{R} Puromycin^{R}
M630G	Attempts: 2	NA	Attempts: 1
M630A	Attempts: 2	NA	NA
M630V	Attempts: 2	NA	Attempts: 1
M630L	Attempts: 2	NA	Attempts: 1
M630F	Attempts: 2	NA	Attempts: 1
WT	Attempts: 1	Attempts: 1	NA

Table 6-2 Summary of attempts to generate mutant Polε HT1080 cell lines
Summary of attempts to generate heterozygously-targeted (Blasticidin^{R} or Puromycin^{R}) and homozygously-targeted (Blasticidin^{R}-Puromycin^{R}) Polε mutant HT1080 cell lines. White = succesfully targeted clones generated subject to further screening; red = unsuccessful; NA = not attempted.

within the puromycin resistance gene and a reverse primer that annealed downstream of the right homology arm; despite several stages of optimization we were unable to verify that these clones had been successfully targeted. To verify that the clones that survived selection did so due to the presence of the puromycin resistance marker somewhere in their genomes, a PCR screen was carried out using forward and reverse primers that both annealed within the puromycin^R fragment (Figure 6-6). All 13 clones screened were found to contain the puromycin^R fragment but due to time constraints we were unable to carry out further screening on these clones.

6.4 Discussion

In this chapter I have shown that it is possible to generate heterozygous L606M and L606K mutations in human Polδ. Whilst we cannot be certain from these results that it is not possible to generate homozygous Polδ mutants in human cells, mice harbouring homozygous mutations at the corresponding Polδ leucine (L604G and L604K) die *in utero* (Venkatesan et al. 2007) suggesting that in higher eukaryotes it is likely that it is essential to have at least one wild type Polδ allele. The failure to generate an HT1080 knockin cell line harbouring a Polδ-L606G mutation was surprising but it is thought that this failure was due to technical error rather than the mutation not being tolerated in human cells since we also failed to generate an HT1080 Polδ-L606M cell line despite the mutation being tolerated in TK6 cells and since the homologous heterozygous mutation (L604G) is tolerated in mice (Venkatesan et al. 2007); however *in vitro* studies suggest that human Polδ-L606G is 50-fold more error prone than wild type Polδ so it is possible that the presence of one copy of the Polδ-L606G is sufficient to cause a mutational burden that it is too great to be tolerated in human cells.

It is possible that one single ribo-incorporating Polδ mutant allele in combination with our RNase H2C knockdown system will result in a sufficient level of genomic ribonucleotides for human Pu-seq. It will therefore be important to attempt to introduce the Polδ mutations into the RPE1 cell line containing the RNase H2C double

degron system to assess the tolerance of the different mutations and to assess the extent to which the different Polδ mutants incorporate ribonucleotides.

In this chapter I have also described the failure to generate any mutations in Polε. Assuming that at least one of the clones obtained from targeting cells with the wild type Polε knockin construct is correctly targeted and that the slow growth phenotype observed for that clone is a result of the introduction of the knockin construct, it may be possible that the presence of the LoxP-flanked selection marker and promoter in the intron disrupt splicing activity of the resultant transcript. Furthermore, the design of the knockin construct results in the removal of the 20bp gRNA sequence following repair (Figure 6-5). It may therefore be possible that disrupted splicing in tandem with a mutant Polε may not be tolerated by cells. If this is the case, it would be worth trying to knockin mutant Polε cDNA at the endogenous locus using a similar strategy to that used for incorporating Polδ mutants in spite of the large size of the repair template plasmid that this strategy would necessitate. It could also be possible to attempt to mutate the M630 codon in a scarless manner although this method would necessitate a large workload to screen for successfully targeted clones.

In the event that we are unable to generate ribo-incorporating homozygous Polδ and/or Polε mutants or that the presence of a wild type allele in heterozygously targeted cells masks the ribo-incorporating effect of the mutant allele, it is possible to imagine a system to circumvent this problem using the HT1080 cell line described in section 4.6 in which RNase H2C is downregulated transcriptionally by doxycycline and at the protein level via an N-terminal SMASh tag. In this system, wild type Polδ or Polε cDNA would be ectopically expressed under the control of the same Tet-on promoter (pTRE3GS) as used for regulating expression of RNase H2C whilst the endogenous polymerase alleles would be targeted to introduce the ribo-incorporating mutations (either homozygously, or heterozygously with a knockout of the remaining wild type endogenous allele). Culturing these cells in doxycycline would therefore turn off expression of RNase H2C and the wild type ectopically expressed polymerase cDNA, resulting in temporary expression of the mutant endogenous allele(s) only.

Chapter 7

Discussion

This thesis describes progress made towards developing the cell lines necessary to carry out human Polymerase-usage sequencing using CRISPR-Cas9 technology.

The first results chapter showed that RNase H2A is essential in human TK6 cells. The essential nature of RNase H2A is perhaps unsurprising since it contains the catalytic domain of the RNase H2 complex (Reijns et al. 2011). Contrary to our findings, recently published work (Benitez-Guijarro et al. 2018) has been able to knock out RNase H2A in two different human cells lines, Hela and HCT116. In that study, RNase H2A knockout in HCT116 was only possible in cells lacking p53. Although the authors of that study do not mention the p53 status of the Hela cells that they used, it is possible that they were already p53 deficient, or that the low p53 expression in Hela cells (Hoppe-Seyler & Butz 1993) allows tolerance of excessive genomic ribonucleotides. The role of p53 in signalling damage caused by RNase H2 deficiency can be seen by the upregulation of p53-inducible genes in RNase H2 knockout mouse embryos and in mice harbouring a homozygous hypomorphic mutation in RNase H2B (Hiller et al. 2012b; Reijns et al. 2012).

The loss of either of the two RNase H2 accessory subunits, RNase H2B and RNase H2C, abolishes RNase H2 catalytic function and consequently RNase H2B and RNase H2C knockout mice display embryonic lethality (Hiller et al. 2012b; Reijns et al. 2012). In RNase H2B knockout mouse embryos, not only is RNase H2B itself undetectable by immunoblotting, so too are cellular levels of RNase H2A and RNase H2C (Reijns et al. 2012). Similarly, RNase H2A knockout in human cells causes a significant decrease in the levels of cellular RNase H2B and RNase H2C. Complementing RNase H2A knockout cells with wild type RNase H2A restores cellular levels of all three RNase H2 subunits to wild type levels (Benitez-Guijarro et al. 2018). These results suggest that the absence of

one RNase H2 subunit is sufficient to destabilise the other two subunits and therefore that all three subunits are necessary for complex stability and catalytic activity.

Since RNase H2 knockout is probably not possible without concomitant loss, or low expression of p53, in the second results chapter we looked at ways of abolishing RNase H2 activity via targeted degradation of one of the three subunits. This method offers two distinct advantages over permanent loss of RNase H2 function. Firstly, whilst cells can tolerate loss of RNase H2 function in the absence of p53, it is not known whether the additional burden of elevated ribonucleotide incorporation by mutant polymerases would still be tolerated. Secondly, the ability to restore RNase H2 function between experiments would allow for the removal of ribonucleotides misincorporated under experimental conditions. This would help to restrict ribonucleotide-derived DNA damage and reduce background ribonucleotide signal.

Whilst we were able to tag the C-terminus of RNase H2A with a mini auxin-inducible degron (mAID), RNase H2A was not degraded in the presence of IAA, the drug necessary for mAID-induced degradation. We therefore sought to target the two RNase H2 accessory subunits and to investigate different methods of downregulation.

The second approach we took was to attempt to tag the C-termini of RNase H2B and RNase H2C at their endogenous loci with a novel double degron tag consisting of an mAID tag fused with a SMASh tag, termed mAID-SMASh. We were unable to tag the C-terminus of RNase H2B with mAID-SMASh in RPE1 cells, perhaps indicating a requirement for the C-terminus of RNase H2B in mediating interactions with the other two RNase H2 subunits. We were however able to tag the C-terminus of RNase H2C with the mAID-SMASh double degron tag in RPE1 cells and to show that following degradation, cellular RNase H2C levels were undetectable by immunoblotting. Furthermore, in the third results chapter we showed that the knockdown of RNase H2C was sufficiently robust to greatly reduce cellular RNase H2 activity.

The third approach taken to downregulate RNase H2 activity was to ectopically express N-terminally SMASh-tagged RNase H2B and RNase H2C at the ROSA26 locus, under the

control of tetracycline controlled transcriptional activation (TCTA) in HT1080 cells. We were unable to generate cells expressing the tagged RNase H2B, despite the sustained expression of wild type RNase H2B in these cells. We assumed that the tagged RNase H2B was overexpressed sufficiently to cause several-fold excess in comparison to endogenous RNase H2B and that the excessive amounts of tagged RNase H2B somehow had a destabilising effect on RNase H2 complex formation.

We were able to generate HT1080 cells expressing N-terminally SMASh tagged RNase H2C from the ROSA26 under control of TCTA. We were able to show that transcriptional shutoff in the absence of doxycycline/tetracycline was rapid and efficient. It is assumed that SMASh-mediated degradation also contributes to the absence of tagged RNase H2C signal seen by immunoblotting. Further work in this cell line will require the endogenous copies of RNase H2C to be knocked out and for subsequent RNase H2 activity assays to be carried out as described in results chapter 3. The *RNASEH2C* cDNA that was knocked in using this construct was designed with several silent point mutations such that it would not be targeted by the Cas9 system used to target the endogenous, wild type *RNASEH2C* sequences.

We have demonstrated that a C-terminal mAID-SMASh tag can be used to robustly degrade endogenously tagged RNase H2C in RPE1 cells and that TCTA causes rapid shutoff of transcription of N-terminally SMASh-tagged RNase H2C at the ROSA26 locus in HT1080 cells. Further work can therefore investigate knocking in C-terminally mAID-SMASh tagged RNase H2C under TCTA at the ROSA26 locus. This would allow for a rapid and reversible knockdown of RNase H2C in three ways: i) by turning off transcription of the tagged gene, ii) by degrading any protein synthesised after transcriptional shutoff in the event of transcriptional leakiness and iii) by degrading any pre-existing protein via the mAID degron tag. It is hoped that this would further reduce the level of cellular RNase H2 activity to below the ~23% of wild type seen using the endogenously mAID-SMASh tagged RNase H2C system.

The third results chapter describes the steps taken in developing a novel assay for determining genomic ribonucleotide incorporation. This assay would have allowed us

to assess the extent of decreased ribonucleotide excision repair following knockdown of RNase H2 and the effects on ribonucleotide incorporation of different mutations in polymerases δ and ε (Pols δ and ε). Unfortunately, this assay lacked the sensitivity that we had hoped to achieve and after showing that a fluorescence-based oligonucleotide cleavage assay and the alkaline comet assay would allow us assess decreased RNase H2 activity and increased polymerase ribonucleotide incorporation, we ceased further optimization of the assay.

The fourth results chapter outlines progress made towards generating mutant alleles of Pols δ and ε that incorporate excess ribonucleotides. We have shown that it is possible to create human cell lines harbouring heterozygous L606M and L606K mutations in Polδ. Polδ-L606M has been shown to incorporate 7-fold more ribonucleotides than wild type Polδ *in vitro* (Clausen et al. 2013), whereas Polδ-L606K is more accurate than wild type Polδ *in vitro* (Schmitt et al. 2010), however little is known about the ribo-incorporating effects of these mutations *in vivo*.

We were unable to generate human cells containing a Polδ-L606G mutation. This may have been a result of technical error, however the L606G mutation causes Polδ to be highly inaccurate *in vitro* (Schmitt et al. 2010) so it is possible that Polδ-L606G causes levels of mutations that are not tolerated in human cells, even if the mutation is only present at one allele. Interestingly, mice harbouring a heterozygous LG mutation at the corresponding residue (Polδ-L604G) have a longer life expectancy and a lower rate of tumourigenesis than mice harbouring the Polδ-L604K mutation (Venkatesan et al. 2007), suggesting that the accuracy of polymerases *in vitro* may not be a complete indicator of the mutational potential of mutant polymerases *in vivo*.

The fourth results chapter also outlined our failure to generate any mutations at the active site of Polε. It is possible that the introduction of nearly ~1.8kb of DNA from the knockin construct into the intron between exons 16 and 17 negatively affects splicing sufficiently that cells die. Due to time constraints, we were unable to verify that this was the case using a knockin construct containing wild type sequence. Should future

work fail to generate ribonucleotide-incorporating mutant alleles of Polε in human cells, it may be possible to investigate the possibility of using proofreading-deficient Polε. The exonucleolytic proofreading activity of *S. cerevisiae* Polε accounts for the removal of approximately one third of misincorporated ribonucleotides (Williams et al. 2012). It is therefore possible that the absence of Polε proofreading may elevate the level of genomic ribonucleotides incorporated by a mutant Polε to a level that is sufficient to detect differences over wild type Polε ribonucleotide incorporation. Encouragingly, mice harbouring homozygous loss-of-proofreading point mutations in the Polε exonuclease domain are viable but have increased tumourigenesis (Albertson et al. 2009), suggesting that the corresponding loss-of-proofreading mutations may be tolerated in human cells.

The work in this thesis has shown that RNase H2 can be successfully downregulated in two different human cell lines. Furthermore, in RPE1 cells it has been shown that the downregulation of RNase H2C causes a significant decrease in cellular RNase H2 activity. These discoveries, and the verification that previously published assays to determine RNase H2 activity and genomic ribonucleotide presence are suitable for our purposes, lay a good foundation for further development of the human Pu-seq system. The RNase H2 shutoff cell lines can be used to knock in polymerase mutations to assess the extent of ribonucleotide incorporation by each of the mutations. Once the desired polymerase mutations are identified, these cell lines can be used for Pu-seq experiments to help map the location of origins in human cells and to investigate whether responses to replication stress seen by the Carr laboratory in *S. pombe* are also true in human cells.

Bibliography

Abel, S. & Theologis, A., 1996. Early genes and auxin action. *Plant physiology*, 111(1), pp.9–17.

Adli, M., 2018. The CRISPR tool kit for genome editing and beyond. *Nature Communications*, 9(1), p.1911. Available at: https://doi.org/10.1038/s41467-018-04252-2.

Akamatsu, Y. & Kobayashi, T., 2015. The Human RNA Polymerase I Transcription Terminator Complex Acts as a Replication Fork Barrier That Coordinates the Progress of Replication with rRNA Transcription Activity. *Molecular and cellular biology*, 35(10), pp.1871–1881.

Albertson, T.M. et al., 2009. DNA polymerase epsilon and delta proofreading suppress discrete mutator and cancer phenotypes in mice. *Proceedings of the National Academy of Sciences of the United States of America*, 106(40), pp.17101–17104.

Anders, C. et al., 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*. Available at: http://www.nature.com/doifinder/10.1038/nature13579%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/25079318 [Accessed July 28, 2014].

Arlt, M.F. et al., 2006. Common fragile sites as targets for chromosome rearrangements. *DNA repair*, 5(9–10), pp.1126–1135.

Azvolinsky, A. et al., 2009. Highly transcribed RNA polymerase II genes are impediments to replication fork progression in Saccharomyces cerevisiae. *Molecular cell*, 34(6), pp.722–734.

Bahler, J. et al., 1998. Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. *Yeast (Chichester, England)*, 14(10), pp.943–951.

Balakrishnan, L. & Bambara, R.A., 2013. Okazaki fragment metabolism. *Cold Spring Harbor perspectives in biology*, 5(2).

Bartkova, J. et al., 2006. Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature*, 444(7119), pp.633–637.

Benitez-Guijarro, M. et al., 2018. RNase H2, mutated in Aicardi-Goutieres syndrome, promotes LINE-1 retrotransposition. *The EMBO journal*, 37(15).

Bermejo, R. et al., 2011. The replication checkpoint protects fork stability by releasing transcribed genes from nuclear pores. *Cell*, 146(2), pp.233–246.

Bermejo, R., Lai, M.S. & Foiani, M., 2012. Preventing replication stress to maintain genome stability: resolving conflicts between replication and transcription. *Molecular cell*, 45(6), pp.710–718.

Bertoli, C., Klier, S., et al., 2013. Chk1 inhibits E2F6 repressor function in response to replication stress to maintain cell-cycle transcription. *Current biology : CB*, 23(17), pp.1629–1637.

Bertoli, C. et al., 2016. Sustained E2F-Dependent Transcription Is a Key Mechanism to Prevent Replication-Stress-Induced DNA Damage. *Cell reports*, 15(7), pp.1412–1422.

Bertoli, C., Skotheim, J.M. & de Bruin, R.A.M., 2013. Control of cell cycle transcription during G1 and S phases. *Nature reviews. Molecular cell biology*, 14(8), pp.518–528.

Besnard, E. et al., 2012. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nature structural & molecular biology*, 19(8), pp.837–844.

Bester, A.C. et al., 2011. Nucleotide deficiency promotes genomic instability in early stages of cancer development. *Cell*, 145(3), pp.435–446.

Bhaya, D., Davison, M. & Barrangou, R., 2011. CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. *Annual Review of Genetics*, 45(1), pp.273–297. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22060043 [Accessed July 9, 2014].

Bhowmick, R. & Hickson, I.D., 2017. The "enemies within": regions of the genome that are inherently difficult to replicate. *F1000Research*, 6, p.666.

Bianconi, E. et al., 2013. An estimation of the number of cells in the human body. *Annals of human biology*, 40(6), pp.463–471.

Bibikova, M. et al., 2001. Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Molecular and cellular biology*, 21(1), pp.289–297.

Boch, J. et al., 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science (New York, N.Y.)*, 326(5959), pp.1509–1512.

Boehm, E.M., Gildenberg, M.S. & Washington, M.T., 2016. The Many Roles of PCNA in Eukaryotic DNA Replication. *The Enzymes*, 39, pp.231–254.

Boyer, A.-S. et al., 2013. The human specialized DNA polymerases and non-B DNA: vital relationships to preserve genome integrity. *Journal of molecular biology*, 425(23), pp.4767–4781.

Brambati, A. et al., 2015. Replication and transcription on a collision course: eukaryotic regulation mechanisms and implications for DNA stability. *Frontiers in genetics*, 6, p.166.

Brewer, B.J. & Fangman, W.L., 1988. A replication fork barrier at the 3' end of yeast ribosomal RNA genes. *Cell*, 55(4), pp.637–643.

Brown, J.A. & Suo, Z., 2011. Unlocking the sugar "steric gate" of DNA polymerases. *Biochemistry*, 50(7), pp.1135–1142.

Bubeck, D. et al., 2011. PCNA directs type 2 RNase H activity on DNA replication and repair substrates. *Nucleic acids research*, 39(9), pp.3652–3666.

Burge, S. et al., 2006. Quadruplex DNA: sequence, topology and structure. *Nucleic acids research*, 34(19), pp.5402–5415.

Burgers, P.M.J. & Kunkel, T.A., 2017. Eukaryotic DNA Replication Fork. *Annual review of biochemistry*, 86, pp.417–438.

Busen, 1980. Purification, subunit structure, and serologicai analysis of calf thymus ribonuclease H I. *The Journal of biological chemistry*, 255(19), pp.9434–9443.

Byun, T.S. et al., 2005. Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. *Genes & development*, 19(9), pp.1040–1052.

Capecchi, M.R., 1989. Altering the genome by homologous recombination. *Science (New York, N.Y.)*, 244(4910), pp.1288–1292.

Carr, A.M. & Lambert, S., 2013. Replication stress-induced genome instability: the dark side of replication maintenance by homologous recombination. *Journal of molecular biology*, 425(23), pp.4733–4744.

Carvalho, C.M.B. et al., 2015. Absence of heterozygosity due to template switching during replicative rearrangements. *American journal of human genetics*, 96(4),

pp.555–564.
Caskey, C.T. & Kruh, G.D., 1979. The HPRT locus. *Cell*, 16(1), pp.1–9.
Cayrou, C., Gregoire, D., et al., 2012. Genome-scale identification of active DNA replication origins. *Methods (San Diego, Calif.)*, 57(2), pp.158–164.
Cayrou, C., Coulombe, P., et al., 2012. New insights into replication origin characteristics in metazoans. *Cell cycle (Georgetown, Tex.)*, 11(4), pp.658–667.
Cerritelli, S.M. & Crouch, R.J., 2009. Ribonuclease H: the enzymes in eukaryotes. *The FEBS journal*, 276(6), pp.1494–1505.
Chen, Y.-H. et al., 2015. ATR-mediated phosphorylation of FANCI regulates dormant origin firing in response to replication stress. *Molecular cell*, 58(2), pp.323–338.
Chilkova, O. et al., 2007. The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of PCNA. *Nucleic acids research*, 35(19), pp.6588–6597.
Chon, H. et al., 2009. Contributions of the two accessory subunits, RNASEH2B and RNASEH2C, to the activity and properties of the human RNase H2 complex. *Nucleic acids research*, 37(1), pp.96–110.
Christian, M. et al., 2010. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, 186(2), pp.757–761.
Chung, H.K. et al., 2015. Tunable and reversible drug control of protein production via a self-excising degron. *Nature chemical biology*, 11(9), pp.713–720.
Ciccia, A. & Elledge, S.J., 2010. The DNA damage response: making it safe to play with knives. *Molecular cell*, 40(2), pp.179–204.
Clausen, A.R. et al., 2013. Ribonucleotide incorporation , proofreading and bypass by human DNA polymerase ʕ. *DNA Repair*, 12(2), pp.121–127. Available at: http://dx.doi.org/10.1016/j.dnarep.2012.11.006.
Clausen, A.R. et al., 2015. Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation. *Nature structural & molecular biology*, 22(3), pp.185–191.
Clyne, R.K. & Kelly, T.J., 1995. Genetic analysis of an ARS element from the fission yeast Schizosaccharomyces pombe. *The EMBO journal*, 14(24), pp.6348–6357.
Cobb, J.A. et al., 2005. Replisome instability, fork collapse, and gross chromosomal rearrangements arise synergistically from Mec1 kinase and RecQ helicase mutations. *Genes & development*, 19(24), pp.3055–3069.
Cong, L. et al., 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, N.Y.)*, 339(6121), pp.819–23. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3795411&tool=pmcentrez&rendertype=abstract [Accessed July 9, 2014].
Crompton, N.E.A. et al., 2002. A single low dose of X-rays induces high frequencies of genetic instability (aneuploidy) and heritable damage (apoptosis), dependent on cell type and p53 status. *Mutation research*, 517(1–2), pp.173–186.
Crow, Y.J. et al., 2015. Characterization of human disease phenotypes associated with mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and IFIH1. *American journal of medical genetics. Part A*, 167A(2), pp.296–312.
Crow, Y.J., Leitch, A., et al., 2006. Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutieres syndrome and mimic congenital viral brain infection. *Nature genetics*, 38(8), pp.910–916.
Crow, Y.J., Hayward, B.E., et al., 2006. Mutations in the gene encoding the 3’-5’ DNA

exonuclease TREX1 cause Aicardi-Goutieres syndrome at the AGS1 locus. *Nature genetics*, 38(8), pp.917–920.

Daigaku, Y. et al., 2015. A global profile of replicative polymerase usage. , 22(3).

Deegan, T.D. & Diffley, J.F.X., 2016. MCM: one ring to rule them all. *Current opinion in structural biology*, 37, pp.145–151.

Deltcheva, E. et al., 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340), pp.602–607.

Deshpande, A.M. & Newlon, C.S., 1996. DNA replication fork pause sites dependent on transcription. *Science (New York, N.Y.)*, 272(5264), pp.1030–1033.

Doudna, J. a. & Charpentier, E., 2014. The new frontier of genome engineering with CRISPR-Cas9. *Science*, 346(6213), pp.1258096–1258096. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.1258096 [Accessed November 28, 2014].

Douglas, M.E. et al., 2018. The mechanism of eukaryotic CMG helicase activation. *Nature*, 555(7695), pp.265–268.

Dungrawala, H. et al., 2015. The Replication Checkpoint Prevents Two Types of Fork Collapse without Regulating Replisome Stability. *Molecular cell*, 59(6), pp.998–1010.

Eder, P.S., Walder, R.Y. & Walder, J.A., 1993. Substrate specificity of human RNase H1 and its role in excision repair of ribose residues misincorporated in DNA. *Biochimie*, 75(1–2), pp.123–126.

Evrin, C. et al., 2009. A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proceedings of the National Academy of Sciences of the United States of America*, 106(48), pp.20240–20245.

Fernandez-Cid, A. et al., 2013. An ORC/Cdc6/MCM2-7 complex is formed in a multistep reaction to serve as a platform for MCM double-hexamer assembly. *Molecular cell*, 50(4), pp.577–588.

Ferraro, P. et al., 2010. Quantitation of cellular deoxynucleoside triphosphates. *Nucleic acids research*, 38(6), p.e85.

Figiel, M. et al., 2011. The Structural and Biochemical Characterization of Human RNase H2 Complex Reveals the Molecular Basis for Substrate Recognition and Aicardi-Goutie ` res Syndrome Defects * □. , 286(12), pp.10540–10550.

Forstemann, K. & Lingner, J., 2005. Telomerase limits the extent of base pairing between template RNA and telomeric DNA. *EMBO reports*, 6(4), pp.361–366.

Frick, D.N. & Richardson, C.C., 2001. DNA primases. *Annual review of biochemistry*, 70, pp.39–80.

Fu, Y. et al., 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature biotechnology*, 32(3), pp.279–84. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24463574 [Accessed July 9, 2014].

Gadaleta, M.C. & Noguchi, E., 2017. Regulation of DNA Replication through Natural Impediments in the Eukaryotic Genome. *Genes*, 8(3).

Gaj, T. et al., 2012. Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. *Nature methods*, 9(8), pp.805–807.

Gambus, A. et al., 2011. MCM2-7 form double hexamers at licensed origins in Xenopus egg extract. *The Journal of biological chemistry*, 286(13), pp.11855–11864.

Gan, H. et al., 2017. Checkpoint Kinase Rad53 Couples Leading- and Lagging-Strand DNA Synthesis under Replication Stress. *Molecular cell*, 68(2), p.446–455.e3.

Garcia-Benitez, F., Gaillard, H. & Aguilera, A., 2017. Physical proximity of chromatin to nuclear pores prevents harmful R loop accumulation contributing to maintain genome stability. *Proceedings of the National Academy of Sciences of the United States of America*, 114(41), pp.10942–10947.

Garcia-Rubio, M. et al., 2018. Yra1-bound RNA-DNA hybrids cause orientation-independent transcription-replication collisions and telomere instability. *Genes & development*, 32(13–14), pp.965–977.

Garg, P. et al., 2004. Idling by DNA polymerase delta maintains a ligatable nick during lagging-strand DNA replication. *Genes & development*, 18(22), pp.2764–2773.

Gasiunas, G. et al., 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 109(39), pp.E2579-86.

Ge, X.Q. & Blow, J.J., 2010. Chk1 inhibits replication factory activation but allows dormant origin firing in existing factories. *The Journal of cell biology*, 191(7), pp.1285–1297.

Ge, X.Q., Jackson, D.A. & Blow, J.J., 2007. Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress. *Genes & development*, 21(24), pp.3331–3341.

Ghodgaonkar, M.M. et al., 2013. Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair. *Molecular cell*, 50(3), pp.323–332.

Gilson, E. & Geli, V., 2007. How telomeres are replicated. *Nature reviews. Molecular cell biology*, 8(10), pp.825–838.

Glover, T.W. et al., 1984. DNA polymerase alpha inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. *Human genetics*, 67(2), pp.136–142.

Goksenin, A.Y. et al., 2012. Human DNA polymerase epsilon is able to efficiently extend from multiple consecutive ribonucleotides. *The Journal of biological chemistry*, 287(51), pp.42675–42684.

Gossen, M. et al., 1995. Transcriptional activation by tetracyclines in mammalian cells. *Science (New York, N.Y.)*, 268(5218), pp.1766–1769.

Gossen, M. & Bujard, H., 1992. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences of the United States of America*, 89(12), pp.5547–5551.

Gravells, P. et al., 2015. Use of the HPRT gene to study nuclease-induced DNA double-strand break repair. *Human molecular genetics*, 24(24), pp.7097–7110.

Guilinger, J.P., Thompson, D.B. & Liu, D.R., 2014. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nature biotechnology*, 32(6), pp.577–82. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24770324 [Accessed July 9, 2014].

Gunther, C. et al., 2015. Defective removal of ribonucleotides from DNA promotes systemic autoimmunity. *The Journal of clinical investigation*, 125(1), pp.413–424.

Heinz, N. et al., 2011. Retroviral and transposon-based tet-regulated all-in-one vectors with reduced background expression and improved dynamic range. *Human gene therapy*, 22(2), pp.166–176.

Helmrich, A. et al., 2006. Common fragile sites are conserved features of human and mouse chromosomes and relate to large active genes. *Genome research*, 16(10),

pp.1222–1230.

Helmrich, A., Ballarino, M. & Tora, L., 2011. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Molecular cell*, 44(6), pp.966–977.

Hiller, B. et al., 2012a. Mammalian RNase H2 removes ribonucleotides from DNA to maintain genome integrity. *The Journal of experimental medicine*, 209(8), pp.1419–1426.

Hiller, B. et al., 2012b. Mammalian RNase H2 removes ribonucleotides from DNA to maintain genome integrity. , 209(8), pp.1419–1426.

Hoa, N.N. et al., 2015. Relative contribution of four nucleases, CtIP, Dna2, Exo1 and Mre11, to the initial step of DNA double-strand break repair by homologous recombination in both the chicken DT40 and human TK6 cell lines. *Genes to cells : devoted to molecular & cellular mechanisms*, 20(12), pp.1059–1076.

Hoppe-Seyler, F. & Butz, K., 1993. Repression of endogenous p53 transactivation function in HeLa cervical carcinoma cells by human papillomavirus type 16 E6, human mdm-2, and mutant p53. *Journal of virology*, 67(6), pp.3111–3117.

Hsu, P.D. et al., 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology*, 31(9), pp.827–32. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3969858&tool=pmcentrez&rendertype=abstract [Accessed July 10, 2014].

Hua, Y. et al., 2017. A simple and efficient method for CRISPR/Cas9-induced mutant screening. *Journal of genetics and genomics = Yi chuan xue bao*, 44(4), pp.207–213.

Huang, S.-Y.N. et al., 2017. Topoisomerase I-mediated cleavage at unrepaired ribonucleotides generates DNA double-strand breaks. *The EMBO journal*, 36(3), pp.361–373.

Hubscher, U. & Maga, G., 2011. DNA replication and repair bypass machines. *Current opinion in chemical biology*, 15(5), pp.627–635.

Im, J.-S. et al., 2015. RecQL4 is required for the association of Mcm10 and Ctf4 with replication origins in human cells. *Cell cycle (Georgetown, Tex.)*, 14(7), pp.1001–1009.

Inagaki, H. et al., 2013. Two sequential cleavage reactions on cruciform DNA structures cause palindrome-mediated chromosomal translocations. *Nature communications*, 4, p.1592.

Jensen, R.B., Carreira, A. & Kowalczykowski, S.C., 2010. Purified human BRCA2 stimulates RAD51-mediated recombination. *Nature*, 467(7316), pp.678–683.

Jeong, H.-S. et al., 2004. RNase H2 of Saccharomyces cerevisiae is a complex of three proteins. *Nucleic acids research*, 32(2), pp.407–414.

Jiang, W. et al., 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology*, 31(3), pp.233–239.

Jiang, Y. et al., 2009. Common fragile sites are characterized by histone hypoacetylation. *Human molecular genetics*, 18(23), pp.4501–4512.

Jiang, Y. et al., 2014. Discovery of danoprevir (ITMN-191/R7227), a highly selective and potent inhibitor of hepatitis C virus (HCV) NS3/4A protease. *Journal of medicinal chemistry*, 57(5), pp.1753–1769.

Jinek, M. et al., 2012. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, 337(6096), pp.816–821. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22745249 [Accessed July 9, 2014].

Jinek, M. et al., 2013. RNA-programmed genome editing in human cells. *eLife*, 2, p.e00471.

Kanke, M. et al., 2012. Mcm10 plays an essential role in origin DNA unwinding after loading of the CMG components. *The EMBO journal*, 31(9), pp.2182–2194.

Kara, N. et al., 2015. Orc1 Binding to Mitotic Chromosomes Precedes Spatial Patterning during G1 Phase and Assembly of the Origin Recognition Complex in Human Cells. *The Journal of biological chemistry*, 290(19), pp.12355–12369.

Katou, Y. et al., 2003. S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex. *Nature*, 424(6952), pp.1078–1083.

Keszthelyi, A. et al., 2015. Mapping ribonucleotides in genomic DNA and exploring replication dynamics by polymerase usage sequencing (Pu-seq). *Nature protocols*, 10(11), pp.1786–1801.

Keszthelyi, A., Minchell, N.E. & Baxter, J., 2016. The Causes and Consequences of Topological Stress during DNA Replication. *Genes*, 7(12).

Kidane, D., Murphy, D.L. & Sweasy, J.B., 2014. Accumulation of abasic sites induces genomic instability in normal human gastric epithelial cells during Helicobacter pylori infection. *Oncogenesis*, 3, p.e128.

Kim, J.H. et al., 2011. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PloS one*, 6(4), p.e18556.

Kim, N. et al., 2011. Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I. *Science (New York, N.Y.)*, 332(6037), pp.1561–1564.

Kim, S., Kim, D., Cho, S.W., Kim, J. & Kim, J.-S., 2014. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome research*, 24(6), pp.1012–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4032847&tool=pmcentrez&rendertype=abstract [Accessed July 9, 2014].

Kim, S., Kim, D., Cho, S.W., Kim, J. & Kim, J., 2014. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. , pp.1012–1019.

Kim, Y.G., Cha, J. & Chandrasegaran, S., 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America*, 93(3), pp.1156–1160.

Kind, B. et al., 2018. Single Cell Gel Electrophoresis for the Detection of Genomic Ribonucleotides. *Methods in molecular biology (Clifton, N.J.)*, 1672, pp.311–318.

Koh, K.D. et al., 2015. Ribose-seq: global mapping of ribonucleotides embedded in genomic DNA. *Nature methods*, 12(3), p.251–7, 3 p following 257.

Kuchta, R.D. et al., 1992. Inhibition of DNA primase and polymerase alpha by arabinofuranosylnucleoside triphosphates and related compounds. *Biochemistry*, 31(19), pp.4720–4728.

Labib, K., 2010. How do Cdc7 and cyclin-dependent kinases trigger the initiation of chromosome replication in eukaryotic cells? *Genes & development*, 24(12), pp.1208–1219.

Ladenburger, E.-M., Keller, C. & Knippers, R., 2002. Identification of a binding region for human origin recognition complex proteins 1 and 2 that coincides with an origin of DNA replication. *Molecular and cellular biology*, 22(4), pp.1036–1048.

Lalle, P. et al., 1995. Genomic stability and wild-type p53 function of lymphoblastoid cells with germ-line p53 mutation. *Oncogene*, 10(12), pp.2447–2454.

Lamarre, D. et al., 2003. An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus. *Nature*, 426(6963), pp.186–189.

Lambert, S. & Carr, A.M., 2013. Impediments to replication fork movement : stabilisation , reactivation and genome instability. , pp.33–45.

Langston, L.D. et al., 2014. CMG helicase and DNA polymerase epsilon form a functional 15-subunit holoenzyme for eukaryotic leading-strand DNA replication. *Proceedings of the National Academy of Sciences of the United States of America*, 111(43), pp.15390–15395.

Langston, L.D. et al., 2017. Mcm10 promotes rapid isomerization of CMG-DNA for replisome bypass of lagging strand DNA blocks. *eLife*, 6.

Li, P. et al., 2017. Aicardi-Goutieres syndrome protein TREX1 suppresses L1 and maintains genome integrity through exonuclease-independent ORF1p depletion. *Nucleic acids research*, 45(8), pp.4619–4631.

Li, X. et al., 1995. Lagging strand DNA synthesis at the eukaryotic replication fork involves binding and stimulation of FEN-1 by proliferating cell nuclear antigen. *The Journal of biological chemistry*, 270(38), pp.22109–22112.

Li, Y. & Breaker, R.R., 1999. Kinetics of RNA Degradation by Specific Base Catalysis of Transesterification Involving the 2'-Hydroxyl Group. *Journal of the American Chemical Society*, 121(23), pp.5364–5372. Available at: https://doi.org/10.1021/ja990592p.

Liu, L., Rice, M.C. & Kmiec, E.B., 2001. In vivo gene repair of point and frameshift mutations directed by chimeric RNA / DNA oligonucleotides and modified single-stranded oligonucleotides. , 29(20), pp.4238–4250.

Loew, R. et al., 2010. Improved Tet-responsive promoters with minimized background expression. *BMC biotechnology*, 10, p.81.

Lopes, J. et al., 2011. G-quadruplex-induced instability during leading-strand replication. *The EMBO journal*, 30(19), pp.4033–4046.

Lopes, M., Foiani, M. & Sogo, J.M., 2006. Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions. *Molecular cell*, 21(1), pp.15–27.

Lossaint, G. et al., 2013. FANCD2 binds MCM proteins and controls replisome function upon activation of s phase checkpoint signaling. *Molecular cell*, 51(5), pp.678–690.

Lucca, C. et al., 2004. Checkpoint-mediated control of replisome-fork association and signalling in response to replication pausing. *Oncogene*, 23(6), pp.1206–1213.

Lujan, S.A. et al., 2013. Ribonucleotides are signals for mismatch repair of leading-strand replication errors. *Molecular cell*, 50(3), pp.437–443.

Lundin, C. et al., 2002. Different roles for nonhomologous end joining and homologous recombination following replication arrest in mammalian cells. *Molecular and cellular biology*, 22(16), pp.5869–5878.

Lygerou, Z. & Nurse, P., 1999. The fission yeast origin recognition complex is constitutively associated with chromatin and is differentially modified through the cell cycle. *Journal of cell science*, 112 (Pt 2, pp.3703–3712.

Macheret, M. & Halazonetis, T.D., 2015. DNA replication stress as a hallmark of cancer. *Annual review of pathology*, 10, pp.425–448.

Mackenzie, K.J. et al., 2016. Ribonuclease H2 mutations induce a cGAS/STING-dependent innate immune response. *The EMBO journal*, 35(8), pp.831–844.

Makarova, K.S. et al., 2006. A putative RNA-interference-based immune system in

prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology direct*, 1, p.7.

Mali, P., Aach, J., et al., 2013. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature biotechnology*, 31(9), pp.833–838.

Mali, P., Yang, L., et al., 2013. RNA-guided human genome engineering via Cas9. *Science (New York, N.Y.)*, 339(6121), pp.823–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3712628&tool=pmcentrez&rendertype=abstract [Accessed July 9, 2014].

Manzo, S.G. et al., 2018. DNA Topoisomerase I differentially modulates R-loops across the human genome. *Genome biology*, 19(1), p.100.

Marahrens, Y. & Stillman, B., 1992. A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science (New York, N.Y.)*, 255(5046), pp.817–823.

Marchal, C. et al., 2018. Genome-wide analysis of replication timing by next-generation sequencing with E/L Repli-seq. *Nature protocols*, 13(5), pp.819–839.

Martinez-Lopez, A. et al., 2018. SAMHD1 deficient human monocytes autonomously trigger type I interferon. *Molecular immunology*, 101, pp.450–460.

Mazouzi, A., Velimezi, G. & Loizou, J.I., 2014. DNA replication stress: causes, resolution and disease. *Experimental cell research*, 329(1), pp.85–93.

McPhee, F. et al., 2012. Preclinical Profile and Characterization of the Hepatitis C Virus NS3 Protease Inhibitor Asunaprevir (BMS-650032). *Antimicrobial agents and chemotherapy*, 56(10), pp.5387–5396.

Mechali, M., 2010. Eukaryotic DNA replication origins: many choices for appropriate answers. *Nature reviews. Molecular cell biology*, 11(10), pp.728–738.

Mirkin, S.M., 2007. Expandable DNA repeats and human disease. *Nature*, 447(7147), pp.932–940.

Miyabe, I., Kunkel, T.A. & Carr, A.M., 2011. The Major Roles of DNA Polymerases Epsilon and Delta at the Eukaryotic Replication Fork Are Evolutionarily Conserved. , 7(12).

Mizuno, K. et al., 2009. Nearby inverted repeats fuse to generate acentric and dicentric palindromic chromosomes by a replication template exchange mechanism. *Genes & development*, 23(24), pp.2876–2886.

Mizuno, K. et al., 2013. Recombination-restarted replication makes inverted chromosome fusions at inverted repeats. *Nature*, 493(7431), pp.246–249.

Montecucco, A. et al., 1998. DNA ligase I is recruited to sites of DNA replication by an interaction with proliferating cell nuclear antigen: identification of a common targeting mechanism for the assembly of replication factories. *The EMBO journal*, 17(13), pp.3786–3795.

Muramatsu, S. et al., 2010. CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol (epsilon}, and GINS in budding yeast. *Genes & development*, 24(6), pp.602–612.

Naito, Y. et al., 2015. CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics (Oxford, England)*, 31(7), pp.1120–1123.

Natsume, T. et al., 2016. Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. *Cell reports*, 15(1), pp.210–218.

Necsulea, A. et al., 2009. The relationship between DNA replication and human genome

organization. *Molecular biology and evolution*, 26(4), pp.729–741.

Newlon, C.S., 1997. Putting it all together: building a prereplicative complex. *Cell*, 91(6), pp.717–720.

Nguyen, H.D. et al., 2017. Functions of Replication Protein A as a Sensor of R Loops and a Regulator of RNaseH1. *Molecular cell*, 65(5), p.832–847.e4.

Nick McElhinny, S.A., Watts, B.E., et al., 2010. Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proceedings of the National Academy of Sciences of the United States of America*, 107(11), pp.4949–4954.

Nick McElhinny, S.A. et al., 2008. Division of labor at the eukaryotic replication fork. *Molecular cell*, 30(2), pp.137–144.

Nick McElhinny, S.A., Kumar, D., et al., 2010. Genome instability due to ribonucleotide incorporation into DNA. *Nature chemical biology*, 6(10), pp.774–781.

Nieduszynski, C.A., Knox, Y. & Donaldson, A.D., 2006. Genome-wide identification of replication origins in yeast by comparative genomics. *Genes & development*, 20(14), pp.1874–1879.

Nishimura, K. et al., 2009. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature methods*, 6(12), pp.917–922.

Niu, H. et al., 2016. Roles of DNA helicases and Exo1 in the avoidance of mutations induced by Top1-mediated cleavage at ribonucleotides in DNA. *Cell cycle (Georgetown, Tex.)*, 15(3), pp.331–336.

Nowotny, M. et al., 2007. Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. *Molecular cell*, 28(2), pp.264–276.

O'Connell, K., Jinks-Robertson, S. & Petes, T.D., 2015. Elevated Genome-Wide Instability in Yeast Mutants Lacking RNase H Activity. *Genetics*, 201(3), pp.963–975.

O'Donnell, M., Langston, L. & Stillman, B., 2013. Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harbor perspectives in biology*, 5(7).

O'Driscoll, M., 2017. The pathological consequences of impaired genome integrity in humans; disorders of the DNA replication machinery. *The Journal of pathology*, 241(2), pp.192–207.

Okuno, Y. et al., 1999. Clustered adenine/thymine stretches are essential for function of a fission yeast replication origin. *Molecular and cellular biology*, 19(10), pp.6699–6709.

Olavarrieta, L. et al., 2002. Knotting dynamics during DNA replication. *Molecular microbiology*, 46(3), pp.699–707.

On, K.F. et al., 2014. Prereplicative complexes assembled in vitro support origin-dependent and independent DNA replication. *The EMBO journal*, 33(6), pp.605–620.

Ozer, O. & Hickson, I.D., 2018. Pathways for maintenance of telomeres and common fragile sites during DNA replication stress. *Open biology*, 8(4).

Parajuli, S. et al., 2017. Human ribonuclease H1 resolves R-loops and thereby enables progression of the DNA replication fork. *The Journal of biological chemistry*, 292(37), pp.15216–15224.

Peace, J.M. et al., 2016. Quantitative BrdU immunoprecipitation method demonstrates that Fkh1 and Fkh2 are rate-limiting activators of replication origins that reprogram replication timing in G1 phase. *Genome research*, 26(3), pp.365–375.

De Piccoli, G. et al., 2012. Replisome stability at defective DNA replication forks is

independent of S phase checkpoint kinases. *Molecular cell*, 45(5), pp.696–704.

Pizzi, S. et al., 2015. Reduction of hRNase H2 activity in Aicardi-Goutieres syndrome cells leads to replication stress and genome instability. *Human molecular genetics*, 24(3), pp.649–658.

Prado, F. & Aguilera, A., 2005. Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *The EMBO journal*, 24(6), pp.1267–1276.

Pursell, Z.F. et al., 2007. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science (New York, N.Y.)*, 317(5834), pp.127–130.

Qi, L.S. et al., 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152(5), pp.1173–1183.

Qiu, J. et al., 1999. Saccharomyces cerevisiae RNase H(35) functions in RNA primer removal during lagging-strand DNA synthesis, most efficiently in cooperation with Rad27 nuclease. *Molecular and cellular biology*, 19(12), pp.8361–8371.

Ran, F.A. et al., 2013. Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell*, 154(6), pp.1380–1389. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23992846 [Accessed July 10, 2014].

Ran, F.A. et al., 2013. Genome engineering using the CRISPR-Cas9 system. *Nature protocols*, 8(11), pp.2281–308. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3969860&tool=pmcentrez&rendertype=abstract [Accessed July 9, 2014].

Ran, F.A. et al., 2015. In vivo genome editing using Staphylococcus aureus Cas9. *Nature*. Available at: http://www.nature.com/doifinder/10.1038/nature14299 [Accessed April 1, 2015].

Randell, J.C.W. et al., 2006. Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. *Molecular cell*, 21(1), pp.29–39.

Randerath, K. et al., 1992. Formation of ribonucleotides in DNA modified by oxidative damage in vitro and in vivo. Characterization by 32P-postlabeling. *Mutation research*, 275(3–6), pp.355–366.

Rao, H., Marahrens, Y. & Stillman, B., 1994. Functional conservation of multiple elements in yeast chromosomal replicators. *Molecular and cellular biology*, 14(11), pp.7643–7651.

Reijns, M.A.M. et al., 2012. Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell*, 149(5), pp.1008–1022.

Reijns, M.A.M. et al., 2011. The structure of the human RNase H2 complex defines key interaction interfaces relevant to enzyme function and human disease. *The Journal of biological chemistry*, 286(12), pp.10530–10539.

Reijns, M.A.M. & Jackson, A.P., 2014. Ribonuclease H2 in health and disease. *Biochemical Society transactions*, 42(4), pp.717–725.

Remus, D. et al., 2009. Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell*, 139(4), pp.719–730.

Rice, G. et al., 2007. Clinical and molecular phenotype of Aicardi-Goutieres syndrome. *American journal of human genetics*, 81(4), pp.713–725.

Rice, G.I. et al., 2014. Gain-of-function mutations in IFIH1 cause a spectrum of human disease phenotypes associated with upregulated type I interferon signaling. *Nature genetics*, 46(5), pp.503–509.

Rice, G.I. et al., 2012. Mutations in ADAR1 cause Aicardi-Goutieres syndrome associated

with a type I interferon signature. *Nature genetics*, 44(11), pp.1243–1248.

Rice, G.I. et al., 2009. Mutations involved in Aicardi-Goutieres syndrome implicate SAMHD1 as regulator of the innate immune response. *Nature genetics*, 41(7), pp.829–832.

Richardson, F.C. et al., 2000. Polymerization of 2'-fluoro- and 2'-O-methyl-dNTPs by human DNA polymerase alpha, polymerase gamma, and primase. *Biochemical pharmacology*, 59(9), pp.1045–1052.

Riera, A., Fernandez-Cid, A. & Speck, C., 2013. The ORC/Cdc6/MCM2-7 complex, a new power player for regulated helicase loading. *Cell cycle (Georgetown, Tex.)*, 12(14), pp.2155–2156.

Rouet, P., Smih, F. & Jasin, M., 1994. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Molecular and cellular biology*, 14(12), pp.8096–8106.

Rychlik, M.P. et al., 2010. Crystal structures of RNase H2 in complex with nucleic acid reveal the mechanism of RNA-DNA junction recognition and cleavage. *Molecular cell*, 40(4), pp.658–670.

Rydberg, B. & Game, J., 2002. Excision of misincorporated ribonucleotides in DNA by RNase H (type 2) and FEN-1 in cell-free extracts. *Proceedings of the National Academy of Sciences of the United States of America*, 99(26), pp.16654–16659.

Samadashwily, G.M., Raca, G. & Mirkin, S.M., 1997. Trinucleotide repeats affect DNA replication in vivo. *Nature genetics*, 17(3), pp.298–304.

Schmitt, M.W. et al., 2010. Active site mutations in mammalian DNA polymerase delta alter accuracy and replication fork progression. *The Journal of biological chemistry*, 285(42), pp.32264–32272.

Schneider, C.A., Rasband, W.S. & Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature methods*, 9(7), pp.671–675.

Schuster-Böckler, B. & Lehner, B., 2012. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature*, 488, p.504. Available at: http://dx.doi.org/10.1038/nature11273.

Schutz, K., Hesselberth, J.R. & Fields, S., 2010. Capture and sequence analysis of RNAs with terminal 2',3'-cyclic phosphates. *RNA (New York, N.Y.)*, 16(3), pp.621–631.

Segurado, M., de Luis, A. & Antequera, F., 2003. Genome-wide distribution of DNA replication origins at A+T-rich islands in Schizosaccharomyces pombe. *EMBO reports*, 4(11), pp.1048–1053.

Sequeira-Mendes, J. et al., 2009. Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS genetics*, 5(4), p.e1000446.

Shaban, N.M. et al., 2010. The structure of the mammalian RNase H2 complex provides insight into RNA.NA hybrid processing to prevent immune dysfunction. *The Journal of biological chemistry*, 285(6), pp.3617–3624.

Shinbrot, E. et al., 2014. Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. *Genome research*, 24(11), pp.1740–1750.

Shirahige, K. et al., 1993. Location and characterization of autonomously replicating sequences from chromosome VI of Saccharomyces cerevisiae. *Molecular and cellular biology*, 13(8), pp.5043–5056.

Siddiqui, K. & Stillman, B., 2007. ATP-dependent assembly of the human origin recognition complex. *The Journal of biological chemistry*, 282(44), pp.32370–

32383.
Smithies, O. et al., 1985. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature*, 317(6034), pp.230–234.
Sparks, J.L. et al., 2012. RNase H2-initiated ribonucleotide excision repair. *Molecular cell*, 47(6), pp.980–986.
Sparks, J.L. & Burgers, P.M., 2015. Error-free and mutagenic processing of topoisomerase 1-provoked damage at genomic ribonucleotides. *The EMBO journal*, 34(9), pp.1259–1269.
Stinchcomb, D.T., Struhl, K. & Davis, R.W., 1979. Isolation and characterisation of a yeast chromosomal replicator. *Nature*, 282(5734), pp.39–43.
Stirling, P.C. et al., 2012. R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes & development*, 26(2), pp.163–175.
Stout, J.T. & Caskey, C.T., 1985. HPRT: gene structure, expression, and mutation. *Annual review of genetics*, 19, pp.127–148.
Sugimoto, N. et al., 2015. Cdt1-binding protein GRWD1 is a novel histone-binding protein that facilitates MCM loading through its influence on chromatin architecture. *Nucleic acids research*, 43(12), pp.5898–5911.
Sun, J. et al., 2012. Cdc6-induced conformational changes in ORC bound to origin DNA revealed by cryo-electron microscopy. *Structure (London, England : 1993)*, 20(3), pp.534–544.
Sun, J. et al., 2013. Cryo-EM structure of a helicase loading intermediate containing ORC-Cdc6-Cdt1-MCM2-7 bound to DNA. *Nature structural & molecular biology*, 20(8), pp.944–951.
Sun, J. et al., 2015. The architecture of a eukaryotic replisome. *Nature structural & molecular biology*, 22(12), pp.976–982.
Sun, L. et al., 2013. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science (New York, N.Y.)*, 339(6121), pp.786–791.
Talwani, R. et al., 2013. Simeprevir: a macrocyclic HCV protease inhibitor. *Drugs of today (Barcelona, Spain : 1998)*, 49(12), pp.769–779.
Tan, X. et al., 2007. Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature*, 446(7136), pp.640–645.
Tanaka, S. & Araki, H., 2013. Helicase activation and establishment of replication forks at chromosomal origins of replication. *Cold Spring Harbor perspectives in biology*, 5(12), p.a010371.
Tatsumi, Y. et al., 2003. The ORC1 cycle in human cells: I. cell cycle-regulated oscillation of human ORC1. *The Journal of biological chemistry*, 278(42), pp.41528–41534.
Teale, W.D., Paponov, I.A. & Palme, K., 2006. Auxin in action: signalling, transport and the control of plant growth and development. *Nature reviews. Molecular cell biology*, 7(11), pp.847–859.
Terns, M.P. & Terns, R.M., 2011. CRISPR-based adaptive immune systems. *Current opinion in microbiology*, 14(3), pp.321–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3119747&tool=pmcentrez&rendertype=abstract [Accessed April 8, 2015].
Thomas, K.R., Folger, K.R. & Capecchi, M.R., 1986. High frequency targeting of genes to specific sites in the mammalian genome. *Cell*, 44(3), pp.419–428.
Thorslund, T. et al., 2010. The breast cancer tumor suppressor BRCA2 promotes the specific targeting of RAD51 to single-stranded DNA. *Nature Publishing Group*,

17(10), pp.1263–1265. Available at: http://dx.doi.org/10.1038/nsmb.1905.

Toledo, L.I. et al., 2013. ATR prohibits replication catastrophe by preventing global exhaustion of RPA. *Cell*, 155(5), pp.1088–1103.

Traut, T.W., 1994. Physiological concentrations of purines and pyrimidines. *Molecular and cellular biochemistry*, 140(1), pp.1–22.

Tsai, S.Q. et al., 2014. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature biotechnology*, 32(6), pp.569–76. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24770325 [Accessed July 9, 2014].

Tsurimoto, T. & Stillman, B., 1990. Functions of replication factor C and proliferating-cell nuclear antigen: functional similarity of DNA polymerase accessory proteins from human cells and bacteriophage T4. *Proceedings of the National Academy of Sciences of the United States of America*, 87(3), pp.1023–1027.

Tubbs, A. & Nussenzweig, A., 2017. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell*, 168(4), pp.644–656.

Tuduri, S. et al., 2009. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nature cell biology*, 11(11), pp.1315–1324.

Ul Ain, Q., Chung, J.Y. & Kim, Y.-H., 2015. Current and future delivery systems for engineered nucleases: ZFN, TALEN and RGEN. *Journal of controlled release : official journal of the Controlled Release Society*, 205, pp.120–127.

Valton, A.-L. et al., 2014. G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *The EMBO journal*, 33(7), pp.732–746.

Venkatesan, R.N. et al., 2007. Mutation at the polymerase active site of mouse DNA polymerase delta increases genomic instability and accelerates tumorigenesis. *Molecular and cellular biology*, 27(21), pp.7669–7682.

Voineagu, I. et al., 2008. Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 105(29), pp.9936–9941.

White, R. et al., 2013. Identification of small-molecule inhibitors of the ribonuclease H2 enzyme. *Journal of biomolecular screening*, 18(5), pp.610–620.

Wiedenheft, B., Sternberg, S.H. & Doudna, J. a., 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature*, 482(7385), pp.331–338. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22337052 [Accessed July 9, 2014].

Williams, J.S. et al., 2012. Proofreading of ribonucleotides inserted into DNA by yeast DNA polymerase varepsilon. *DNA repair*, 11(8), pp.649–656.

Williams, J.S. & Kunkel, T.A., 2014. Ribonucleotides in DNA: origins, repair and consequences. *DNA repair*, 19, pp.27–37.

Williams, J.S., Lujan, S.A. & Kunkel, T.A., 2016. Processing ribonucleotides incorporated during eukaryotic DNA replication. *Nature reviews. Molecular cell biology*, 17(6), pp.350–363.

Woodward, A.M. et al., 2006. Excess Mcm2-7 license dormant origins of replication that can be used under conditions of replicative stress. *The Journal of cell biology*, 173(5), pp.673–683.

Yardimci, H. & Walter, J.C., 2014. Prereplication-complex formation: a molecular double take? *Nature structural & molecular biology*, 21(1), pp.20–25.

Yeeles, J.T.P. et al., 2013. Rescuing stalled or damaged replication forks. *Cold Spring Harbor perspectives in biology*, 5(5), p.a012815.

Zahn, K.E., Wallace, S.S. & Doublie, S., 2011. DNA polymerases provide a canon of strategies for translesion synthesis past oxidatively generated lesions. *Current opinion in structural biology*, 21(3), pp.358–369.

Zeman, M.K. & Cimprich, K.A., 2014. Causes and consequences of replication stress. *Nature cell biology*, 16(1), pp.2–9.

Zerbino, D.R. et al., 2018. Ensembl 2018. *Nucleic acids research*, 46(D1), pp.D754–D761.

Zhang, H. & Freudenreich, C.H., 2007. An AT-rich sequence in human common fragile site FRA16D causes fork stalling and chromosome breakage in S. cerevisiae. *Molecular cell*, 27(3), pp.367–379.

Zhu, W. et al., 2007. Mcm10 and And-1/CTF4 recruit DNA polymerase alpha to chromatin for initiation of DNA replication. *Genes & development*, 21(18), pp.2288–2299.

Zou, L. & Elledge, S.J., 2003. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science (New York, N.Y.)*, 300(5625), pp.1542–1548.

Zou, X. et al., 2017. Short inverted repeats contribute to localized mutability in human somatic cells. *Nucleic acids research*, 45(19), pp.11213–11221.