Understanding the Cellular Consequences of Meier-Gorlin Syndrome

A thesis

submitted in partial fulfilment

of the requirements for the Degree of

Master of Science

at

University of Otago

by

Rosemary Sullivan

University of Otago

March 2020

Abstract

Meier-Gorlin syndrome (MGS) is a rare autosomal recessive disorder characterised by short stature, microtia and patella aplasia/hypoplasia. Genetic variants, which cause MGS, have previously been found in genes involved in the initiation of DNA replication. This study examined the cellular consequences of novel variants in MGS genes; *ORC1*, *CDC45* and *DONSON*.

The *ORC1* gene encodes an essential component of the pre-replication complex and functions during late mitosis/early G1 phase to initiate DNA replication. MGS individuals previously reported to have variants in *ORC1* have had at least one variant in the bromo-adjacent homology (BAH) domain at the N-terminus of the protein, a region suggested to be important for protein-protein interactions. In this study we report a patient with a novel homozygous variant (c.1865T>C, p.L622P) in the ATPase Associated with a wide range of cellular Activities (AAA) domain at the C-terminus of ORC1, and attempted to investigate how variants in this region of the protein lead to a MGS phenotype using CRISPR-Cas9 genome editing and a minigene splicing assay. Due to time constraints, results were inconclusive and more work is required to better understand variants in this area of *ORC1*.

CDC45 encodes an essential component of the pre-initiation and CDC45-MCM-GINS (CMG) complex, required during G1 and S-phase. The novel homozygous *CDC45* variant (c. 1441-2 A>G) under investigation in this study was found in two siblings presenting with a severe MGS phenotype alongside a range of secondary phenotypes some of which are not typically associated with MGS. It was hypothesised that this novel variant may represent the extreme end of the CDC45 phenotypic spectrum. A splicing assay showed that the variant, located within the canonical splice acceptor site for *CDC45* exon 16, caused aberrant splicing and use of an alternative 5′ splice acceptor within exon 16. This resulted in a two amino acid deletion (p.Thr481_Lys482del) and a 72.9% reduction in *CDC45* mRNA levels, which was confirmed by RT-qPCR analysis using patient fibroblasts.

DONSON is required for stabilising replication forks during S-phase when replication stress is encountered. Biallelic variants in this gene have previously been described in patients presenting with Microcephaly-Micromelia syndrome (MIMIS) and Microcephaly Short Stature, and Limb Abnormalities (MISSLA), both of which are characterized by severe microcephaly and a slight reduction in height. In this study we discovered a novel

DONSON variant (c.631C>T, p.R211C), for the first time in clinically diagnosed MGS patients, presenting with a global reduction in size. This project used CRISPR-Cas9 genome editing as well as a range of cellular techniques including subcellular localisation, immunocytochemistry and DNA fibre combing to better understand the cellular consequences of this novel MGS gene. Taken together, results confirmed *DONSON* as a novel MGS gene and showed that MGS variants led to more subtle changes in subcellular localisation, DNA damage, and replication events than variants seen in non-MGS *DONSON* patients, reflecting the difference in phenotype.

This project aimed to study the cellular consequences of novel MGS variants using CRISPR-Cas9 and a variety of cellular techniques to understand the effect these novel variants had on initiation of DNA replication.

Acknowledgements

First and foremost I would like to thank my supervisor Dr Louise Bicknell for your ongoing support, advice, encouragement and always being ready to help with a smile. I would also like to extend thanks to Dr Karen Knapp for your constant support in the lab and somehow always knowing how to troubleshoot my next failed experiment. As well as everyone else in the Bicknell lab, I have enjoyed every minute of working with you guys (even the minutes I hated!).

Thanks to everyone in the RoBiMark group for your continued support and advice. As well as the pathology lunch group, thanks for my daily work distraction, support and advice. I would also like to say a big thank you to anyone who sat down and wrote with me, there's no way I could have done this without you guys!

I would like to thank my fellow pathology postgrads and 400 level genetics class for your support, laughs and belief in me. Shout out to my master's buddy, Sean Driver who was there with me through all the late night 4th year lab and 5th year writing sessions. I would also like to thank Josh Clark for being the best undergraduate lab partner and friend a girl could have.

A huge thank you to office 113 for welcoming me into the department and allowing me to come back and finish writing my thesis in the same place I started. A special shout out Amarni Thomas for supporting me from day one (2nd year) right through to writing with me in the last few weeks. You're the best!

I would like to thank my mum for your continued love and support, my dad for trying his best to proof read my work even when he had no idea what he was reading, my dog Ginger for your unconditional love and adorableness, and my sister who decided to distract me in the last few days of thesis writing by making me drive to Palmerston at 5am so you could have a baby (oh yeah it's in writing now, I'm never gonna let you forget that one).

Last but definitely not least I would like to thank my partner Henry Golden Arnold for somehow tolerating me through this whole process, ya fucking loser.

List of Symbols and Abbreviations

AAA	ATPase Associated with a wide range of cellular Activities
APC	Anaphase-promoting complex
ATLD2	Ataxia-telangiectasia-like disorder-2
ATM	Ataxia telangiectasia mutated
ATP	Adenosine triphosphate
ATPase	Adenylpyrophosphatase
ATR	ATM and Rad3 related
ATR–ATRIP	ATR-interacting protein
ATR–CHK1	ATR-Checkpoint kinase 1
BAH	Bromo-adjacent homology domain
Bp	Base pair
BrdU	5-bromo-2'-deoxyuridine
BSA	Bovine Serum Albumin
CDC	Cell Division Cycle
CDC45	Cell Division Cycle 45
CDC6	Cell division control protein 6
CDK	Cyclin dependent kinase
cDNA	Complementary DNA
CDT1	Chromatin licensing and DNA replication factor 1
CENPJ	Centromere Protein J
CEP152	Centrosomal Protein 152

CEP63	Centrosomal Protein 63
CldU	5-chloro-2'-deoxyuridine
CMG	CDC45-MCM-GINS
CRISPR	Clustered regularly interspaced short palindromic repeats
DAPI	4′,6-diamidino-2-phenylindole
DDK	cdc7-dbf4 kinase
DMEM	Dulbecco's Modified Eagle Medium
DMEM/F12	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12
DMSO	Dimethly sulfoxide
DNA	Deoxyribonucleic Acid
DNA2	DNA Replication Helicase/Nuclease 2
DNA-PKcs	DNA-dependent protein kinase
dNTPS	Deoxyribonucleotide triphosphate
DONSON	Downstream neighbour of SON
FACS	Fluorescence-activated cell sorting
FBS	Fetal bovine serum
FILS	Facial dysmorphism-immunodeficiency-livedo-short stature
g	Centrifuge speed (times gravity)
G1	Gap 1
G2	Gap 2
gDNA	Genomic DNA
GFP	Green fluorescent protein

GMNN	Geminin
hTERT	Human telomerase reverse transcriptase
IdU	5-Iodo-2′-deoxyuridine
IMAGEI	Intrauterine growth retardation, metaphyseal dysplasia, adrenal hypoplasia congenita, genital anomalies, and immunodeficiency
IMD54	Immunodeficiency 54
IMD55	Immunodeficiency 55
LB	Luria Broth
MCM	Minichromosome Maintenance
MGS	Meier-Gorlin syndrome
MIMIS	Microcephaly-micromelia syndrome
MISSLA	Microcephaly and Short Stature, with Limb Anomalies
MOPD	Microcephalic Osteodysplastic Primordial Dwarfism
M-phase	Mitosis phase
mRNA	Messenger Ribonucleic Acid
mSA	Monomeric streptavidin
mV	Millivolt
NGS	Normal Goat Serum
NIN	Ninein
NSMCE2	E3 SUMO-protein transferase NSE2Curated
OFC	Occipitofrontal Circumference
ORC	Origin Recognition complex

ORC1	Origin Recognition Complex Subunit 1
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase Chain Reaction
PenStrep	Penicillin streptomycin
Pol α	Polymerase alpha
Pol δ	Polymerase delta
Pol ε	Polymerase epsilon
PreIC	Pre-initiation complex
PreLC	Pre-loading complex
PreRC	Pre-replication complex
Primer	Oligonucleotide
RT-qPCR	Quantitative reverse transcriptase PCR
RAPADILINO	Radial hypoplasia, patella hypoplasia and cleft or arched palate, diarrhoea and dislocated joints, little size and limb malformation, slender nose and normal intelligence
RBBP8RB	Binding Protein 8, Endonuclease
RECQL4	ATP-dependent DNA helicase Q4
RNA	Ribonucleic Acid
RNU4ATAC	RNA, U4atac Small Nuclear (U12-Dependent Splicing
RPA proteins	Replication protein A
RT-PCR	Reverse transcriptase PCR
SD	Standard deviation
sgRNA	Single guide RNA

S-phase	Synthesis phase
TopBP1	DNA topoisomerase 2-binding protein 1
TRAIP	TRAF Interacting Protein
Treslin	TopBP1-interacting, replication-stimulating protein
UV	Ultra violet
WT	Wild Type

Table of Contents

List of Tables

List of Figures

*All figures were made by me unless stated otherwise.

1 Introduction

This project aimed to investigate three novel Meier-Gorlin syndrome (MGS) variants using CRISPR-Cas9 and a variety of cellular assays. MGS is a type of primordial dwarfism caused by variants in genes important in the initiation of DNA replication, a highly regulated process. This project describes an investigation of three novel variants in three MGS genes (*ORC1, CDC45* and *DONSON*) whose protein products work at different stages of DNA replication. The *ORC1* gene encodes an essential subunit of the origin recognition complex (ORC), a component of the pre-replication complex (preRC), required for the initiation of DNA replication during the late M/early G1 phase of the cell cycle. CDC45 is required for firing of the pre-initiation complex (preIC) during the G1/S interphase of the cell cycle and *DONSON* is a novel MGS gene known to be involved in replication fork stability during S-phase.

1.1 DNA replication and cell cycle progression

The cell cycle is an important, highly regulated process. During the cell cycle, a cell must copy its genetic material and grow in size before dividing into two identical daughter cells. DNA replication is tightly regulated in eukaryotes and is only undertaken once per cell cycle in healthy cells (Benmerzouga et al., 2012; Kara et al., 2015). There are four phases in the cell cycle, the first is gap 1 (G1) phase. This is the longest phase where the cell grows and DNA replication is initiated, next is S phase where the DNA is copied; this is followed by gap 2 (G2) phase where the cell grows in size in preparation for the final phase mitosis (M-phase); where the cell divides. The process by which DNA is copied during the S phase is complex. DNA replication starts at origins of replication; in larger genomes there are many of these sites across the genome in order to copy the DNA quickly (Remus et al., 2009). Licencing proteins such as those that make up the preRC are recruited to the origins of replication in the late M phase/early G1 phase, before the initiation of replication in the S phase by the preIC (Klein and Gilbert, 2016). The process of DNA replication licencing is conserved in Eukaryotes, Bacteria and Archaea (Dueber et al., 2011, 2007; O'Donnell et al., 2013; de la Paz Sanchez and Gutierrez, 2009). In prokaryotes, there are few origins of replication due to the small genome size. The human genome has over 6,000,000,000 base pairs (bp) which must be copied correctly in a relatively short

amount of time in order for the cell to progress through the cell cycle and divide, therefore many origin of replication sites (~50,000) are required (Gambus, 2017).

1.1.1 Pre-replication complex

The eukaryotic preRC is required for the initiation of DNA replication and is highly conserved from humans through to *Saccharomyces cerevisiae* (Bell, 2002; Bleichert et al., 2013; Dueber et al., 2011; Stillman, 2001). The preRC contains origin recognition complex (ORC) subunits 1-6, cell division control protein 6 (CDC6), chromatin licensing and DNA replication factor 1 (CDT1) and minichromosome maintenance DNA helicase (MCM) subunits 2-7. ORC1-6 subunits encircle the chromatin at origin of replication start sites to form the hexameric ORC complex. After the ORC has been loaded onto the chromatin, CDC6 is recruited to the complex and loaded with the help of ORC1 (Wang et al., 1999). One MCM2-7 complex is loaded with CDT1 and bound by the ORC and CDC6 complex. The second MCM2-7 complex is then loaded along with the help of CDC6 and CDT1 (Ticau et al., 2017). Together these form the pre-replication complex (Figure 1.1).

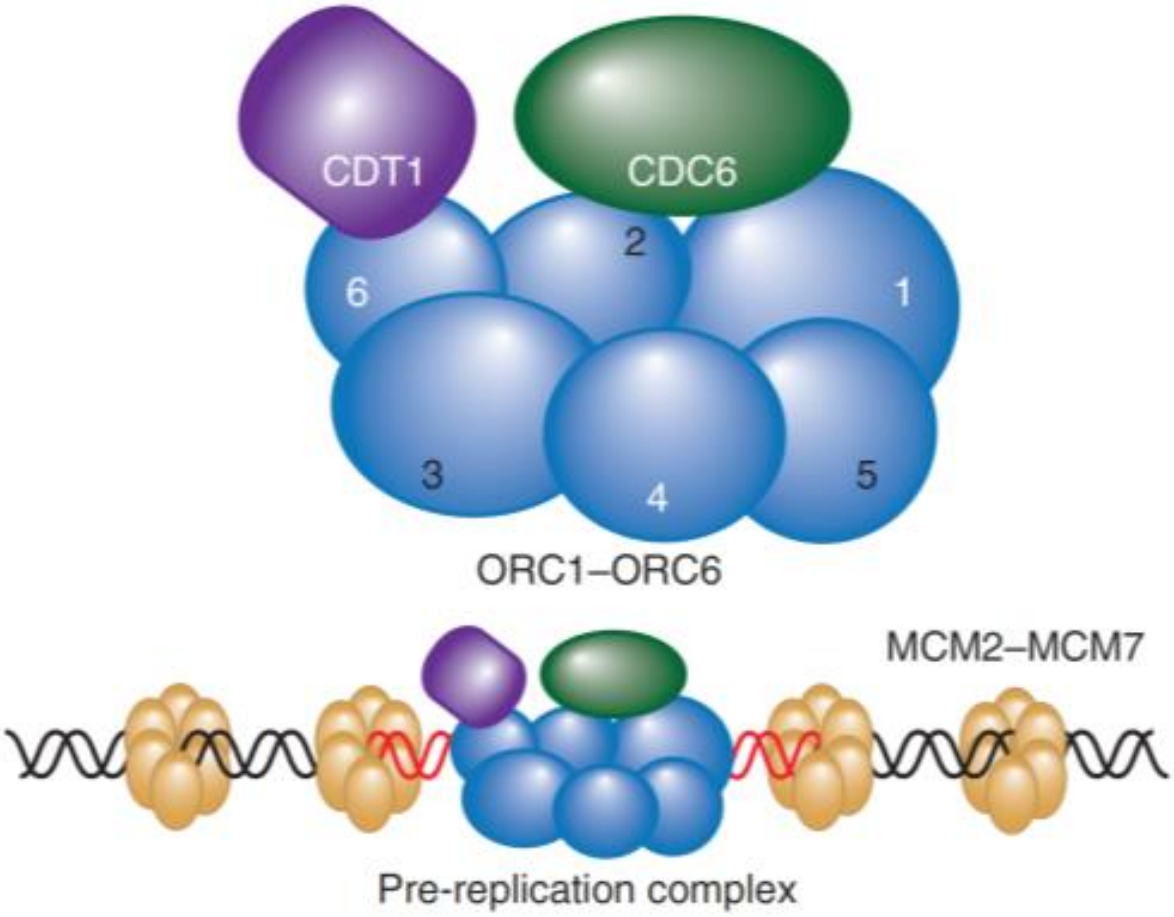


Figure 1.1 **Pre-replication complex assembly.** The six ORC subunits encircle and bind to chromatin at origin of replication sites. CDC6 is loaded to the ORC complex with the help of ORC1; CDT1 is loaded next to the complex with the help of ORC6. CDT1 associates with the C-terminal of CDC6 to load the two MCM 2-7 helicases, creating the pre-replication complex. Figure from (Bicknell et al., 2011b), copyright obtained from Springer Nature - 4850520608800.

1.1.1.1 ORC complex

The ORC complex is the first to be recruited to the origin of replication sites, at the initiation of DNA replication. Six subunits make up the ORC complex (Figure 1.1). ORC1 is the biggest subunit and ORC6 is the smallest subunit. In yeast ORC is permanently associated with the chromatin (Kolesnikova, 2013), while in other eukaryotes the ORC complex is recruited to the origin of replication sites in late mitosis/early G1 phase (Mendoza-Maldonado et al., 2010). This is the start of DNA replication (Bell and Stillman, 1992; Bicknell et al., 2011a). The ORC complex uses ATPase activity to bind to the DNA independently of other proteins (Bell, 2002; Bell and Stillman, 1992; Méchali, 2010). The six ORC subunits assemble around the DNA encircling it and binding to each other.

1.1.1.2 Origins of Replication

Origins of replication in metazoan are not chosen based on sequence but on epigenetic feature and sequence content (Aladjem, 2007). Previous studies have found that origins are enriched at transcription start sites, CpG islands and G-rich motifs (Cayrou et al., 2011; Delgado et al., 1998). It is thought that G-rich motifs may be enriched for as the asymmetry in the G/C, A/T content causes distortion in the double helix leading to the formation of non-B DNA. Non-B DNA is more likely to unwind than B-form DNA, allowing the DNA to unwind during S-phase (Arnott et al., 1983; Bartholdy et al., 2015). Kuo et al., 2012 found a direct link between chromatin modifications and replication licensing. It was found that the bromo-adjacent homology (BAH) domain of human ORC1 binds to H4K20me2 modifications and therefore may act as a selector of origin sites in humans. This complex is extremely important in DNA replication as any problems or delay with the first step of DNA replication will likely lead to a delay in DNA replication and cell cycle progression (discussed further in section 1.2.3).

1.1.1.3 CDC6, CDT1 and GMNN interactions

CDC6 is an essential pre-replication complex protein, which is only active in G1 phase. It is rapidly degraded during S-phase in budding and fission yeast, but in metazoan it's activity is down-regulated by being exported from the nucleus (Kim and Kipreos, 2007; Kim et al., 2007; Musiałek and Rybaczek, 2015; Piatti et al., 1995). CDC6 is loaded with the help of ORC1 and is important in maintaining pre-replication complex formation (Chen et al., 2008; Saha et al., 1998; Wang et al., 1999). Like ORC1, CDC6 is an (ATPase Associated with a wide range of cellular Activities) AAA+ ATPase.

CDT1 plays an important role in replication; it is loaded to the pre-replication complex with the help of ORC6. CDT1 then associates with the C-terminus of CDC6 to load the MCM2-7 helicase (Fernández-Cid et al., 2013; Khayrutdinov et al., 2009; Liu et al., 2012; Takara and Bell, 2011). CDT1 also plays a role in mitosis localizing to kinetochores through interaction with Hec1, although its exact function is not known (Pozo and Cook, 2016; Varma et al., 2012). CDT1's vital role in origin licencing makes it a target for DNA replication regulation, as without regulation CDT1 can load the MCM2-7 complex in G2-phase leading to inappropriate replication licensing and DNA damage (Blow and Dutta, 2005; Truong and Wu, 2011). CDT1 is regulated by both degradation and inactivation.

CDT1 levels are high in G1 phase where it is involved in MCM loading, significantly lower during S phase where it is actively degraded to prevent re-replication and increased again in G2 and M phase to prepare for DNA replication initiation in G1 phase. In humans CDT1 is degraded in S phase by two ubiquitin-mediated E3 ligases, CUL4-DDB1^{CDT2} and SCF^{Skp2}. (Kim and Kipreos, 2007; Lin et al., 2010; Machida et al., 2005; Sansam et al., 2006).

The second mechanism by which CDT1 is regulated is inactivation by Geminin (GMNN). GMNN is an inhibitor of DNA replication, active during S, G2 and M phase to prevent the incorporation of MCM to the pre-replication complex (Burrage et al., 2015; Yoshida et al., 2004). GMNN is quickly degraded during G1 phase to allow CDT1 to load MCM2-7 onto the chromatin (Figure 1.2). It is thought that the majority of CDT1 is degraded during S-phase and any low levels remaining are inactivated by GMNN. However CDT1 must accumulate in G2 phase in order to be at the high level required for DNA replication initiation in the G1 phase. Therefore GMNN is thought to play a role in increasing and inhibiting CDT1 during G2 phase in preparation for M and G1 phase (Ballabeni et al., 2013).

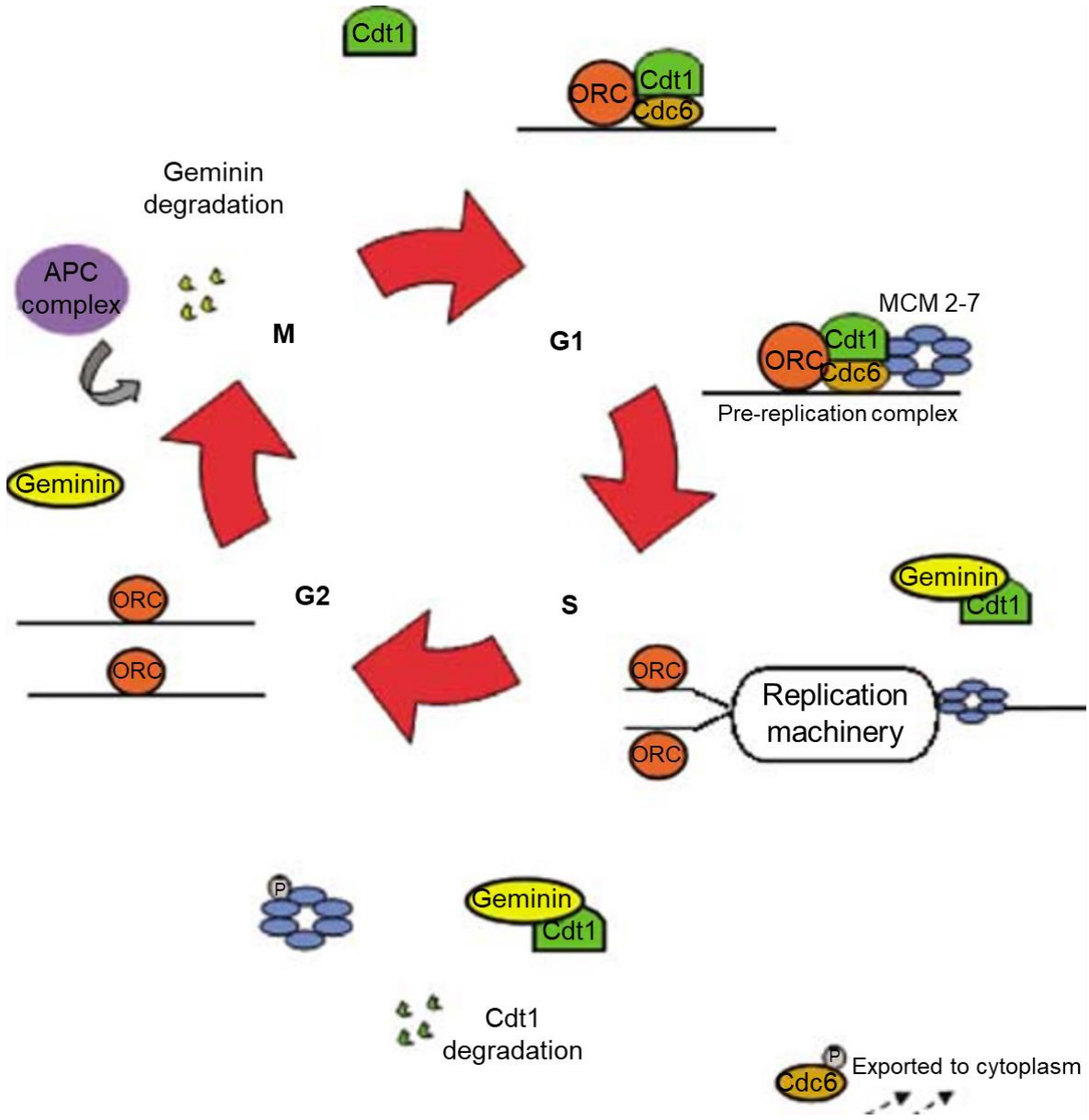


Figure 1.2 **Cdt1, Cdc6 and GMNN role and interaction during cell cycle in yeast.** Cdt1 and Cdc6 are required during G1 phase to load MCM2-7 complex. GMNN binds any Cdt1, which had not been degraded during S phase to prevent re-replication. Cdc6 is exported to the cytoplasm until the next G1 phase. GMNN levels increases during G2/M phase to bind Cdt1 in order for it to act quickly in G1 phase. GMNN is degraded at the end of M phase to allow Cdt1 to work quickly in G1 phase. Figure from (Montanari et al., 2006), copyright obtained from Springer Nature 4850520916840.

1.1.1.4 Minichromosome maintenance (MCM) 2-7 complex

The MCM complex comprises six subunits (MCM2-7), all of which are part of the AAA+ superfamily (Bell and Labib, 2016). These six subunits are important for ATP hydrolysis, which is required for conformational changes leading to the unwinding of the DNA duplex (Costa et al., 2011; Ilves et al., 2010). The MCM complex functions in S-phase as a DNA helicase, where it plays two important roles. First, it works in the initiation of DNA

replication and second it works to elongate DNA during S-phase. MCM2-7 acts by unwinding DNA ahead of the replication fork (Musiałek and Rybaczek, 2015). DNA helicase activity *in vitro* only requires three out of the six subunits (MCM4, 6 and 7), but all six are required for DNA replication (You et al., 1999). MCM2-7 exists as a single hexamer in solution but forms a double hexamer when loaded onto DNA (Evrin et al., 2009; Gambus et al., 2011). The double hexamer formation is important as this allows the MCM complex to move and unwind DNA in both directions from the origin of replication. MCM2-7 is loaded by ORC-CDC6 and chaperoned by CDT1 in an inactive state (Hua and Orr-Weaver, 2017; Parker et al., 2017). The MCM complex does not have the required helicase activity to initiate DNA replication alone and requires other proteins and kinases, which form the pre-initiation complex (Parker et al., 2017). This helps to regulate DNA replication so it occurs only once during S-phase.

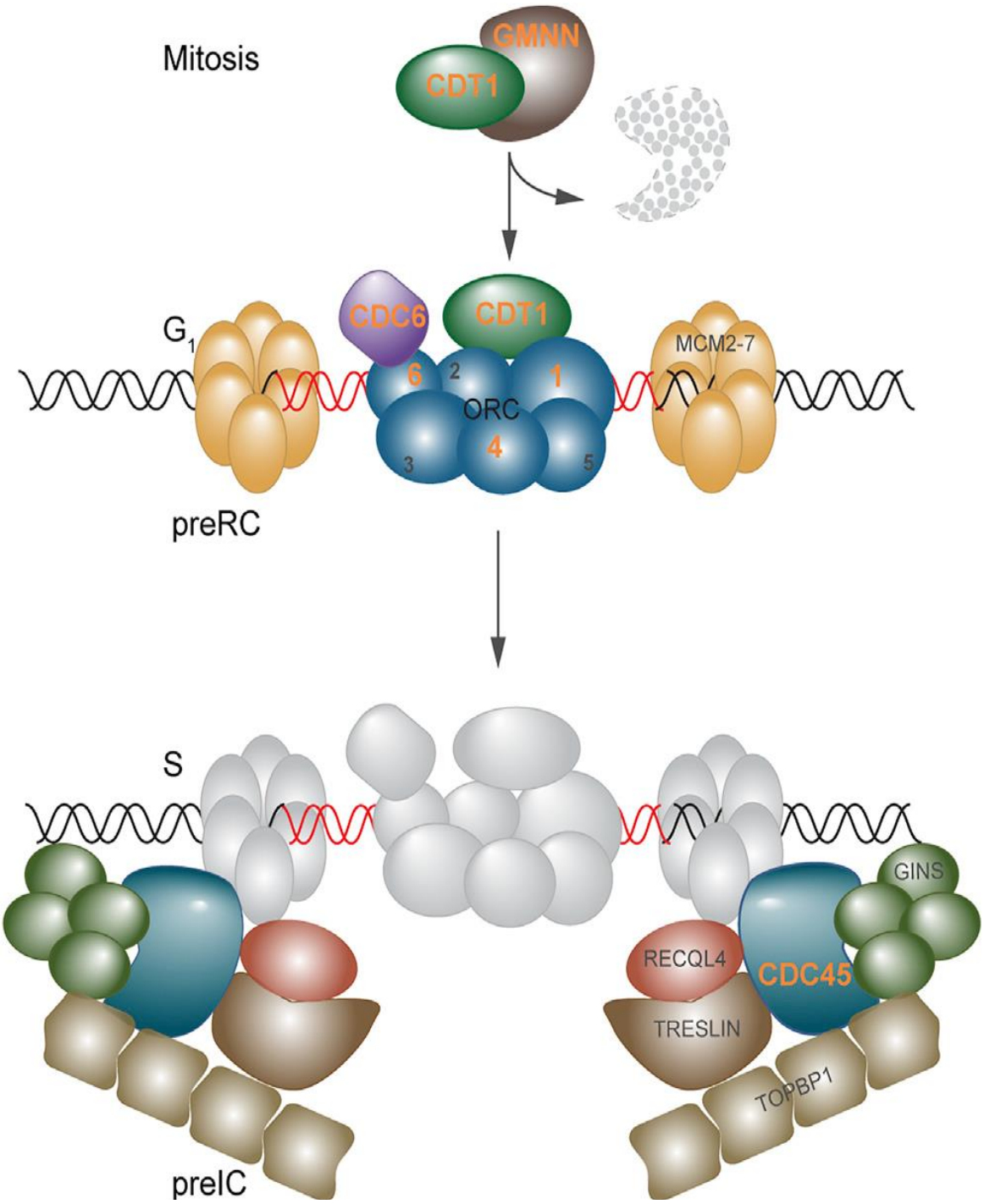


Figure 1.3 **Composition of the pre-initiation complex in humans.** The ORC complex is loaded onto the chromatin at the origins of replication, CDC6, CDT1 and MCM 2-7 are recruited to the complex. The pre-replication complex becomes the pre-initiation complex when CDC45 along with the GINS complex and many other proteins are recruited. All known MGS genes play a role in the initiation of DNA replication. Published MGS genes are written in orange. Figure from (Fenwick et al., 2016), copyright obtained from Elsevier - 4850521104495.

1.1.2 Pre-Initiation complex

The pre-initiation complex forms at the G1/S phase transition. CDC45 and the GINS complex (GINS subunits 1-4) are among the many proteins which bind to the pre-replication complex to form the pre-initiation complex (Miyazawa-Onami et al., 2017). This is regulated by S-phase Cyclin-Dependent Kinase (S-CDK) and Dbf4-Dependent Kinase (DDK) during the transition from G1 to S-phase (Musiałek and Rybaczek, 2015; Uzcanga et al., 2017). The addition of CDC45 and the GINS complex leads to the assembly of the CDC45-MCM2–7-GINS (CMG) complex, an inactive functional replicative helicase (Bleichert et al., 2017; Hua and Orr-Weaver, 2017; Parker et al., 2017). MCM10 is required for activation of this complex (Ricke and Bielinsky, 2004). The MCM2-7 double hexamer becomes separated into single hexamers, where both MCM2-7 complexes form an active CMG helicase complex, which unwind DNA in opposite directions initiating at the origin of replication (Figure 1.3).

1.1.2.1 Replication fork initiation

DDK phosphorylates the N-terminus of several MCM2-7 subunits (specifically MCM2, MCM4 and MCM6) (Francis et al., 2009; Sheu and Stillman, 2006), thereby creating a binding site for Treslin allowing the subsequent recruitment of CDC45 (Bell and Labib, 2016) (Figure 1.4). Next, the S-CDK phosphorylates Treslin and RecQL4, allowing the recruitment of TopBP1 to the phosphorylated RecQL4 protein (Tanaka et al., 2007; Yeeles et al., 2015; Zegerman and Diffley, 2007). This phosphorylation event also leads to the recruitment of Pol ε and the GINS complex (Figure 1.4). RecQL4-TopBP1-GINS-Pol ε complex is call the pre-loading complex (preLC) (Muramatsu et al., 2010). The pre-loading complex is recruited to the MCM helicase by interaction with TopBP1 and Treslin (Tanaka et al., 2007; Zegerman and Diffley, 2007).

RecQL4, Treslin and TopBP1 disassociate from the complex leaving just the CMG complex remaining (CDC45/MCM2-7/GINS). Although the CMG is ultimately the active helicase it does not have DNA unwinding abilities until it is activated by MCM10 (Figure 1.4) (van Deursen et al., 2012; Quan et al., 2015; Ricke and Bielinsky, 2004; Watase et al., 2012).

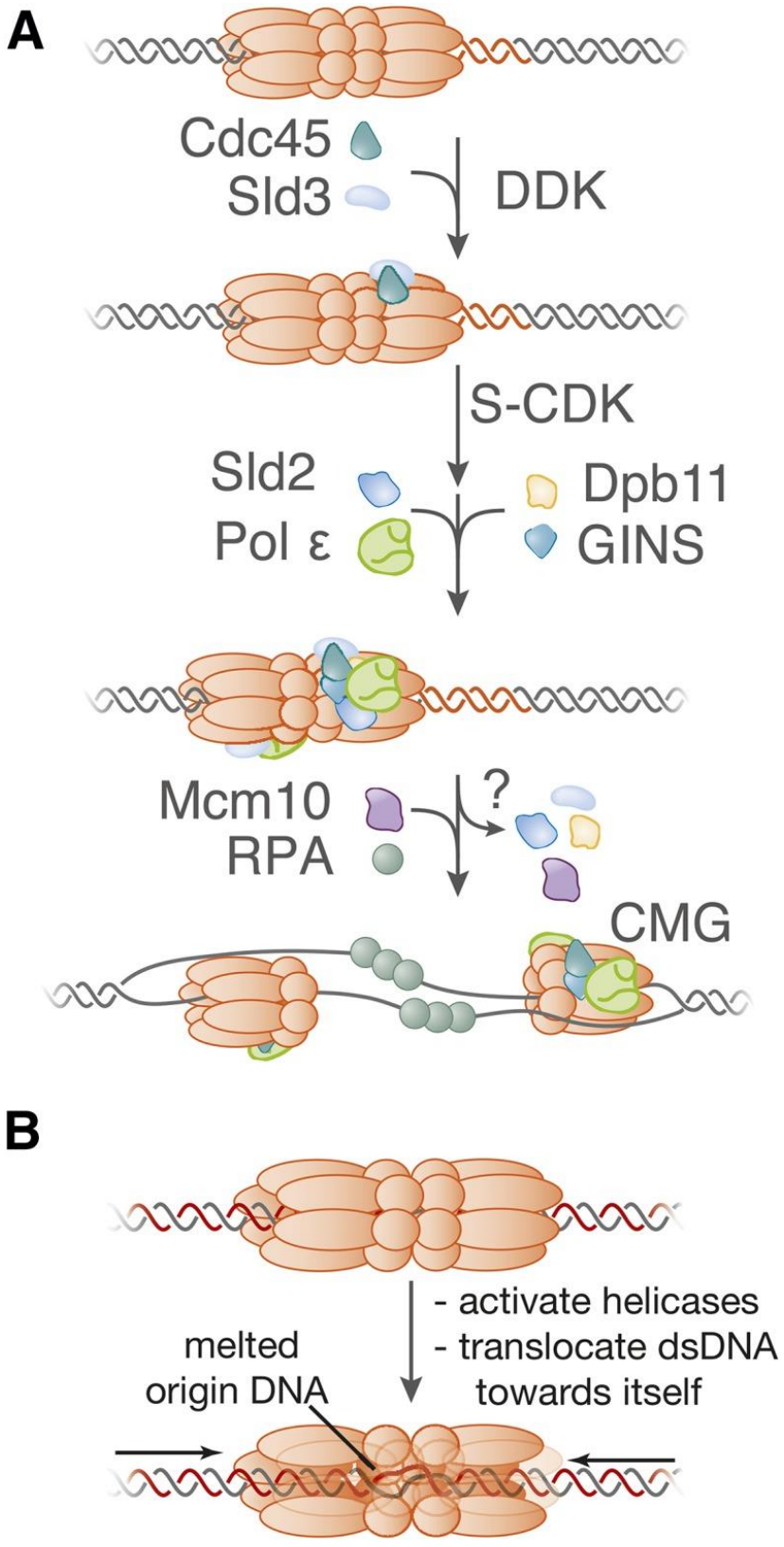


Figure 1.4 **Detailed assembly of the preIC and an active replication fork.** A) Depicts order and method by which each protein is acquired to preIC complex. Proteins in the figure represent the yeast system, human orthologues are in brackets. Sld2 (RecQL4), Sld3 (Treslin) and Dpb11 (TopBP1) are not part of the final replisome; however when they are released is unknown. B) Model mechanism for origin DNA melting and MCM2-7 non-translocating strand exportation. Figure from (Bell and Labib, 2016) reuse allowed under the Creative Commons Attribution 4.0 International License.

Once the CMG complex has been activated by MCM10, RPA proteins bind the single-stranded DNA to prevent re annealing (Figure 1.4A), the origin DNA is melted away to allow access by replisome components (Figure 1.4B) (Bell and Labib, 2016). As the MCM complexes originally bind double-stranded DNA the non-translocating strand must be exported from the MCM complex, via a conformational change of the CMG complex, as each of the MCM2-7 hexamers initiates DNA replication on single-stranded DNA in opposite directions (bidirectional replication fork) (Bell and Labib, 2016; Riera and Speck, 2015).

Next, Polymerase (Pol) α and primase are recruited by interactions with Ctf4 and the GINS complex (Gambus et al., 2009; Tanaka et al., 2009; Zhu et al., 2007). The primase synthesizes the first RNA primers to support initiation of elongation by Pol α (Bell and Labib, 2016). Pol δ or Pol ε undertake the majority of DNA replication, depending on whether DNA synthesis is occurring on the leading or lagging strand (Kunkel and Burgers, 2014; Pellegrini, 2012). Once a primer is placed on the leading strand, extension is continuous in the 5′ to 3′ direction by Pol ε (Kunkel and Burgers, 2014). Multiple primers must be placed on the lagging strand and small Okazaki fragments are elongated by pol δ (Kunkel and Burgers, 2014; Ogawa and Okazaki, 1980). When Pol δ reaches the 5′ end of the preceding Okazaki fragment it can continue synthesis by displacing part of the primer leaving a 5′ flap, before the two DNA fragments are joined by DNA ligase I (Garg et al., 2004).

1.1.2.2 CDC45

CDC45 is an essential DNA replication gene as it is required for formation of the CMG complex along with the GINS complex when it binds to the origin bound MCM2-7 complex (Fenwick et al., 2016; Hardy, 1997; Zou and Stillman, 2000; Zou et al., 1997). CDC45 is required for elongation during S-phase and during replication fork stalling events (Karnani and Dutta, 2011; Moyer et al., 2006; Pacek and Walter, 2004). CDC45 binds the leading strand of DNA at the replication fork, facilitating strand displacement (Fenwick et al., 2016; Parker et al., 2017). CDC45 is a limiting factor in origin firing based on low abundance, thus determining the number of origins of replication that can be active at any one time (Köhler et al., 2016). Compared to other proteins in the preRC there is only ~0.35 total molecules of CDC45, based on a preRC containing one ORC1-6 hexamer, four-five

MCM2-7 hexamers and two CDC6 proteins spaced every 100 kb (Wong et al., 2011). Due to the very low levels of CDC45, later origins cannot initiate until early replicons terminate and release CDC45 to be recycled (Musiałek and Rybaczek, 2015).

1.1.3 Replication Stress

During DNA replication a proportion of active replication forks may stall; this is called replication stress, and can lead to single or double-stranded breaks and chromosomal instability (Fragkos and Naim, 2017). As the replication fork moves along unwinding and copying DNA, there are many endogenous and exogenous stressors, which may cause the replication fork to slow down or stall. These stressors can include hard to copy areas of DNA such as highly repetitive regions, DNA lesions including nicks, gaps or long stretches of single-stranded DNA, DNA secondary structures, limited nucleotides or highly compact chromatin (Zeman and Cimprich, 2014).

When a replication fork stalls, there are different stress pathways and checkpoints activated depending on the level of stress. Normally during fork arrest the ATM and Rad3 related (ATR)-interacting protein (ATR–ATRIP) complex would be activated, and would subsequently phosphorylate and activate proteins required to inhibit cell cycle progression and promote DNA repair (Fragkos and Naim, 2017; Kemp et al., 2010; Sancar et al., 2004). Stalled replication forks must be stabilised to prevent collapse and recently DONSON was found to be involved in stabilising these replication forks (Reynolds et al., 2017).

1.1.3.1 DONSON

Downstream neighbour of SON (DONSON) was first described in 2000 by Wynn et al., but only had its role in replication fork stability described in 2017 first by Reynolds et al., 2017 in Nature Genetics, then confirmed by Evrony et al., 2017 in Genome Research. *DONSON* is a highly conserved gene with orthologues conserved as distantly as *Caenorhabditis elegans* (Bandura et al., 2005; Richardson, 2006). Genetic knockouts of the *DONSON* gene have been found to be embryonic lethal in both mouse and fly (*Drosophila melanogaster)* model systems tested (Dickinson et al., 2016; Lesly et al., 2017; Reynolds

et al., 2017), thus suggesting that *DONSON* is an essential DNA replication gene. The *Drosophila* orthologue of DONSON (humpty dumpty) has been shown to be involved in both DNA replication and cell cycle proliferation during development (Bandura et al., 2005). Both *humpty dumpty* and *DONSON* expression peaks during G1/S phase, suggesting they are involved in DNA replication and S-phase progression (Bandura et al., 2005; Evrony et al., 2017; Fuchs et al., 2010). DONSON responds to replication stress through ATR signalling and is thought to help stabilise the replication fork during replication (Reynolds et al., 2017). In the absence of DONSON, stalled replication forks are cleaved resulting in double-stranded breaks, chromosome instability and cell death (Reynolds et al., 2017) (Figure 1.5). While Reynolds et al., 2017 also proposed that DONSON may be involved in activating the ATR–CHK1 replication stress response, how DONSON ensures fork stability and promotes checkpoint activation is not known yet and more research is required to understand this.

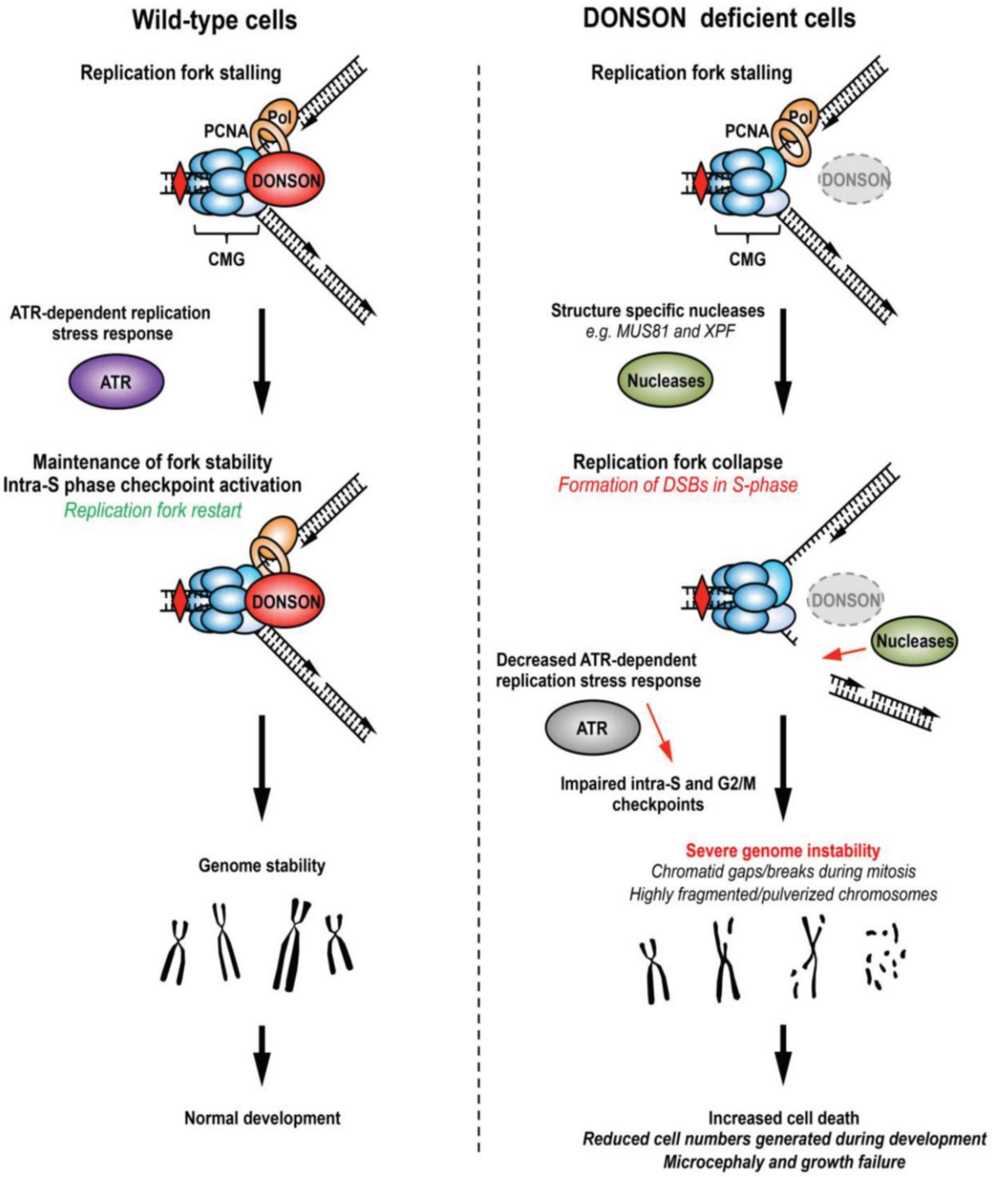


Figure 1.5 **Proposed model depicting how loss of DONSON may lead to genome instability.** DONSON is thought to help stabilise replication forks during replication (left) which leads to normal replication and genome stability. In the absence (or reduced function) of DONSON, stalled replication forks are not stabilised long enough for replication forks to reactivate, leading to double-stranded breaks, genome instability and cell death Figure from (Reynolds et al., 2017), copyright obtained from Springer Nature - 4850530174277.

1.2 DNA replication disorders

The inability to replicate DNA appropriately leads to disease. Complications in replicating DNA cause replication stress, DNA damage and double-stranded breaks. DNA replication is a tightly regulated process, but issues can occur at many points; the initiation of replication, during replication, restarting stalled replication forks and cell cycle checkpoints. It is essential that cells replicate their DNA and progress through the cell cycle at the correct rate. Cells which divide too quickly can cause overgrowth and lead to tumours or cancers (Jackson et al., 2014; O'Driscoll, 2017). Cells which do not divide fast enough can lead to delayed growth of part or all of an organism (Mazouzi et al., 2014). There are many disorders associated with the proteins involved in DNA replication (O'Driscoll, 2017). Many of these lead to reduced organism size as the cells fail to progress through the cell cycle at the expected rate (Klingseisen and Jackson, 2011).

DNA replication disorders cause a range of phenotypes with some genes causing multiple disorders (Harley et al., 2015; Karaca et al., 2019; Knapp et al., 2019; Mo et al., 2018) while some single gene disorders can be caused by variants in multiple different genes (Bicknell et al., 2011b; de Munnik et al., 2015). While there are many DNA replication disorders, the focus of this study is on Meier-Gorlin syndrome and related Primordial Dwarfism disorders (see: Sections 1.1.2-1.3), and only a selection of other replication disorders relevant to this study.

DNA replication disorders resulting from variants within essential components of the pre-RC and pre-IC include: Immunodeficiency 54 and Immunodeficiency 55 (Gennery et al., 2000). Immunodeficiency 54 (IMD54) is caused by biallelic variants in the *MCM4* gene, and patients present with delayed growth, adrenal insufficiency and natural killer cell deficiency (Casey et al., 2012; Eidenschenk et al., 2006; Gineau et al., 2012; Hughes et al., 2012). Increased genomic instability, double-stranded breaks and impaired proliferation has been observed in patient cells (Casey et al., 2012; Eidenschenk et al., 2006; Gineau et al., 2012; Hughes et al., 2012). Immunodeficiency 55 (IMD55) is caused by biallelic variants in genes encoding the GINS complex, specifically *GINS1*, and again patients present with delayed growth and natural killer cell deficiency but have additional chronic neutropenia (Cottineau et al., 2017). Unsurprisingly, patient cells showed impaired GINS complex assembly leading to replication stress, genome instability, impaired checkpoint signalling and cell cycle control (Cottineau et al., 2017).

Additional published DNA replication disorders result from variants within the *PCNA, POLε* and *RECQL4* genes. Ataxia-telangiectasia-like disorder-2 (ATLD2) is caused by biallelic variants in the *PCNA* gene. PCNA forms a sliding clamp around DNA and interacts with many proteins (Baple et al., 2014; O'Driscoll, 2017). ATLD2 patients present with a wide range of phenotypes including, telangiectasia, sensorineural hearing loss, cerebellar atrophy, short stature, neurodegeneration, photosensitivity and ataxia, due to the role of PCNA in many genomic stability pathways (Baple et al., 2014; O'Driscoll, 2017). Variants in DNA Polymerase Epsilon (*POLε*) are associated with a range of disorders including increased susceptibility to colorectal cancer (Bellido et al., 2016; Elsayed et al., 2015; Palles et al., 2013), facial dysmorphism-immunodeficiency-livedo-short stature (FILS) syndrome (Pachlopnik Schmid et al., 2012; Thiffault et al., 2015) and intrauterine growth retardation, metaphyseal dysplasia, adrenal hypoplasia congenita, genital anomalies, and immunodeficiency (IMAGEI) syndrome (Logan et al., 2018). Lastly variants in *RECQL4* also causes a range of syndromes (Siitonen et al., 2009), including Rothmund-Thomson syndrome (Kitao et al., 1999), Ballier–Gerold syndrome (Van Maldergem et al., 2006) and radial hypoplasia, patella hypoplasia and cleft or arched palate, diarrhoea and dislocated joints, little size and limb malformation, slender nose and normal intelligence (RAPADILINO) syndrome (Siitonen et al., 2003, 2009). RECQL4 is a DNA helicase required for genomic stability (Capp et al., 2010; Khakhar et al., 2003; Kitao et al., 1998; Matsuno et al., 2006; Sangrithi et al., 2005; Siitonen et al., 2009).

Rothmund-Thomson syndrome patients present with short stature, congenital bone defects, skin rash (poikiloderma), premature ageing and increased risk of skin cancer later in life (Beghini et al., 2003; Kitao et al., 1999; Larizza et al., 2010; Lindor et al., 2000; Wang et al., 2003). As suggested by the name RAPADILINO syndrome is associated with a range of phenotypes (Kääriäinen et al., 1989; Siitonen et al., 2003, 2009; Vargas et al., 1992). Ballier–Gerold syndrome is characterised by craniosynostosis and radial aplasia (Galea and Tolmie, 1990). *RECQL4* is an example of a single replication gene which causes a wide range and severity of phenotypes and syndromes (Siitonen et al., 2009).

1.1.4 Primordial Dwarfism

Primordial dwarfism is a group of rare genetic disorders characterised by severe pre- and post-natal growth restriction affecting both males and females equally (Codd et al., 2009;

Khetarpal et al., 2016). Primordial dwarfism is largely caused by variants in genes encoding proteins involved in DNA replication, cell cycle progression, proliferation, differentiation and apoptosis (Khetarpal et al., 2016; Müller et al., 2012).

Microcephalic primordial dwarfisms are a group of single gene disorders, in which causative gene variants discovered to date, are involved in DNA replication and cell cycle progression. A common phenotype of these disorders is prenatal and postnatal growth delay, as well as microcephaly and skeletal abnormalities (Klingseisen and Jackson, 2011). Disorders include microcephalic osteodysplastic primordial dwarfism (MOPD) types I, II and III, microcephaly-micromelia syndrome (MIMIS), microcephaly, short stature and limb abnormalities (MISSLA), ATR–Seckel syndrome (SCKL1), and Meier-Gorlin syndrome (MGS) (Evrony et al., 2017; Karaca et al., 2019; Reynolds et al., 2017). These disorders are all caused by variants in important DNA replication/ cell cycle progression genes (Figure 1.6).

MOPD type I and III are caused by variants in the *RNU4ATAC* gene, which encodes a component of the minor spliceosome that is responsible for splicing mRNA transcripts including those from DNA replication genes (Edery et al., 2011; He et al., 2011). MOPD type II is caused by variants in the *PCNT* gene, which encodes a key centrosomal protein essential for the formation of mitotic spindle and chromosome segregation (Kantaputra et al., 2011; Petry, 2016; Rauch et al., 2008).

Seckel syndrome is caused by variants in *TRAIP, CEP63, ATR, NSMCE2, DNA2, CENPJ, NIN, CEP152,* and *RBBP8*, the encoded proteins have crucial roles in DNA damage response signalling, DNA repair, chromosome segregation and genome stability (Harley et al., 2015; Karaca et al., 2019). MGS (see section 1.2.3 and 1.3) is caused by genes involved in the initiation of DNA replication – specifically the pre-replication and pre-initiation complex, *ORC1, ORC4, ORC6, CDT1, CDC6, CDC45, MCM5, GMNN* (Bicknell et al., 2011a, 2011b; Burrage et al., 2015; Fenwick et al., 2016; Vetro et al., 2017). These proteins are essential for licencing replication origins, mediating loading and function of the replication fork. Lastly MISSLA and MIMIS are both caused by variants in *DONSON* (Evrony et al., 2017; Reynolds et al., 2017; Schulz et al., 2018).

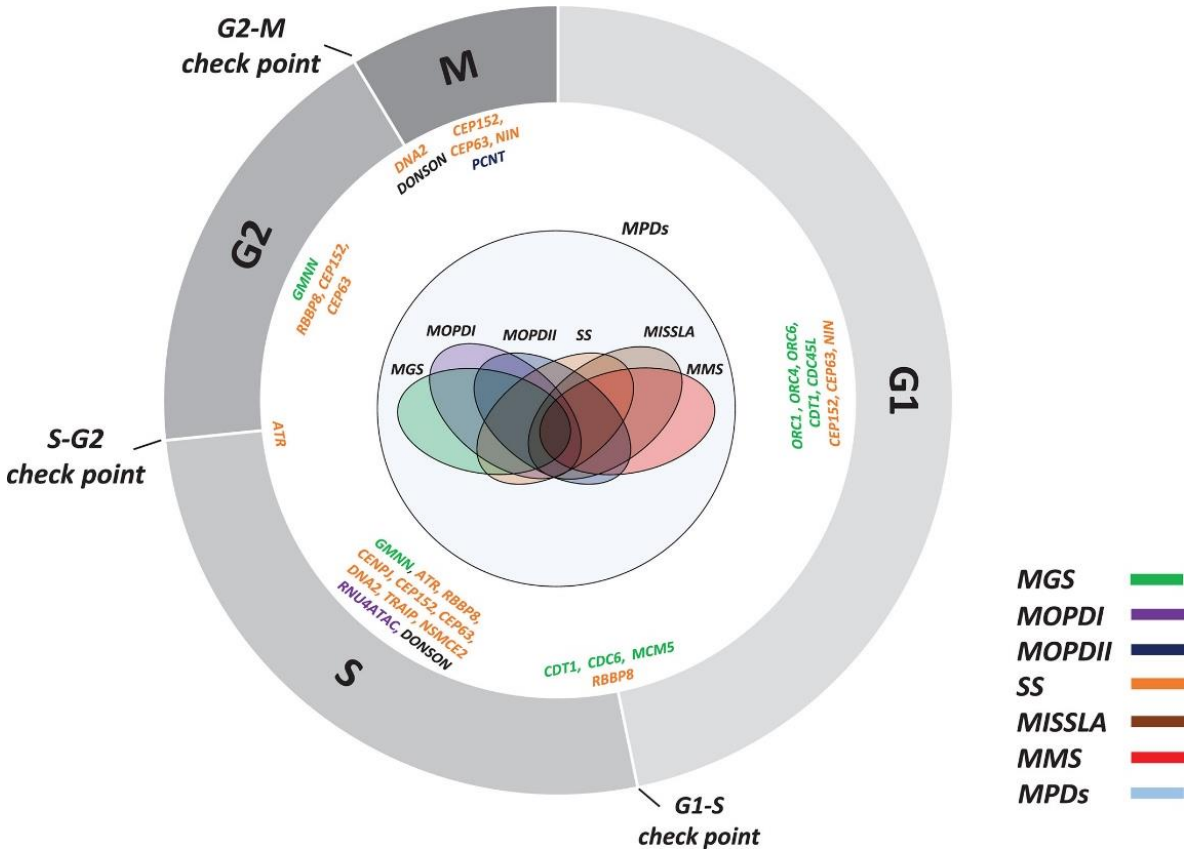


Figure 1.6 Schematic visualising where different genes causing microcephalic primordial dwarfism act in the cell cycle. Genes are colour coded based on primary associated disease; Meier-Gorlin syndrome (MGS) green, microcephalic osteodysplastic primordial dwarfism type 1 (MOPDI) purple, microcephalic osteodysplastic primordial dwarfism type 2 (MOPDII) navy blue, Seckel Syndrome (SS) orange, Microcephaly, short stature, and limb abnormalities (MISSLA) brown, Microcephaly Micromelia syndrome (MIMIS) red, microcephalic primordial dwarfisms light blue. Note genes involved in MGS are present in three out of four cell cycle phases. Phenotypic overlap of diseases is depicted in Venn diagram. Figure adapted from (Karaca et al., 2019), copyright obtained from John Wiley and Sons - 4850530640988.

1.1.5 *DONSON* disorders

DONSON is a gene, which has recently been implicated in a range of microcephalic primordial dwarfism disorders. Pathogenic variants in *DONSON* were found in 29 patients presenting with MIcrocephaly and Short Stature, with Limb Anomalies (MISSLA, OMIM #617604) (Reynolds et al., 2017). These patients all had severe microcephaly and a mild reduction in height as well as assorted other anomalies including a number of skeletal defects. A more recent publication reported a pair of German siblings presenting with MISSLA (Schulz et al., 2018). These siblings had previously been diagnosed with Fanconi anaemia-like syndrome, a disorder which Reynolds et al., 2017 also noted in three

affected siblings from one family in their cohort. Evrony et al., 2017 identified a splicing variant in *DONSON* (c.1047-9A>G) underlying microcephaly-micromelia syndrome (MIMIS) segregating in a First Nations community in Canada. Homozygosity for this variant caused a more severe phenotype with severe intrauterine growth restriction, severe microcephaly, craniofacial dysmorphism, and marked limb malformations. The majority of the patients are stillborn or die within 24 hours of birth but all die within 7 days. The same variant has been found to cause MIMIS in two Saudi Arabian and two Emirati patients (Abdelrahman et al., 2019; Reynolds et al., 2017). More recently, Karaca et al., 2019 reported five affected individuals across four families with variants in *DONSON*. Two individuals presented with MGS, and two siblings with Seckel-like phenotype. The remaining individual presenting with Femoral-Facial syndrome was found to have a *de novo DONSON* variant, but as all other syndromes associated with DONSON are recessive it is possible this last patient has an undetected variant on the other allele.

1.1.6 Meier-Gorlin Syndrome

Meier-Gorlin syndrome (MGS, OMIM 224690) is a disorder caused by variants in genes encoding proteins with essential roles in DNA replication. It is a monogenetic autosomal recessive disorder that presents with an equal occurrence in males and females (Bongers et al., 2001; Hossain and Stillman, 2016; de Munnik et al., 2012a). MGS was originally named ear-patella-short stature syndrome after the diagnostic features (Cohen et al., 1991), but was later renamed by Boles et al., 1994 after the first two authors to describe the syndrome, Meier et al., 1959 and Gorlin et al., 1975. Unlike other forms of dwarfism (such as achondroplasia), patients are proportionally small, with delayed growth, both prenatally and postnatally in the first year of life (Bicknell et al., 2011b; Klingseisen and Jackson, 2011; de Munnik et al., 2012b, 2015). MGS patients do not tend to respond to growth hormone treatments, except for a small number of cases where the patients also presented with low levels of IGF1 (de Munnik et al., 2012b). Ninety-seven percent of patients have two of the following; short stature, microtia (small malformed ears) and patellar aplasia/hypoplasia (small or absent kneecaps) (de Munnik et al., 2015). Associated features of the disease include feeding problems (de Munnik et al., 2015), mammary hypoplasia in females (de Munnik et al., 2012b), urogenital anomalies (Kora et al., 2016) and proportional microcephaly (de Munnik et al., 2012a). Characteristic facial

features such as a small mouth with full lips, micrognathia (undersized jaw) and beaked nose are often present in infants, this changes in adults to a high forehead with a prominent and narrow nose with a broad nasal bridge (Bongers et al., 2003, 2005; de Munnik et al., 2012b). Despite an often small head size, the majority of patients have normal intellect (Bicknell et al., 2011b; Shawky et al., 2014).

Variants underlying MGS have been found in genes involved in DNA replication. Specifically genes in the pre-replication complex, *ORC1, ORC4, ORC6, CDT1, CDC6, MCM5, GMNN* and genes in the pre-initiation complex *CDC45*. Variants in any of these genes lead to disruptions in initiation of DNA replication, therefore variants are hypomorphic as null homozygous null alleles would be lethal (Bicknell et al., 2011a, 2011b; Burrage et al., 2015; Fenwick et al., 2016; Guernsey et al., 2011; Kerzendorfer et al., 2013; Vetro et al., 2017). *GMNN* is the only gene known to harbour *de novo* gain of function variants causing the disorder. GMNN normally functions as a negative regulator for CDT1. MGS variants in GMNN are found in the destruction box motif within the N-terminus of GMNN, a motif is recognised by anaphase-promoting complex (APC) which degrades GMNN during the metaphase-anaphase transition (Burrage et al., 2015). MGS-related variants prevent the degradation of GMNN, effectively reducing available CDT1 for DNA replication licensing, and therefore acting in a similar fashion to MGS variants in other preRC and preIC encoding genes.

All of the previously mentioned MGS genes are essential for the initiation of DNA replication. There are two main hypotheses as to how the variants cause disease. The first is that reduced protein function is likely to lead to slower initiation of DNA replication, leading to slower cell cycle progression and division, therefore resulting in delayed growth, leading to the severe short stature phenotype (Klingseisen and Jackson, 2011). The other hypothesis is that normally during replication there are many more licenced origins than required, giving rise to 'dormant origins' (Ge et al., 2007). The majority of the time these won't be required to fire, but they are able to fire in times of replication stress leading to replication being completed as efficiently as possible (Blow and Ge, 2009; Blow et al., 2011; Ge et al., 2007; Woodward et al., 2006). It is thought that MGS patients may not have the same reservoir of dormant origins and so during critical points in development when rapid proliferation is required, delays can occur which slow cell division, ultimately reducing the number of cells and reducing the size of an organism. The critical point(s) in development have not yet been identified.

1.3 Meier-Gorlin syndrome individuals participating in this study

Whole genome sequencing of our MGS patient cohort revealed two patients to have novel homozygous variants in either the known MGS gene, *ORC1* or a new MGS gene *DONSON* (Dr Karen Knapp, Department of Pathology, University of Otago). The two MGS patients we are focusing on in this study are LB005 (*ORC1* c. 1865T>C, p.Leu622Pro) and LB008 (*DONSON* c.631C>T, p.Arg211Cys). Both come from consanguineous families, and have a reduction in head size, measured as occipitofrontal circumference (OFC) of -4.1 and -4.0 standard deviations (SD) below the mean and also have a reduction in height of -4.2 and -3.5 SD for LB005 and LB008 respectively. They also have small hypoplastic/dysplastic ears, micrognathia (small jaw) and absent patella which is characteristic of MGS. LB005 has no apparent intellectual disability. LB008 has mild hearing loss and a high arched palate.

Two additional patients LB191 and LB192 with a predicted homozygous splicing variant in *CDC45* (c.1441-2A>G) were also studied. These two siblings were referred after both pregnancies were terminated ~ 20-week's gestation due to severe growth restriction and malformations. The parents were healthy and consanguineous (mother's mother and father's father were siblings); they had no family history of inherited conditions, congenital abnormalities, recurrent miscarriages or mental retardation. An ultrasound at 17 weeks gestation showed the first sibling's growth was lagging 3 weeks behind expected. A follow up ultrasound at 21.5 weeks showed continued growth retardation and a variety of other malformations. The pregnancy was terminated and an autopsy confirmed observations. The second pregnancy was monitored from the beginning and an ultrasound at 16.5 weeks showed growth was lagging two weeks behind, the next ultrasound at 20.1 weeks confirmed growth was lagging 2.9 weeks behind along with other malformations similar to the first pregnancy. Both siblings presented with many complications including absent eyelids, intrauterine growth approximately three weeks behind, cardiac abnormalities, oligohydramnios, dolichocephaly, cerebral ventriculomegaly, kyphotic lumbar spine, lordotic thoracic spine, thin ribs, absent clavicles and absent genitalia.

1.1.7 Novel variants found in *ORC1*

MGS patients have previously been reported to have variants in *ORC1* (Bicknell et al., 2011a, 2011b; Guernsey et al., 2011). All of these patients have at least one deleterious allele in the BAH domain at the N-terminal domain of the protein (Figure 1.7). ORC1 is the only ORC subunit that contains a BAH domain (Noguchi et al., 2006), which is involved in protein-protein interaction, as well as other roles such as chromatin binding (Callebaut et al., 1999; Kuo et al., 2012; Noguchi et al., 2006; Zhang et al., 2015). Variants in the BAH domain were previously thought to be causing the severe MGS phenotype (Kuo et al., 2012). The novel homozygous *ORC1* variant (c. 1865T>C) is however found at the 3′ end of the *ORC1* gene, distant from the BAH domain. Two other MGS patients have previously been found to have compound heterozygous variants involving a null allele and a missense variant (c.1996C>T, p.Arg666Trp) located within the same area of *ORC1*. The first of these two patients was published in Guernsey et al., 2011 with the phenotype previously described in Shalev and Hall, 2003, and comes from a non-consanguineous family. At 25 years old she presents with a reduction in head size (-4.8 SD), a reduction in height (-5.5 SD), absent kneecaps, small but not misshapen ears (-7 SD) and breast hypoplasia. She presented with dislocation of her hips and knees at birth, had feeding problems, normal intellect and has completed three years of college (Guernsey et al., 2011; Shalev and Hall, 2003). The second patient is unpublished with no data available except a clinical diagnosis of MGS (Dr Louise Bicknell, personal communication).

The two missense variants (p.Leu622Pro and p.Arg666Trp) are located in the AAA domain toward the C-terminal end of the protein (Bicknell et al., 2011a). The AAA domain is important for chromatin binding, partner protein recruitment, ATP hydrolysis, oligomerization and double-stranded DNA binding (Bleichert et al., 2017). ORC1 plays an essential role and mouse model *ORC1* knockouts have been found to be lethal (Cooper, 2016; Meehan et al., 2017). MGS patient cells also show a reduced amount of both chromatin-bound and unbound ORC1, suggesting variants in the BAH domain impact protein stability (Bicknell et al., 2011a). This suggests that variants in this gene cause a knockdown of function rather than complete knockout.

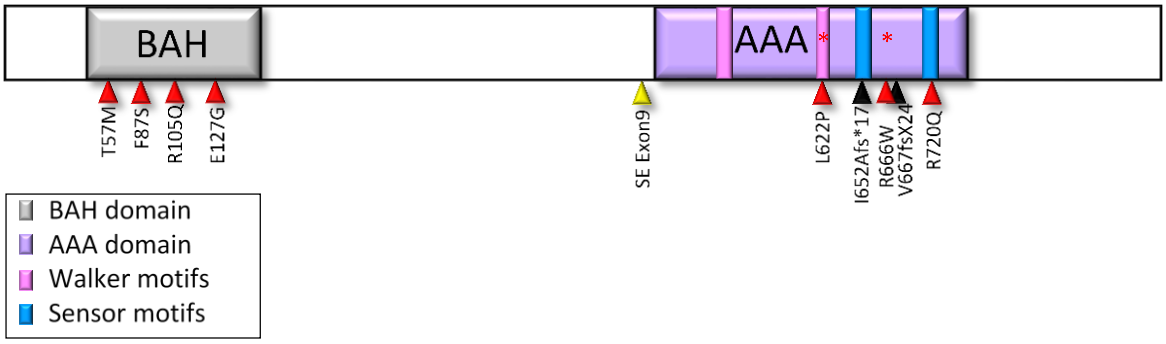


Figure 1.7 **Important ORC1 domains and previous MGS variants.** *ORC1* MGS variants are indicated by arrowheads. Missense in red, splicing in yellow, truncation in black. The two variants studied in this thesis, p.L622P and p.R666W are indicated with a red asterisk. Other important features of ORC1 include the BAH domain (grey), AAA domain (purple), walker motifs (pink) and sensor motifs (blue).

Previous research found that ORC1 depletion in zebrafish leads to a MGS-like phenotype which can be rescued by the addition of wild type (WT) human ORC1 (Bicknell et al., 2011a; Kuo et al., 2012). ORC1 has also been reported to have roles in cilia formation (Stiff et al., 2013) and regulation of centriole and centrosome reduplication (Hemerly et al., 2009; Kuo et al., 2012), suggesting ORC1 could be involved in more than just the initiation of DNA replication. The BAH domain of ORC1 has also been found to be an H4K20me2 binding molecule, suggesting a role in chromatin modifications (Kuo et al., 2012; Zhang et al., 2015). Kuo et al., 2012 suggested that this H4K20me2-BAH domain interaction is the underlying cause of the MGS phenotype. This interaction was hypothesized to be causing the MGS phenotype, as all reported patients with *ORC1* variants had at least one allele with a BAH domain variant (Kuo et al., 2012). However, the novel *ORC1* p.L622P missense variant is homozygous and is located in the AAA domain, not in the BAH domain of the protein. This previous theory of the BAH domain being the cause of the MGS phenotype therefore does not explain why the same phenotype is observed in patients who do not have variants in the BAH domain.

1.1.8 Novel *CDC45* variants

Biallelic *CDC45* variants have previously been reported in patients presenting with craniosynostosis and MGS (Fenwick et al., 2016). The majority (80%) of CDC45-MGS patients presented with craniosynostosis compared to just 4% of MGS patients with variants in other genes (Fenwick et al., 2016; de Munnik et al., 2012a; Ting et al., 2019).

Although there is a slight difference in phenotype with these patients, besides the addition of craniosynostosis the secondary phenotypes are not considered more severe. In this study two severely affected foetuses presenting with a range of severe secondary phenotypes, including absent eyelids, cardiac abnormalities and craniosynostosis are presented. Both siblings were terminated at approximately 20 weeks gestation due to severe growth restrictions and secondary phenotypes. These siblings harbour a homozygous variant (c.1441-2A>G) within an essential splice site in *CDC45*. We suspected this variant may cause a more severe reduction in CDC45 protein levels compared to previously described patients with compound heterozygous splicing variants in *CDC45*.

1.1.9 Novel MGS variants in *DONSON*

DONSON's role in replication fork stability and the S-phase checkpoint has previously been described by Reynolds et al., 2017. DONSON has been found to be involved in a range of MPD disorders, with patients with variants in *DONSON* tending to present with moderate to severe microcephaly and a slight reduction in height (Evrony et al., 2017; Reynolds et al., 2017). While MGS patients tend to be proportionally small, MISSLA and MIMIS patients tend to have a greater reduction in head size (Figure 1.8). It was thought that the brain is more severely affected as it undergoes a period of rapid proliferation with a very short G1 phase (Kalogeropoulou et al., 2019), and is therefore disproportionally affected by chromosomal instability and cell death. MGS patients tend to be proportionally small as all other MGS genes play a role early on in DNA replication effecting all cells evenly. It is hypothesised that DONSON may play another role earlier in DNA replication, which would be consistent with other MGS genes. Reynolds et al., 2017 showed that patient cell lines have reduced DONSON protein levels and mislocalisation of protein in the cytoplasm. *DONSON* is highly conserved between species and a homozygous mouse model is lethal (Evrony et al., 2017; Reynolds et al., 2017). This suggests that DONSON plays an important role in the cells and patient variants are likely to be knockdown rather than complete loss of function.

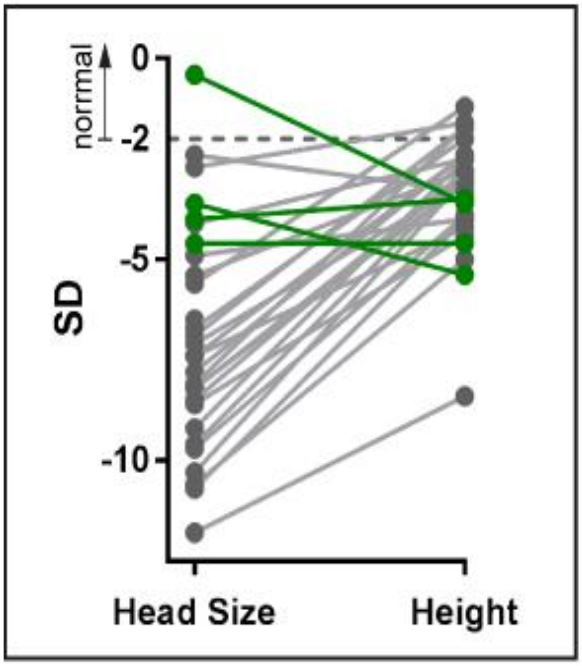


Figure 1.8 Differences in phenotype between MISSLA patients and MGS-DONSON patients. Phenotypic data for MGS patients (green) and MISSLA patients (grey) show differences in phenotype between the two groups. MGS patients are proportionally small or have a larger reduction in height than head size whereas DONSON patients have a larger reduction in head size than height. Figure adapted from (Knapp et al., 2019) (figure created by Dr Louise Bicknell for Bicknell Marsden grant).

1.1.9.1 DONSON in MGS patients

Through whole genome sequencing a novel homozygous (c.631C>T, p.R211C) variant was found in a patient (LB008) presenting with MGS. A total of four patients with five missense and one splicing variant were found in the cohort (24%), making *DONSON* a relatively common MGS gene (Knapp et al., 2019). At the time of discovery, there were no published *DONSON* variants in MGS patients. The c.631C > T (p.Arg211Cys) *DONSON* variant has subsequently been published by Karaca et al., 2019 (Figure 1.9). Interestingly these authors also found a second MGS patient presenting with the same missense variant – both patients are from Turkey. LB008 was clinically diagnosed with MGS with an almost proportionate reduction in head size (-3.97 SD) and height (-3.48 SD), absent patella and small dysplastic ears, the three main markers of MGS. This phenotype is very clinically typical of MGS and different from the severe microcephaly seen with most other DONSON disorders (Figure 1.9). LB008 was chosen from the four MGS-DONSON patients, as this was the only patient with a homozygous missense variant, more suitable for CRISPR editing by homology directed repair (HDR) than a compound heterozygous variant.

Other functional studies on DONSON were carried out using fibroblasts derived from LB128, the only patient for which cells were available. LB128 is compound heterozygous for missense variants c.670C>T (p.Pro224Ser) and c.809A>G (p.Tyr270Cys). This patient also presented with the triad of MGS markers, microtia, absent patella and almost proportionate reduction in size, (-5.26 SD) head size and (-4.35 SD) height (Figure 1.9).

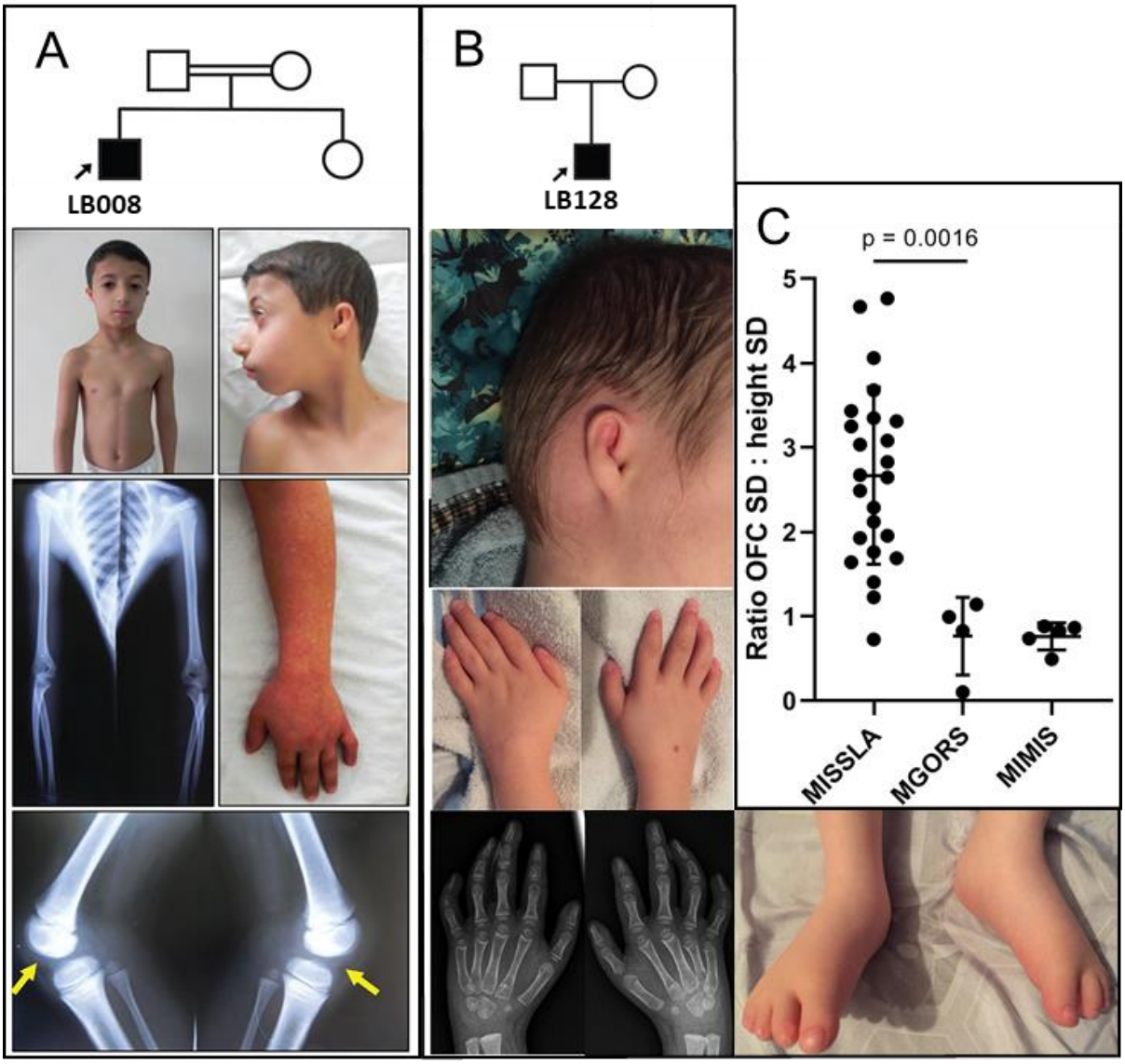


Figure 1.9 **MGS patients with variants in *DONSON* show the classical triad of MGS symptoms.** Note the absent kneecaps in LB008, small-misshaped ears and bowed long bones (secondary phenotype) in both patients. A) LB008 comes from healthy consanguineous parents and presents with the classical triad of MGS phenotypes. B) LB128 comes from healthy unrelated parents and shows the classical triad of MGS phenotypes. C). Head to height ratio is plotted for MISSLA patients, MGS patients and MIMS patients. A proportionally small head to height ratio is seen in both MGS and MIMIS patients while MISSLA patients show a much more diverse spread of head to height ratios. Figure adapted from (Karaca et al., 2019 & Knapp et al., 2019), copyright obtained from John Wiley and Sons – 4850530640988 and under the Creative Commons Attribution 4.0 International License.

1.4 Aims and objectives

This project aims to use CRISPR-Cas9 HDR to introduce Meier-Gorlin syndrome patient variants into hTERT-RPE-1 cell lines, a commonly used cell model for studying DNA replication and DNA damage. These cells will be used to investigate how these MGS variants affect the cells ability to load chromatin and progress through the cell cycle. The aims of this project changed slightly as it progressed, and instead focused on using transfected HeLa and HEK293FT cells and patient fibroblasts to investigate the cellular consequences of these variants.

The overall aim of this project was to understand the cellular consequences of novel Meier-Gorlin syndrome variants using CRISPR-Cas9 and a variety of cellular assays.

Aims:

Explore whether CRISPR-Cas9 can be used to knock in patient MGS variants (essential genes) into hTERT-RPE-1 cells.

Use a variety of cellular techniques to look at chromatin loading, DNA replication, and cell cycle progression.

Investigate the novel consequences of a novel *CDC45* splicing variant.

Investigate the cellular effects of MGS-*DONSON* variants.

2 Methods

Families were recruited to the study through Dr Louise Bicknell's international collaborations on Meier-Gorlin syndrome. Ethical approval for this study was provided by the New Zealand Health and Disability Ethics Committee (16/STH/3).

A full list of chemicals/reagents (Appendix A.1) and buffers (Appendix A.2) used is this project is available in the appendix.

2.1 Molecular biology

2.1.1 DNA extraction

2.1.1.1 Quick DNA extraction from cultured cells

One µL Lucigen QuickExtract (Mediray) was added to 50-200 µL trypsinized cell culture in 8 well PCR strip tubes. Cells were vortexed for 15 seconds before being incubated at 65 °C for 6 minutes. Samples were vortexed again for 15 seconds before being incubated at 98 °C for 2 minutes to deactivate the enzymes. Samples were left in the fridge overnight for cell components to settle. DNA was then diluted 1:10 for use in PCR reactions (2.1.2) or an ethanol precipitation reaction (2.1.1.4) was performed to clean up the DNA.

2.1.1.2 Crude DNA extraction

Fifty µL 1 M NaOH, 10 mM EDTA pH12 was added to ~150 µL trypsinized cells. Cells were incubated at 95 °C for 15 minutes before being cooled to 4 °C for 5 minutes and neutralised with 50 µL neutralising solution (2 M Tris-HCl, pH5 – Appendix A.2). DNA samples were stored at -20 °C and used diluted 1:10 in PCR reactions (2.1.2).

2.1.1.3 High quality DNA extraction

Qiagen DNeasy Blood & Tissue Kit (Bio-Strategy) was used to extract high quality control DNA from cells cultured in a T25 or T75 flasks, following the manufacturer's instructions.

2.1.1.4 **Ethanol precipitation of DNA**

Ten % sodium acetate (Thermo Scientific) was added to each tube. Two and a half volumes of absolute ethanol was added to each DNA sample. Samples were incubated at -20 °C for 24 hours before being centrifuged at max speed (3220 g) for 40 minutes. Supernatant was removed and samples washed with 70% ethanol, centrifuged again at max speed for 15-20 minutes. Pellets were left to air dry for a few hours before being resuspended in 10-20 µL resuspension buffer or MilliQ water for at least 10 minutes at 37 °C or a few hours at room temperature. Samples were stored long term at 4 °C.

2.1.2 Polymerase chain reaction (PCR)

Primers were designed using primer3 (Koressaar and Remm, 2007; Untergasser et al., 2012) or primer BLAST (Ye et al., 2012). A full list of primers used is avalible in Appendix A.3.

2.1.2.1 **Standard PCR**

Standard PCRs were set up using (2x) DreamTaq Green PCR Master Mix (Thermo Scientific). Five µL DreamTaq, 1 µL 20 nM primer mix or 2 µL 10 nM primer mix, 1-4 µL 10 nM DNA made up to 10 µL with dH_2O.

PCRs were first optimised by altering the annealing temperature or changing to Phusion Flash High-Fidelity master mix (2.1.2.2). If these methods failed 1 µL 5x C.E.F. was added to improve amplification. In the case of long PCR products (>2.5 kb) the extension temperature was reduced from 72 °C to 68 °C.

2.1.2.2 **Phusion flash high-fidelity PCR**

2x Phusion Flash High-Fidelity PCR Master mix (Thermo Scientific) was used for large fragments or optimisation. PCR was set up with 5 µL master mix, 1 µL F primer (10 nM), 1 µL R primer (10 nM), 2 µL dH_2O and 1 µL DNA (10 nM).

2.1.2.3 **PCR clean up**

PCR products were purified to remove unincorporated dNTPS before sequencing to reduce noise. One µL Exonuclease I (NEB) and 0.5 µL Shrimp Alkaline Phosphatase (rSAP) (NEB) were added to 5 µL of PCR product, and incubated at 37 °C for 15 minutes before being inactivated by incubating at 80 °C for 15 minutes. Appropriate concentration of samples were added to BigDye sequencing reaction (2.1.4.1) or stored at 4 °C.

2.1.2.4 **PCR cycling conditions**

PCR reactions were carried out on a BIO-RAD DNA Engine Dyad Peltier Thermal Cycler. Cycling conditions are listed below.

2.1.2.4.1 Standard PCR cycling conditions

Initial denaturing:	95 °C for 10 minutes
Denaturing:	95 °C for 30 seconds
Annealing:	50-65 °C for 30 seconds
Extension:	72 °C for 30 seconds
Cycle:	40 times
Final Extension:	72 °C for 10 minutes
Hold	10 °C forever

2.1.2.4.2 Touchdown PCR cycling conditions

Initial denaturing:	95 °C for 10 minutes
Denaturing:	95 °C for 30 seconds
Annealing:	65 °C for 30 seconds
Extension:	72 °C for 30 seconds
Cycle:	3 times
Denaturing:	95 °C for 30 seconds
Annealing:	62 °C for 30 seconds
Extension:	72 °C for 30 seconds
Cycle:	3 times

Denaturing:	95 °C for 30 seconds
Annealing:	59 °C for 30 seconds
Extension:	72 °C for 30 seconds
Cycle:	3 times
Denaturing:	95 °C for 30 seconds
Annealing:	56 °C for 30 seconds
Extension:	72 °C for 30 seconds
Cycle:	36 times
Final Extension:	72 °C for 10 minutes
Hold	10 °C forever

2.1.2.4.3 Phusion flash PCR cycling conditions

Initial denaturing:	98 °C for 10 seconds
Denaturing:	98 °C for 1 Second
Annealing:	55 °C or 60 °C for 5 seconds
Extension:	72 °C for 30 seconds (15 second per kb)
Cycle:	30 times
Final Extension:	72 °C for 1 minute
Hold	4 °C forever

2.1.2.4.4 Mutagenesis cycling conditions

Initial denaturing:	98 °C for 10 minutes
Denaturing:	98 °C for 30 seconds
Annealing:	55 °C for 30 seconds
Extension:	68 °C for 8 minutes
Cycle:	18 times
Final Extension:	68 °C for 10 minutes
Hold:	10 °C forever

2.1.2.4.5 BigDye

Initial denaturing:	96 °C for 1 minute, 15 seconds
Denaturing:	96 °C for 45 seconds
Annealing:	50 °C for 45 seconds
Extension:	60 °C for 8 minutes, 30 seconds
Cycle:	25 times
Hold:	10 °C forever

2.1.2.4.6 Allele-Specific PCR

Initial denaturing:	95 °C for 10 minutes
Denaturing:	95 °C for 30 seconds
Annealing:	70 °C for 30 seconds
Extension:	72 °C for 30 seconds
Cycle:	40 times
Final Extension:	72 °C for 10 minutes
Hold	10 °C forever

2.1.3 Restriction endonuclease digests

Restriction endonuclease digests were performed using restriction enzymes from NEB or Thermo Scientific. One µg or less DNA was incubated with 1 µL restriction enzyme, 1 µL buffer made up to 10 µL in dH_2O. Restriction digests were performed following the manufacturer's instructions. BbsI (Thermo Scientific) or BsrGI (NEB) were used to linearize PX458 for insertion of sgRNA sequence or Gibson assembly. Taq (α) 1 or SacI (NEB) were used to genotype CRISPR clones. DpnI *(*NEB) was used to degrade any remaining *E.coli* plasmid after mutagenesis. DpnI cleaves methylation sites present in *E.coli* amplified plasmid but not present in mutagenesis plasmids amplified by PCR.

2.1.4 Agarose gel electrophoresis

PCR products and restriction endonuclease digests were analysed by electrophoresis on 1-2.5% agarose gels. Agarose power (MediRay) was added to a conical flask with 50 mL

0.5x TBE (Appendix A.2), and microwaved for 2 minutes or until dissolved. Molten agarose was left to cool for a minute before 5 µL SYBR™ Safe DNA Gel Stain (Invitrogen) was added and swirled to mix. The mixture was transferred to a 50 mL falcon tube and 0.5x TBE was added to make up to 50 mL. The gel was poured into a minigel tank containing 1-2, 16 or 20 well combs, and left to set for half an hour. Once set combs were removed and the gel was covered with 0.5x TBE, samples and 100 bp or 1 kb ladder (NEB) were loaded. One µL of loading dye (NEB) was mixed with 3 µL of restriction digest samples or PCR samples run with Phusion flash master mix, before being loaded on the gel. Two-9 µL DreamTaq PCR reactions were loaded on the gel directly without the need to add loading dye.

Electrophoresis was carried out at 80 V/cm for 25-50 minutes before being imaged on the UVITEC Cambridge gel doc or BioRad UV light transilluminator.

2.1.4.1 **Purify DNA from agarose gel**

Bands were extracted from agarose gels using the GeneJET Gel Extraction Kit (Thermo Scientific) following the manufacturer's instructions.

2.1.5 Sanger sequencing

2.1.5.1 **Big dye reaction**

BigDye sequencing reactions were set up using 1.75 µL 5x sequencing buffer (Genetic Analysis Services (GAS) Otago), 0.5 µL Big Dye Terminator (GAS Otago), 1 µL (3.2 µM) Primer and >200 ng plasmid DNA or 1 ng/100 bp cleaned up PCR product (2.1.2.3) and made up to total volume of 10 µL with distilled water. Sequencing reactions were performed using the BigDye program (2.1.2.4.5) in a thermocycler.

2.1.5.2 **Sequencing clean up and calling**

Two µL 3 M Sodium acetate (Thermo Scientific), 10 MilliQ water and 50 µL absolute ethanol was added to each sample. Samples were incubated for 10 minutes at room temperature before being centrifuged at 3220 g for 20 minutes. Samples were then washed twice with 150 µL, 70% ethanol. Pellets were left to air dry before being

transported to Genetic Analysis Services Otago (Department of Anatomy) for analysis on their Applied Biosystems™ 3730xl DNA Analyzer.

2.1.5.3 Sanger sequencing analysis

Sanger sequencing results were analysed using Sequencher® version 5.4.6 DNA sequence analysis software, Gene Codes Corporation, Ann Arbor, MI USA http://www.genecodes.com or Geneious 11.1.3 (https://www.geneious.com).

2.1.6 Cloning

2.1.6.1 Plasmids

Table 2.1 Table of plasmids used in this project.

Plasmid	Source
p.EGFP-empty	(Reynolds et al., 2017)
p.EGFP-DONSON	(Reynolds et al., 2017)
pSpCas9(BB)-2A-GFP (PX458) empty, *ORC1*, *DONSON*	Addgene (#48138) (Ran et al., 2013a)
PCS2+Cas9-mSA	Addgene (#103882) (Gu et al., 2018)
Flag-tag-CDC45	(Dr Karen Knapp)
pSpliceExpress	Addgene (#32485) (Kishore et al., 2008)

2.1.6.2 Chemically competent *E.coli* top 10

Chemically competent *E.coli* Top10 were used for all transformations. Fifty mM $CaCl_2$, 10 mM Tris-HCL pH 7.4 solution was used to make chemically competent *E.coli* top10 (protocol Appendix B.1).

2.1.6.3 *E.coli* transformations

Fifty µL aliquots of *E.coli* were thawed on ice ~20 minutes. Fifty µL *E.coli* was incubated with 1-5 µL plasmid or cloning reaction for 10 minutes on ice. Vials were heat shocked for 30 seconds at 42 °C then incubated on ice for 2 minutes. The transformation reaction was either plated straight away or 300 µL LB medium (Appendix A.2) added and incubated shaking for 30-45 minutes before – centrifuging, removing 300 µL supernatant and resuspending in the remaining 50 µL.

2.1.6.4 Colony PCR

Colony PCR (cPCR) was used to screen *E.coli* colonies for correct plasmids. These were set up as previously described (2.1.2.1) using one plasmid-specific and one insert-specific primer. *E.coli* colonies were picked using a P200 tip and suspended in 20 µL dH_2O, 3 µL of the suspension was added to each PCR reaction and 2 µL was spotted on an agar plate containing appropriate antibiotic and incubated overnight 37 °C. A scraping of colonies containing the correct insert were taken for a mini plasmid prep followed by Sanger sequencing to confirm the correct sequence.

2.1.6.5 Plasmid purification and glycerol stock preperation

Glycerol stocks were prepared for every plasmid by adding 250 µL overnight culture to 250 µL 50% glycerol (Thermo Scientific) and storing in the -80 °C freezer. Small-scale plasmid preparations were performed using the GeneJET miniprep kit (Thermo Scientific) following the manufacturer's instructions. Large-scale plasmid preparations were performed using the GeneJET midiprep kit (Thermo Scientific) following the manufacturer's instructions. Plasmid concentrations were measured by NanoDrop (2.1.6.6) or Qubit (2.2.6.7).

2.1.6.6 Initial quantification of DNA, RNA and *E.coli* concentrations

DNA, RNA or *E.coli* concentrations were measured using the NanoDrop™ OneC UV-Vis Spectrophotometer. The NanoDrop was set to zero using elution buffer, RNAase free water or LB medium. Two-3 µL of sample was added to the clean pedestal and measurements taken. DNA and RNA concentrations were measured by absorbance at 260

nm. Absorbance at 230 nm and 280 nm was used to calculate A260/A230 and A260/A280 purity ratios, an A260/A280 ratio of ~1.8 (DNA) or ~2.0 (RNA) and an A260/A230 ratio between 1.8 and 2.2 was considered pure. *E.coli* concentrations were measured using OD600 (600 nm) reading. Two-three replicates were performed to confirm concentration.

2.1.6.7 **Conformation of plasmid concentrations for transformations**

Qubit was used to confirm plasmid concentrations for transfections. Other members of the laboratory group had found plasmid concentrations had been overestimated by the NanoDrop in some circumstances. The plasmid concentrations were measured using the Qubit™ dsDNA BR Assay Kit (Thermo Scientific) following the manufacturer's instructions.

2.1.6.8 **sgRNA cloning**

sgRNA were designed using http://crispr.mit.edu/, however this tool is no longer functional. Guides were chosen based on proximity of cut site to desired knock-in variant, a high off-target score (few off-targets in exons and introns) and fewest variants required to disrupt the PAM site, add or remove a restriction enzyme site and insert a patient variant.

DNA oligonucleotides representing the sgRNAs were phosphorylated using NEB T4 PNK and annealed. Double-stranded DNAs representing the sgRNA sequences were then ligated into the PX458 (Addgene #48138) vector at the BbsI restriction cut site using NEB T4 ligase as per the manufacturer's instructions.

Repair templates (ssODN) were designed with homology arms of at least 88 bp on each side of the variant site (Yang et al., 2013).

2.1.6.9 **Site directed mutagenesis**

Primers were designed using the QuikChange® Primer Design tool. This tool allows 1-7 bp to be deleted or substituted from the original plasmid sequence.

Site directed mutagenesis was used to introduce a patient c.631C>T missense variant into the pEGFP-DONSON vector. The pEGFP-DONSON and pEGFP-empty vectors were created and used by Reynolds et al., 2017. The forward and reverse mutagenesis primers contained the variant in the middle (in red) CTCTCCTCTGAGCTCTGTTGTACCTTCCAGC. The pEGFP-DONSON plasmid was amplified using the Phusion Flash high-fidelity PCR protocol as previously described (2.1.2.2) and Phusion flash PCR cycling conditions (2.1.2.4.3).

As the pEGFP vector contained kanamycin resistance, transformed *E.coli* (2.1.6.3) was plated on kanamycin containing agar plates. Mini plasmid preps (2.1.6.5) were performed on two colonies, these were Sanger sequenced (2.1.4) to confirm variant.

Mutagenesis was also used to delete 6 bp of cDNA from flag tagged *CDC45* cDNA plasmid. This allowed over expression of the cDNA from a patient with a splicing variant in HeLa cells.

2.1.6.10 **Gibson assembly**

Gibson assembly was performed following the manufacturer's instructions. First a restriction digest was performed (2.1.5) using BsrGI to linearize the PX458 plasmid at the end of the Cas9-GFP complex.

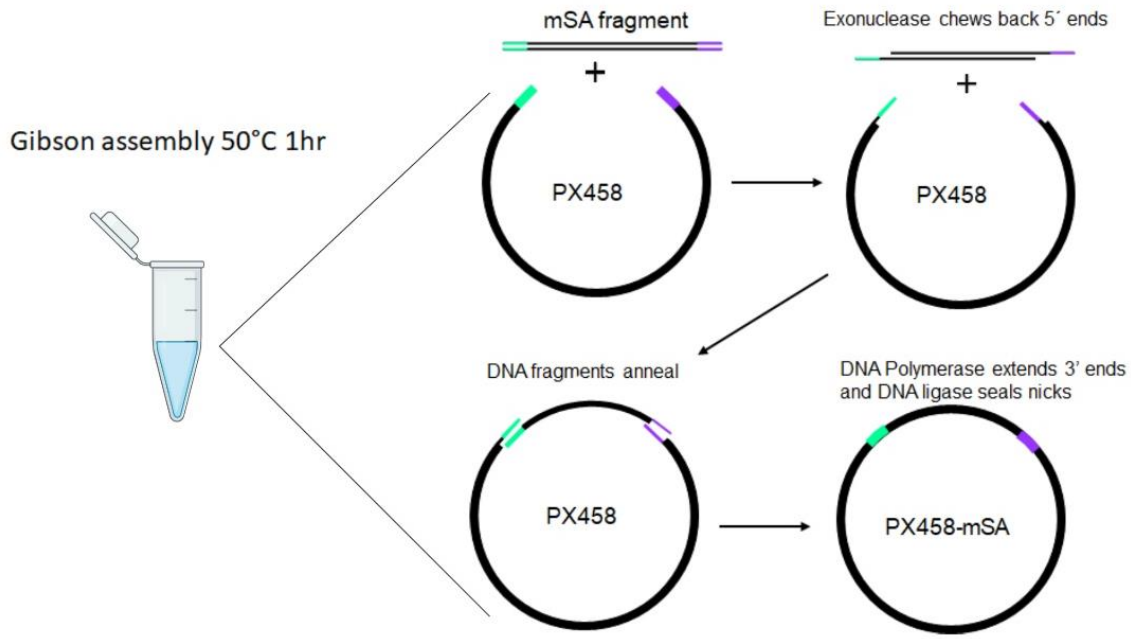


Figure 2.1 **Gibson assembly master mix assembles two overlapping fragments.** The Gibson assembly reaction takes place in a single tube at 50 °C and can be used to combine multiple fragments with overlapping ends. First, the mSA fragment was PCR amplified with primers containing overlapping PX458 sequence. Next, the endonuclease digests the 5' ends of the PCR product and plasmid. The two complimentary strands anneal before DNA polymerase extends the 3' ends and DNA ligase seals the nicks. (Figure created with BioRender).

Next, the monomeric streptavidin (mSA) fragment was amplified by PCR from PCS2 + Cas9-mSA (2.1.2.2). The end of the Gibson assembly primers contain part of the PX458 plasmid. The Gibson assembly reaction was then performed to incorporate this fragment into the PX458 plasmid. This created a plasmid containing the specific sgRNA encoding sequence under the U6 promotor and a Cas9-GFP-mSA complex under a CAG promoter (Figure 2.2).

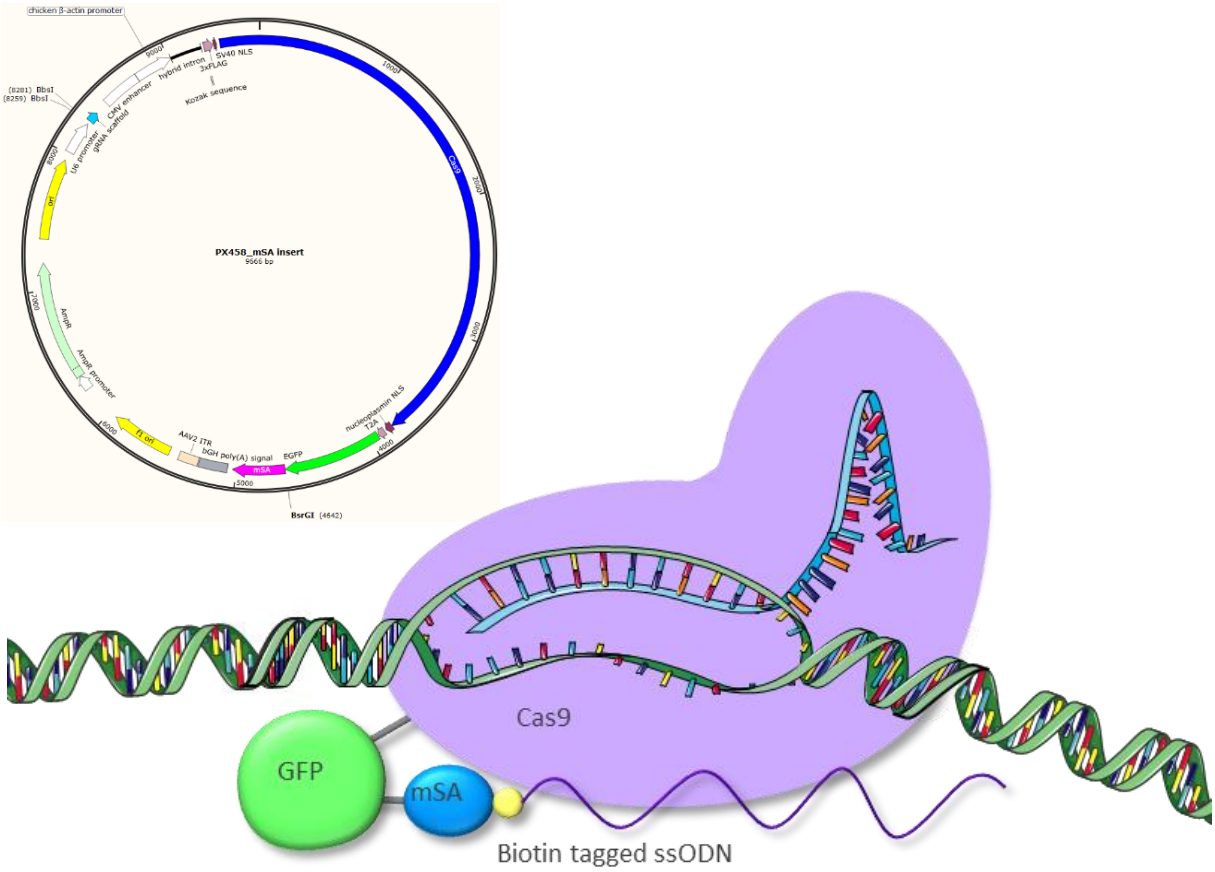


Figure 2.2 **PX458 plasmid modified by Gibson assembly to include mSA fragment.** Plasmid map shows insertion site of the mSA fragment (BsrGI), alongside the orientation of the Cas9-GFP-mSA construct in the new PX458-mSA plasmid. Schematic shows how this new Cas9-GFP-mSA complex should interact with the sgRNA, DNA and biotin tagged ssODN at the desired cut site (DNA adapted from https://smart.servier.com/ template).

2.1.6.11 **Gateway cloning**

Gateway cloning was used to create a p.SpliceExpress vector containing a small part of the *ORC1* gene. This allowed assessment of potential splicing variants between WT *ORC1* and LB005 homozygous variant p.L622P. Ex12-14 and >200 bp of intron either side were amplified by Phusion PCR (2.1.2.2). Gateway cloning was performed following the manufacturer's instructions and *E.coli* was transformed as previously described (2.1.6.3) and plated on ampicillin containing agar. Gateway cloning uses attB/attP sites for site-specific recombination to insert the desired fragment (Figure 2.3). Intact plasmid is toxic to *E.coli* top10, which selects for any *E.coli* with the original plasmid. p.SpliceExpress also contains the ampicillin resistance gene allowing it to grow on the ampicillin containing plates.

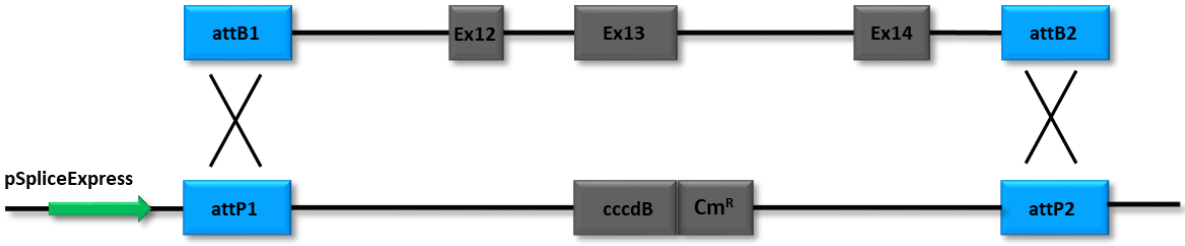


Figure 2.3 Gateway cloning uses attB/attP recombination sites to insert region of interest into p.SpliceExpress vector.

2.1.7 Expression level analysis:

2.1.7.1 RNA extraction

Cells lysates were homogenised by passing through a blunt 20-gauge (0.9 mm) needle 5-10 times. RNA was then extracted using a Qiagen RNA mini kit (Bio-Strategy) following the manufacturer's instructions. cDNA was synthesised (2.1.7.2) from the extracted RNA immediately. Any remaining RNA was stored at -80 °C.

2.1.7.2 cDNA synthesis

cDNA was synthesised using SuperScript™ IV VILO™ Master Mix with ezDNase™ Enzyme (Thermo Scientific) following the manufacturer's instructions. cDNA was stored at -20 °C.

2.1.7.3 RT-qPCR

CDC45 expression levels were measured by relative quantification from RT-qPCR, using PowerUp™ SYBR™ Green Master Mix (Thermo Scientific) following the manufacturer's instructions, using the Roche LightCycler® 480 Instrument II.

UDG activation	50 °C for 2 minutes
Dual-lock DNA polymerase	95 °C for 2 minutes
Amplification	95 °C for 15 seconds,
	55 °C for 15 seconds,
	72 °C for 1 minute
	x 40 cycles
Dissociation curve	95 °C for 15 seconds, 60 °C for 15 seconds, 95 °C continuous
Cooling	40 °C for 30 seconds

Standard curves were performed for all primer pairs (appendix C.1) and primer efficiencies were calculated to be between 92-109%.

CDC45 expression was normalised to Hypoxanthine-guanine phosphoribosyltransferase (*HPRT*) and Ribosomal Protein S18 (*RPS18*) levels as housekeeping genes, although after investigation it was decided to only use *RPS18* as *HPRT* expression in fibroblasts is variable (Nielsen et al., 2018).

Expression was compared using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001).

2.1.7.4 **RT-PCR**

RT-PCR was set up using DreamTaq a described (2.1.2.1) using cDNA synthesised above (2.1.7.2) and a standard PCR cycle (2.1.2.4.1).

2.1.8 Protein Analysis

2.1.8.1 **Protein Extraction and quantification**

A western blot was performed on fibroblast cell lysate to measure protein levels. KDM14.9 (healthy control) and patient (LB192) fibroblasts, as well as transfected (Section 2.3.7) and untransfected HEK293FT cells were used for western blots (2.1.8.3). Protein was extracted using 1x cOmplete™ Mini, EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich) in RIPA buffer (Thermo Scientific) on ice for 10 minutes (100 uL per 5.0 x 10^5 cells). Samples were then sonicated (20% Amp, 30 seconds) to disrupt the cell membranes and release the cells contents.

Protein levels were measured using Pierce Rapid Gold BCA Protein Assay kit (Thermo Scientific) following the manufacturer's instructions together with a BSA standard curve.

2.1.8.2 **SDS-PAGE**

40-80 µg protein in RIPA buffer was added to Laemmli buffer (containing 5% β-mercaptoethanol), and incubated at 95 °C for 5 minutes to denature protein. Samples were stored at -20 °C until needed. A 10 % mini protein SDS-PAGE gel and stacking gel was prepared (appendix B.2), upper and lower reservoirs were filled with SDS-PAGE

running buffer: 0.1 % SDS, 1x western gel buffer. Protein samples along with 10 µL Precision Plus Protein Dual Colour Standard (Bio-Rad) were loaded, and electrophoresis was carried out at 200 V for ~1 hour.

2.1.8.3 **Immunoblotting**

A semidry transfer was performed, the gel was placed onto a piece of 0.2 µm nitrocellulose membrane layered between 9 pieces of Whatman paper and western transfer buffer (10% iso-propanol, 1x western buffer (Appendix A.2)) using a Bio-Rad Trans-blot Turbo transfer system.

1. 12 V – 1.0 A – 10 minutes
2. 25 V – 1.6 A – 45 minutes

Membranes were blocked with 5% milk powder in PBS (Appendix A.2), with shaking at room temperature for 10 minutes. The membrane was probed with mouse anti-CDC45 (Bio-Strategy cat. no. sc-55569), diluted 1:50, 1:100 or 1:200 and incubated with shaking for 1.5 hours at room temperature or 4 °C overnight. Nitrocellulose membranes were rinsed twice before being probed with IRDye 800CW goat anti-mouse secondary antibody diluted 1:25,000 (Millenium Science NZ) for 1 hour at room temperature (shaking). The membrane was washed three times in PBS, dried in the dark and imaged on a LI-COR Odyssey CLx instrument. The image was viewed and edited using Image Studio software.

2.2 Cell culture Media and reagents

2.2.1 hTERT-RPE-1 media (Complete)

hTERT-RPE-1 cells were grown in complete growth media (CGM): 10% fetal bovine serum (FBS) (Morgate Biotech), 1% penicillin-streptomycin (PenStrep) (Thermo Scientific), 0.01 mg/mL Hygromycin B (Sigma-Aldrich) in Dulbecco's modified Eagle medium/nutrient mixture F-12 (DMEM/F12) 1:1, with 2.50 mM L-glutamine, 15 mM HEPES (Thermo Scientific or Hyclone).

2.2.2 Fibroblast Media (Complete)

Primary fibroblasts were grown in DMEM (Thermo Scientific) supplemented with 10% FBS, 1x penicillin/streptomycin (Thermo Scientific).

LB128 CDC45 patient fibroblasts were grown in DMEM (Thermo Scientific) supplemented 20% FBS, 1% penicillin-streptomycin and Fungizone (1:100 Dilution Factor) due to fungal contamination on arrival. KDM14.9 control fibroblasts grown at the same time were also grown with this medium.

2.2.3 HeLa/ HEK293F media (complete)

HeLa and HEK293FT were growth in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% FBS (Morgate Biotech), 1% penicillin-streptomycin (Thermo Scientific).

2.2.4 Freezing mix hTERT-RPE-1 and Fibroblasts

Complete growth media as above 2.2.1/2.2.2 supplemented with 10% DMSO (Sigma-Aldrich) and 60% FBS.

2.2.5 HeLa and HEK293FT freezing mix

Complete growth media as above 2.2.3 supplemented with 10% DMSO and 10% FBS.

2.2.6 Transfection media

All cell types were transfected in required media (Table 2.2) supplemented with 10% FBS (no antibiotics). HeLa and HEK293FT cells were transfected using Lipofectamine 2000 (Thermo Scientific), hTERT-RPE-1 cells were transfected using Lipofectamine 3000 (Thermo Scientific).

2.2.7 MACS buffer

MACS buffer was generated in the clean hood from autoclaved and EDTA stocks. FBS was added to give a final concentration of 1x PBS pH 7.2, 2 mM EDTA, FBS 1%. The buffer was filter sterilised (0.2 μm) into sterile 50 mL falcon tubes and stored at 4 °C.

2.3 Cell culture methods

2.3.1 Cell Lines

Aseptic technique was used to maintain a sterile environment in a Class II Microbiological Safety Cabinet. All cell lines were grown at 37 °C in a humidified incubator 5% CO_2. Frozen hTERT-RPE-1 and fibroblasts were stored long term in the liquid nitrogen Dewar, while HeLa and HEK293FT cells were stored in the -80 °C freezer. Cell lines were grown in different media conditions and centrifuged at different speeds outlined in Table 2.2.

Patient fibroblasts were derived from skin biopsies, collected and established in local diagnostic laboratories, using standard protocols.

Table 2.2 Cell line sources, media, centrifuge speeds and long term storage.

Cell line	Media	Centrifuge speed	Long term Storage	Source
hTERT-RPE-1	DMEM/F12	150 g-300 g	Liquid nitrogen Dewar	In Vitro Technologies (ATCCRL4000)
HeLa	DMEM	1000 g	-80 °C freezer	Robertson Laboratory
HEK293FT	DMEM	400 rpm (32 g)	-80 °C freezer	Robertson Laboratory
KDM14.9 (control Fibroblasts)	DMEM	300 rpm (18 g)	Liquid nitrogen Dewar	Robertson Laboratory
LB128 (DONSON patient Fibroblasts)	DMEM	300 rpm (18 g)	Liquid nitrogen Dewar	USA
LB192 (CDC45 patient Fibroblast*)	DMEM	300 rpm (18 g)	Liquid nitrogen Dewar	Canada

* note fungal contamination was present in CDC45 patient fibroblast on arrival

2.3.2 Initiating cell culture from frozen stock

One mL cryogenic vials were removed from liquid nitrogen or the -80 °C freezer and transported on ice to a 37 °C water bath. Cells were rapidly warmed for 2 minutes by gentle agitation in the 37 °C water bath before being added to 9 mL CGM (2.2.1, 2.2.2, 2.2.3). Cells were pelleted by centrifugation at the appropriate speed (Table 2.2), supernatant was carefully aspirated to remove any DMSO, and cells gently resuspended in 6 mL CGM before being plated in a T25 flask.

2.3.3 Cell maintancence

Cells were visualised using the inverted cell culture room microscope daily to check confluency and health. Culture media was changed every three days until cells had grown in a monolayer reaching 60-95% confluency. To passage cells CGM was removed and 0.6 mL 0.25% Trypsin-EDTA (Gibco) was added to detach the cells (*cells were only washed with sterile PBS when passaging from a 96 well plate). After 5-10 minutes, the cells were visualised under the microscope to check all had detached, 19.4 mL CGM was added and cells gently resuspended to give a 1:3 dilution. Cells were transferred to a T75 flask.

The above protocol was also followed for passaging cells in T75 flasks (2 mL trypsin). hTERT-RPE-1 and HeLa cells were passaged using a dilution between 1:2-1:20, while HEK293FT and primary fibroblasts were split between 1:2-1:6.

2.3.4 Cyropreservation

Cells were detached as previously described and resuspended in 10 mL CGM, centrifuged at the recommended speed (Table 2.2) and supernatant removed. Cells from an 80% confluent T75 flask were resuspended in 4-20 mL freezing mix (2.2.4/2.2.5) and 1 mL aliquots were transferred to cryovials. Primary fibroblasts were placed in a Mr Frosty (Thermo Scientific) freezing vessel in the -80 °C freezer overnight to slowly cool the cells before being transferred to the liquid nitrogen Dewar. hTERT-RPE-1 cells were transferred directly to the liquid nitrogen Dewar, while HeLa and HEK293FT cells were placed directly in the -80 °C freezer.

2.3.5 Cell counting

Cells were counted manually using a haemocytometer before freezing and transfections to confirm appropriate number of cells (Figure 2.4).

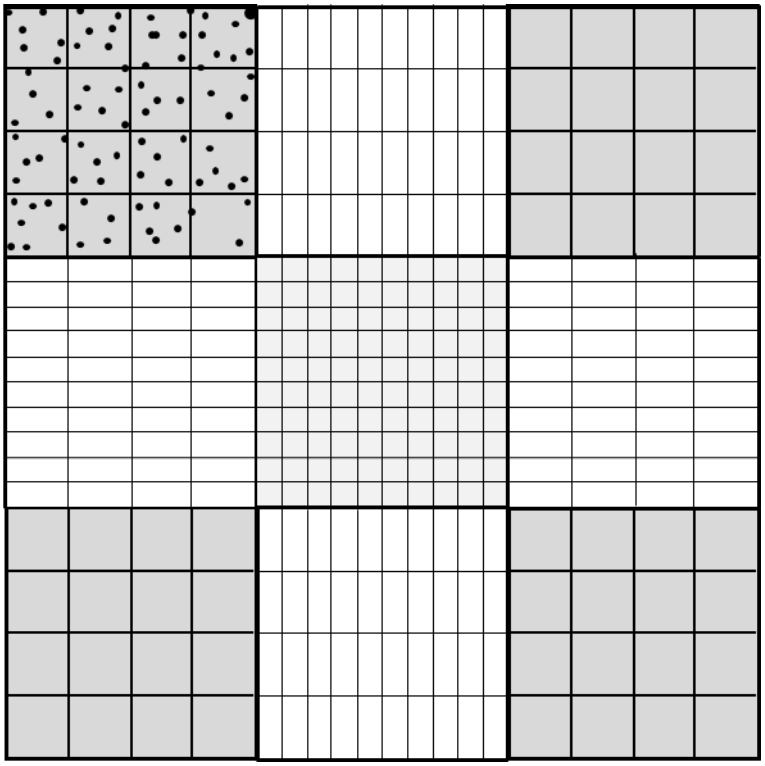


Figure 2.4 Haemocytometer used for counting the number of cells in a cell suspension. Grey squares indicate 1x1 mm grids. The number of cells (represented by black circles) in each shaded grid were counted, and an average obtained by dividing the total number of cells by 5. The average was then multiplied by the dilution factor (if diluted) x 10,000 to get to the total number of cells per mL. e.g. (y+y+y+y+y)/5 x DF x 10, 000 = cells/mL

2.3.6 Transfection of HeLa cells

Transfections were performed with Lipofectamine 2000 (Thermo Scientific). Cells were plated on glass 16 mm coverslips in a 24 well plate 8-24 hours before transfection to allow cells time to attach. Media was replaced with transfection media immediately before transfection (2.2.6). Three µL of 1 mg/mL Lipofectamine 2000 (Thermo Scientific) was added to 50 µL OptiMEM (Thermo Scientific) in tube A, while 1000 ng plasmid DNA was mixed with another 50 µL OptiMEM in a tube B. The two tubes were combined and added to the cells. Cells were incubated for 24 hours (37 °C, 5% CO_2).

2.3.7 Transfection of HEK293FT Cells

HEK293FT cells were transfected with Lipofectamine 2000. Cells were plated at 90% confluence in a 6 well plate at the time of transfection. Sixteen µL of 1 mg/mL Lipofectamine 2000 was added to 200 µL of OptiMEM. Four thousand ng plasmid DNA was added to 200 µL OptiMEM in a separate tube. The two tubes were then combined and added to cells, which were incubated for 24 hours (37 °C, 5% CO_2).

2.3.8 Transfections of hTERT-RPE-1 cells

Transfections were performed with Lipofectamine 3000. Fifty µL OptiMEM was combined with 1 µL Lipofectamine 3000, 1 µL P3000, between 250-1000 ng plasmid DNA and 0-4 ng ssODN in an Eppendorf. Tubes were centrifuged briefly before being added to hTERT-RPE-1 cells that were 70-90% confluent in a 24 well plate. Cells were plated 8-24 hours before transfection in antibiotic free media (2.2.6) to allow time for cells to adhere to the plates. Cells were visualised 24-48 hours post transfection.

2.3.9 FACS preparation

Cells were visualised under the fluorescence microscope to check for GFP expression 24 or 48 hours post transfection. Cells were trypsinized, resuspended in CGM (2.3.3) and strained using a 70 µM cell strainer. Cells were centrifuged at 300 g for 5 minutes before being washed with MACS buffer to reduce any background fluorescence. Cells were centrifuged again and resuspended in 100 µL MACS buffer to be stained using Zombie NIR fixable viability kit (MediRay). This stain causes dead cells to fluoresce so they can be sorted out. Cells were washed again before being resuspended in 500 µL MACS buffer and transported on ice to the FACS suite 4th floor Microbiology. Spare MACS buffer was taken to dilute samples if required along with 15 mL Falcon tubes containing 5 mL CGM to collect cells.

2.3.10 Flow Cytometry

FACS was performed with the help of Michelle Wilson (Department of Microbiology and Immunology, University of Otago) on the BD FACS Aria instrument. A GFP positive bulk

cell sort was performed 24-48 hours after transfection with PX458 or PX458-mSA plasmids. A GFP negative single cell sort was performed 6 or 7 days after the bulk sort. Bulk sorted cells were collected in 15 mL Falcon tubes containing 5 mL CGM, which were transported back to the laboratory, centrifuged at 300 g before being resuspended in fresh CGM and plated or used for DNA extraction. For single cell sorts, U-bottom 96 well plates containing 100 µL CGM per well were transported on ice. Cells were not sorted into the outside wells as these are prone to drying out. Ninety-six well plates were transported back to the laboratory on ice and placed in the cell culture incubator.

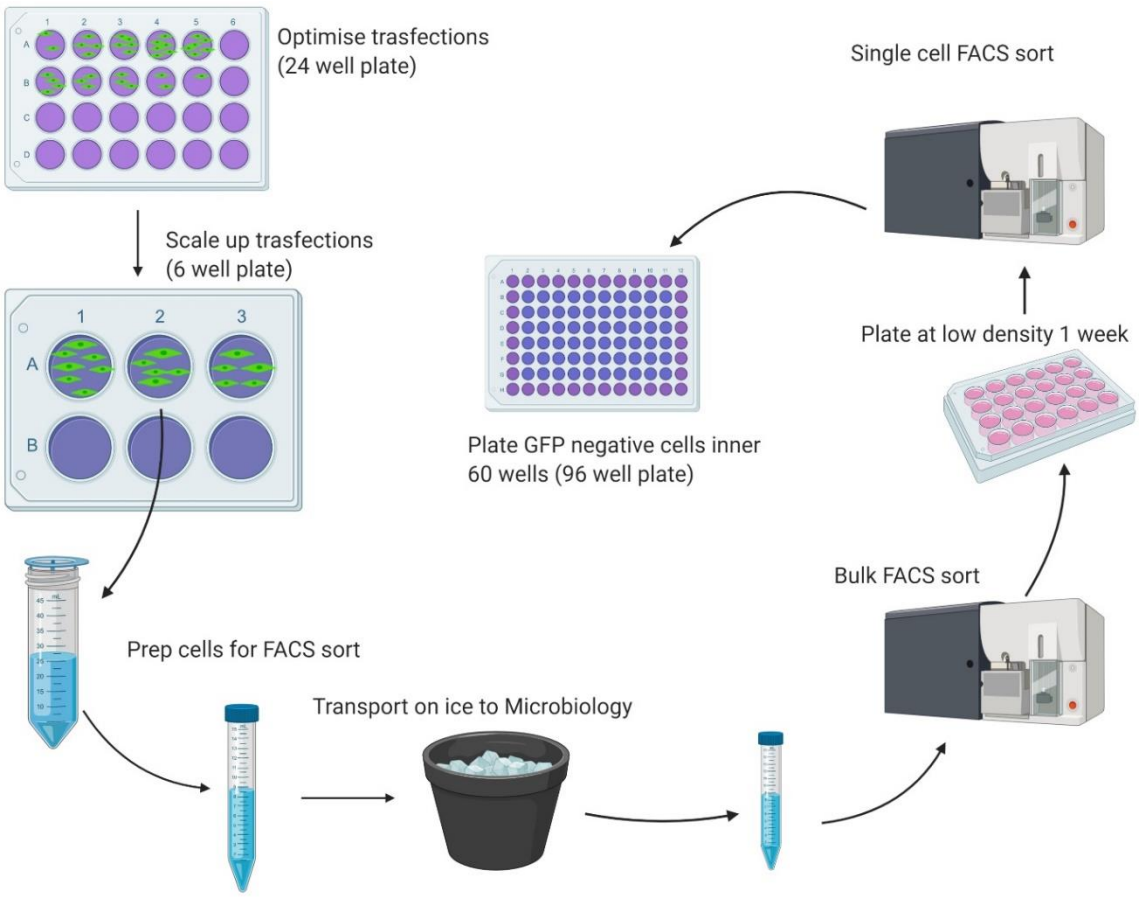


Figure 2.5 **General FACS prep sorting protocol for CRISPR-edited cells.** Transfections optimised in 24 well plate before being scaled up to 6 well plates. Cells were trypsinized, resuspended, strained, washed and stained with Zombie NIR viability dye. GFP-positive Zombie dye negative cells were bulk sorted, plated at a low density and single cell sorted ~1 week later (Figure created with BioRender).

2.4 Immunofluorescence and DNA fibrecombing

2.4.1 Immunocytochemistry

HeLa cells were plated on glass coverslips ~12 hours before transfection (2.3.6). Medium was removed, cells washed in PBS and fixed using ice cold methanol. Cells were blocked with PBS containing 2-7% serum and 0.01-3% triton X-100 (Appendix A.2). If using primary antibodies (anti-phospho-histone H2A.X (Ser139) (Merk Millipore), anti-phospho histone H3 (serine 10) (Merk Millipore), or purified anti-active caspase-3 monoclonal antibody (rabbit IgG) (BD Pharmingen)) cells were incubated in (1:500-1:1500) overnight at 4 °C. Goat anti-mouse IgG (H+L) cross adsorbed secondary antibody Alexa Fluor 568 (Thermo Scientific) or Goat anti-rabbit IgG (H+L) cross adsorbed secondary antibody, Alexa Fluor 568 (Thermo Scientific) were mixed with blocking solution (1:500) and incubated for 1 hour at room temperature. DAPI (Sigma-Aldrich) (1:1500) was then added for 5 minutes. Cells were washed before being mounted with ProLong™ Diamond Antifade Mountant (Invitrogen). Further details available Appendix B.4.

2.4.2 Quantification of DNA replication events and replication fork speed using DNA Fibre combing

Fibroblasts were plated at ~40% confluent the day before. Twenty-five μM 5-chloro-2'-deoxyuridine (CIdU) (Sigma-Aldrich) diluted in CGM (2.2.2) was added to the cells for 20 minutes. Cells were washed in PBS before 250 μM 5-Iodo-2′-deoxyuridine (IdU) (SIGMA - I7125-5 g) was added for 20 minutes. Cells were trypsinized and resuspended in ice cold PBS. Cells were counted (2.3.5) and diluted to the optimal 50 x 10^4 cells/mL concentration. Two μL (1000 cells) were spotted on a clean slide labelled with a pencil and left to become crinkly around the edges. Seven μL spreading buffer (200 mM Tris pH 7.5, 50 mM EDTA, 0.5% SDS) was spotted on top, swirled together with pipette tip and incubated for 2 minutes at room temperature. Slides were tilted to start drop moving down the slide and left at a smaller angle (PCR tube rack) to slowly spread DNA down the slide. Slides were placed in MeOH/ AcOH (3:1) for at least 10 minutes to fix DNA to the slide and de-proteinase the sample. Slides were stored 4 °C (for up to a year).

A PAP pen was used to draw a small square at the top of one of the slides, DNA was spotted into this square as the large amount of DNA in a small area was used for focusing the microscope.

DNA was denatured by adding freshly diluted 2.5 M HCl (Thermo Fisher) to the slides for 1 hour 15 minutes. This makes the ssDNA accessible to the antibody. DNA was blocked with 1% BSA, 0.1% Tween 20, in PBS for 1 hour. One hundred and fifteen µL primary antibodies (1:750 Rat anti-BrdU (anti-CIdU) (Abcam) and 1:750 Mouse anti-BrdU (anti-IdU) (BD Biosciences)) were added to slides and held in place with a long coverslip, slides were incubated at 4 °C overnight. Slides were fixed with 500 µL 4% paraformaldehyde (Sigma-Aldrich) for 10 minutes. This fixes the antibody to the antigen. Slides were washed before 115 µl secondary antibodies (1:500 goat anti-rat Alexa Fluor 594 and 1:500 goat anti-mouse Alexa Fluor 488) was added to the slides and incubated for 1.5 hours at room temperature. Slides were washed and mounted using Fluoroshield histology mounting medium (Sigma-Aldrich) and then stored at -20 °C. Detailed protocol available Appendix B.5.

2.4.3 Microscopy

Transfected cell cultures were visualised using the Olympus IX71 inverted light microscope, both bright field and GFP (488 nm) images were taken.

Slides were visualised using the Olympus BX53 upright microscope. Images were taken using the GFP (488 nm) channel for cells stained with Alexa Fluor 488 or transfected with GFP expressing plasmid (excitation/emission = 499 nm/520 nm), 594 nm (excitation/emission = 590 nm/618 nm) and 568 nm (excitation/emission = 579 nm/603 nm) were used for respective fluorophores and the UV channel was used for DAPI fluorescence (excitation/emission = 359 nm/457 nm). Channels were overlaid using CellSens Dimensions software.

2.4.4 Quantifying images

HeLa immunofluorescence (GFP-DONSON localisation), GFP and DAPI images were taken at 40x or 100x magnification. GFP and DAPI images were overlaid using CellSens software,

cells were classified as diffuse, nuclear or aggregate by eye. Approximately 1200 cells across 5 experiments were counted for each plasmid.

Fibre combing images were taken using the 40x or 100x lens. The 488 and 594 channels were combined and replication events were quantified. Tract lengths were measured using ImageJ software. The length of each tract in pixels was converted to μm using the scale bar on the images, μm was converted to kb based on the common conversion frequency 2.59 kb/1 μm (Daigaku et al., 2010; Jackson and Pombo, 1998; Quinet et al., 2017). The length of tract in kb was divided by pulse time (20 minutes) to give replication fork speed in kb/min. The number of replication events, speed of fork progression and bi directional forks were measured.

3 CRISPR editing to introduce patient variants

3.1 Introduction

The aim for this chapter was to explore whether CRISPR-Cas9 can be used to knock in patient-specific MGS variants in hTERT-RPE-1 cells, disrupting essential DNA replication genes.

CRISPR (clustered regularly interspaced short palindromic repeats) Cas9 genome editing is a new widely used technology that can be used to knockout genes or introduce specific variants into an organism's genome (Adli, 2018; Rozov et al., 2019). This precise technology allows modelling of patient cells when patient cells are unavailable. CRISPR editing also allows creation of cell lines containing only the suspected disease-causing variant, controlling for background variation in a way that would not be possible using primary cells (Royba et al., 2017). CRISPR is commonly used for creating small insertions or deletions, preventing normal protein translation and therefore 'knocking out' a gene, but it is less efficient at knocking in specific variants (Mao et al., 2008; Rozov et al., 2019). This is because knocking in variants has a much lower efficiency than classical knockouts, due to the time and likelihood of the cell repairing the lesion via homology directed repair (HDR) (Mao et al., 2008). After the Cas9 protein has made a double-stranded break in the DNA the cell has two main options for repairing the DNA; error prone non-homologous end joining (NHEJ) or HDR (Mao et al., 2008). NHEJ is the predominant option, as it is active at all points in the cell cycle (Maruyama et al., 2015; Sargent et al., 1997), religating the phosphate ester backbone of two DNA molecules. Processing of the free ends of the DNA molecules can occur as part of this repair, causing removal or insertion of bases (Lieber, 2010). Repair of the Cas9-mediated cleavage site by this pathway leads to insertions or deletions ('indels') at the cut site.

The second pathway is the HDR pathway, where the cell repairs the double-stranded break by copying an available template, either the other copy of the gene or a single-stranded oligonucleotide introduced to insert a specific variant (Horii and Hatada, 2016). Some efficiency problems occur with this method in mammalian cells, as error prone end joining pathways are almost always favoured over homology-directed repair mechanisms (Savic et al., 2018). This is because HDR is more likely to occur during late S/G2-phase as the cell will already have the appropriate machinery active (Symington and Gautier, 2011;

Yao et al., 2017). HDR is more accurate but more time consuming than NHEJ and therefore is less likely to occur unless the cell is already in S-phase (Mao et al., 2008).

Introducing a single missense variant, with CRISPR-Cas9 allows the ability to explore the effect this single bp change has on the protein and pathways, as well as the following consequences this has on DNA replication and cell cycle progression. The ability to use CRISPR Cas-9 to knock in patient variants in established diploid cell lines such as hTERT-RPE-1 would allow functional analysis of the candidate genetic variation an isogenic system. Using patient-derived cells is problematic because individual background genetic variation may affect results or interpretation.

hTERT-RPE-1 cells were chosen for this project as this cell line is near diploid (90% of counted cells contained 46 chromosomes) (https://www.atcc.org), grows in a monolayer and contains one normal and one derivative X chromosome. The derivative X chromosome contains additional chromosomal material at the terminal end of the q-arm, including telomerase reverse transcriptase (hTERT) and Hygromycin B resistance which allows the cell line to grow for at least 15 passages after cryopreservation. These cells represent a strong option for mimicking diploid patient cells, and are a standard cell model for studying DNA replication and DNA damage response (Bielinsky et al., 2019; Ercilla et al., 2016; Feringa et al., 2018; Kuznetsova et al., 2014; Lahtela et al., 2013; Shimada et al., 2017; Sokolova et al., 2017; Sun et al., 2017). hTERT-RPE-1 cells grow well in culture and have successfully been used previously for CRISPR editing (Bodnar et al., 1998; Jiang et al., 1999; Katoh et al., 2017; Kuznetsova et al., 2014; Yiu et al., 2016).

CRISPR utilises a repair system found in the adaptive immune system of bacteria (Cong et al., 2013; Mali et al., 2013a). Although CRISPR can be used to introduce specific variants, this can only be done within close proximity to a PAM site (Paquet et al., 2016), a 3 bp motif recognised by the Cas protein, NGG for Cas9 (Mali et al., 2013a). CRISPR works by creating a double-stranded break in the DNA at a PAM site, the protein is guided to the specific PAM site by a single-stranded guide RNA (sgRNA). The gene specific part of the sgRNA is ~20 bp long, the rest of the RNA forms a hairpin like structure that interacts with the Cas protein.

CRISPR has a variable success rate and has been used in a variety of organisms with a variety of different Cas proteins and methods (Peng et al., 2016; Yu et al., 2018). Editing efficiency also seems to vary widely between areas of the genome (Moon et al., 2019), as

well as having different efficiencies in targeting between individual sgRNAs. Because of this there appears to be little consensus on the most optimal methods to induce HDR in diploid cell lines.

There are many tools designed to help select the best sgRNA for a specific variant as the Cas protein needs to cut as close to desired variant as possible (Paquet et al., 2016) but the sgRNA must be specific and direct the enzyme cut only at this site, not other similar PAM sites.

This project sought to introduce patient variants in the hTERT-RPE-1 cell line. The target genes are essential in the cell, due to the role of the encoded proteins in DNA replication machinery required for proliferation, so biallelic knockouts of these genes are not tolerated. Patient variants are hypomorphic, where they reduce protein function to a lower level but still retain a minimum threshold of functionality to permit cell survival and proliferation. Because of this the less efficient homology directed repair pathway was chosen to try and introduce a single base pair patient variants into the cell lines.

3.2 Patient variants

3.2.1 Novel *ORC1* variant (c. 1865T>C, p.L622P)

The two patient variants chosen to investigate in depth by CRISPR knock in editing in hTERT-RPE-1 cells and cellular assays were both novel MGS variants. The first, LB005 (*ORC1* c. 1865T>C, p.L622P) was found to have a homozygous missense variant in an area of ORC1 not previously associated with MGS. Dr Karen Knapp identified this variant through whole genome sequencing. LB005 presented at the clinic age six, born to parents from the same village in Saudi Arabia, so likely distantly related. He had a proportional reduction in size; head size (OFC -4.1 SD), height (-4.2 SD), weight (-4.6 SD). Absent patella were confirmed by palpitation as patella had not fully ossified to be visualised by X-ray at this point. He also presented with micrognathia (small jaw), very small misshapen ears and a thin face, very typical of MGS. LB005 had otherwise normal development.

Previous MGS patients with variants in ORC1 have all been found to have at least one variant in the BAH domain. This domain is required for important protein-protein interactions essential for ORC1 chromatin binding. Previous research by Kuo et al., 2012, suggested the *ORC1* MGS phenotype was caused by the inability for ORC1 to interact with

H4K20me2. H4K20me2 is a common histone 4 modification present at origins of replication and though to be involved in stabilising the ORC complex at these origins (van Nuland and Gozani, 2016). In this research Kuo et al., 2012 demonstrated that ORC1 depletion in zebrafish leads to a MGS-like phenotype; zebrafish depleted in H4K20me2 also showed a similar reduction in size phenotype. This suggests the MGS phenotype may be caused by the inability of ORC1 to interact with H4K20me2 on chromatin. In this study zebrafish were rescued by the addition of WT human ORC1 but were not rescued by the addition of human ORC1 containing MGS variants in the BAH domain. This research suggests the BAH domain plays an important role in the MGS phenotype in patients with variants in this region. But this does not explain the same *ORC1*-MGS phenotype being seen in patients with neither variant in the BAH domain. The p.L622P variant investigated in this project is found in the Walker II motif of the AAA+ domain, important for ATP hydrolysis.

This variant was chosen for CRISPR as it was a homozygous missense variant, requiring only one ssODN and one screening measure.

3.2.2 Novel *DONSON* variant (c.631C>T, p.R211C)

The second variant was also identified by whole genome sequencing carried out by Dr Karen Knapp. This variant in *DONSON* was, at the time novel. Four patients (24% of our MGS cohort) were found to have variants in *DONSON*, which suggested *DONSON* may be a relatively common novel MGS gene (Knapp et al., 2019). LB008 was chosen, firstly as the only one of four patients to contain a homozygous missense variant (c.631C>T, p.R211C), rather than compound heterozygous variants. Secondly the presentation of a very typical MGS phenotype could not be confused with the MISSLA phenotype. At the most recent examination, the patient (aged 9 years, 6 months) presented with a proportional reduction in size; head size (OFC -3.97 SD) and height (-3.48 SD). LB008 also presented with small dysplastic ears, micrognathia (small jaw) and absent patella. He also presented with delayed development, mild hearing loss, a high arched palate and hypopigmentation. LB008 of Turkey comes from consanguineous parents.

As only one sgRNA and one ssODN was required to knock in this homozygous variant it was chosen for the CRISPR experiments. It had also been mentioned that if the cell does proceed to the HDR pathway it is likely to knock in on both alleles as the cell is most likely

in S-Phase (Personal communication Professor Andrew Jackson, Institute of Genetics and Molecular Medicine, University of Edinburgh).

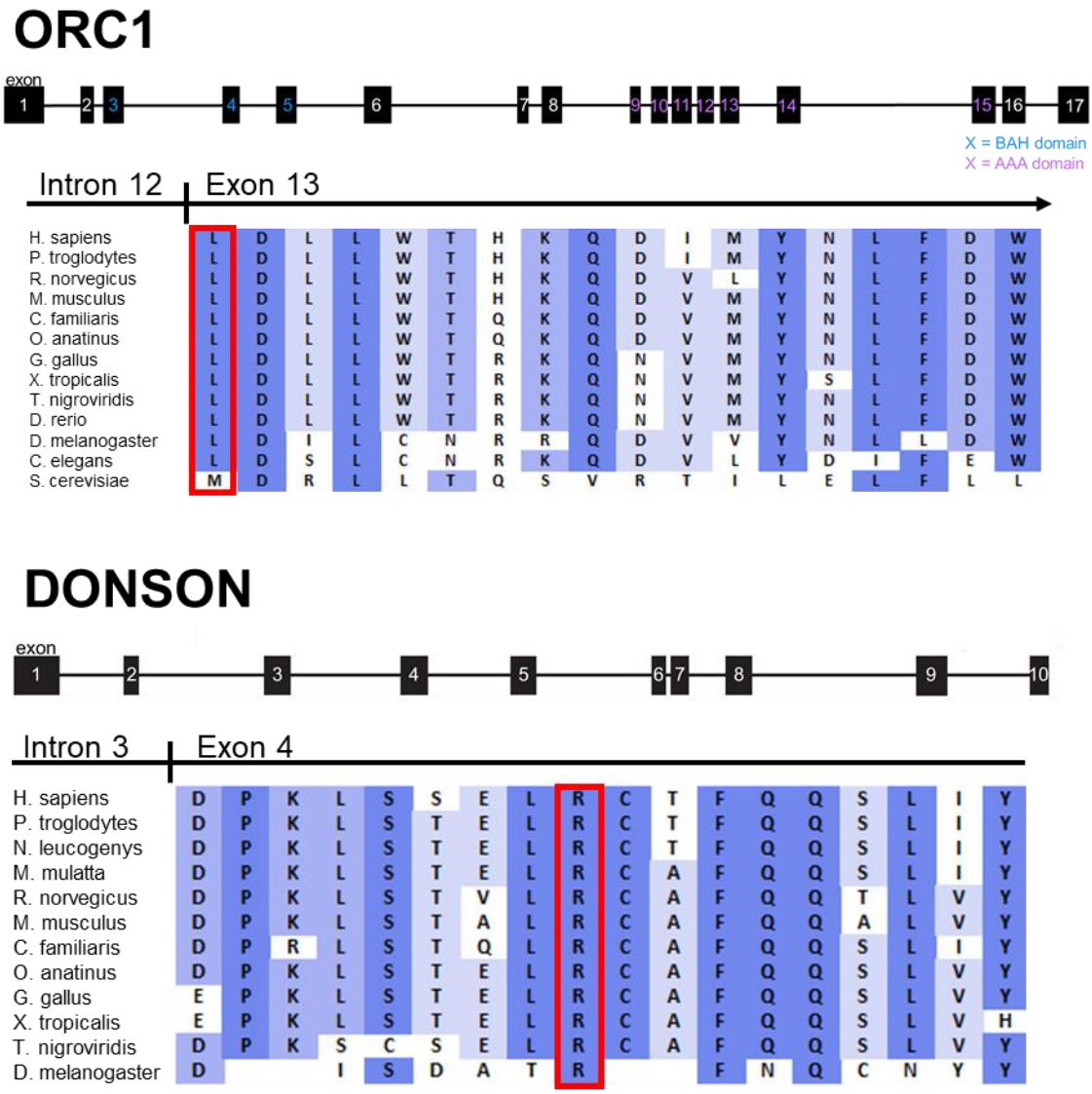


Figure 3.1 **MGS patient variants cause amino acid changes in highly conserved protein regions.** Regions of ORC1 and DONSON contain highly conserved regions between species, as these essential DNA replication genes are required in all eukaryotes. The red boxes indicate amino acid changes caused by homozygous single base pair variants found in MGS patients. *ORC1* - c. 1865T>C, p.L622P, DONSON - c.631C>T, p.R211C. Both amino acids changed are highly conserved between species and in highly conserved areas of the protein**.** Figure adapted from Alamut Visual software (BioSoftware).

Both variants are found in highly conserved regions of the protein (Figure 3.1), and both result in functionally important changes. While both leucine and proline are non-polar, proline is a helix breaker which can disrupt α helical backbone conformation (Li et al.,

1996), this therefore is likely to be a functionally important change. Arginine is a positively charged amino acid while cysteine is a thiol containing amino acid, disulfide bonds are formed between two thiol groups and therefore the addition of a cysteine has the potential to change disulfide bonds in the secondary structure of DONSON (Qiu et al., 2015). These amino acids are conserved through to *C. elegans* (ORC1) or *Drosophila melanogaster* (DONSON). This high conservation of amino acids between species suggests that these amino acids are important for protein function and change is not generally tolerated in this region.

Along with the known role of these genes, indicates these variants are highly likely to be disrupting origin licensing and potentially delaying DNA replication (Bicknell et al., 2011a).

3.3 Designing CRISPR guides and plasmid construction

3.3.1 Guide RNA design

Guide RNAs (sgRNAs) were designed using http://crispr.mit.edu/ a widely used tool at the time, which has since been retired. This tool assigns a score for each potential sgRNA based on a number of factors. The quality score takes into account off-targets in genes and non-coding regions as well as the number and position of mismatches between the sgRNA and the region. Two mismatches proximal to the PAM region will considerably reduce Cas9 activity (Anderson et al., 2015). Three or more interspaced mismatches or 5 consecutive mismatches leads to almost no off-target binding (Hsu et al., 2013). There were limited options of sgRNAs to choose from as the cut site needed to be in close proximity to the desired substitution (Komor et al., 2016; O'Brien et al., 2019; Paquet et al., 2016). The PAM site should ideally also be scrambled to prevent re-cutting by the Cas9 protein (Kwart et al., 2017; Mali et al., 2013a; Paquet et al., 2016). Other tools CHOPCHOP (Labun et al., 2019) and Benchling (Benchling [Biology Software], 2019) were used to confirm the sgRNA efficiencies were similar as the score given by different tools favour off-targets or on target efficiencies.

3.3.1.1 *ORC1* p.L622P sgRNA design

Guide 3 was chosen for *ORC1*-p.L622P for several reasons. First, the cut site was the closest to the desired variant site, maximising efficiency (Paquet et al., 2016). Second, introducing the patient variant would also result in the loss of a Taq (α) I restriction site (Figure 3.2), enabling a simple assay for genotyping cells. Third, only a single synonymous substitution was required to mutate the PAM of guide 3, while two variants would have been required to mutate the PAM site of guide 1, the guide with the highest quality score. Four guide 3 has 226 predicted off-targets but none lie within genes (Figure 3.2B).

Because the PAM site, while in the intron, is in close proximity to the 3′ splice junction of intron 12, *In silco* analysis was undertaken to assess any possible effects on canonical splicing. Alamut Visual version 2.11 (Interactive Biosoftware, Rouen, France) was used to predict any changes to splicing. Alamut is an exploration software which incorporates databases such as NCBI, EBI, UCSC, gnomAD, ESP, Cosmic and ClinVar to make predictions about pathogenic variants. It also uses SpliceSiteFinder-like, MaxEntScan, NNSPLICE, and GeneSplicer to predict changes to splicing. Alamut predicted that a change to the first nucleotide in the PAM sequence would not lead to any change in splicing. Therefore guide 3 was selected.

Figure 3.2 **sgRNA guide selection for *ORC1* p.L622P variant.** A) The top 5 sgRNAs for the region of interest, showing PAM sites (green), variant (highlighted yellow) and Cas9 cut site (dark grey), blue numbers next to guides 1-5 indicate quality score assigned by tool with 100 being the highest and 0 being the lowest. Guide 3 was selected due to the appropriate score, and the cut site being the closest to the desired patient variant. A Taq (α) 1 restriction site is also lost with introduction of the patient variant, important for screening. B) Selected output from design tool showing the selected sgRNA (guide 3) has 226 potential off-targets; none are in exons, the top 20 off-targets are shown here. All contain at least 3-4 mismatches except for two (which only contain only 2 mismatches).

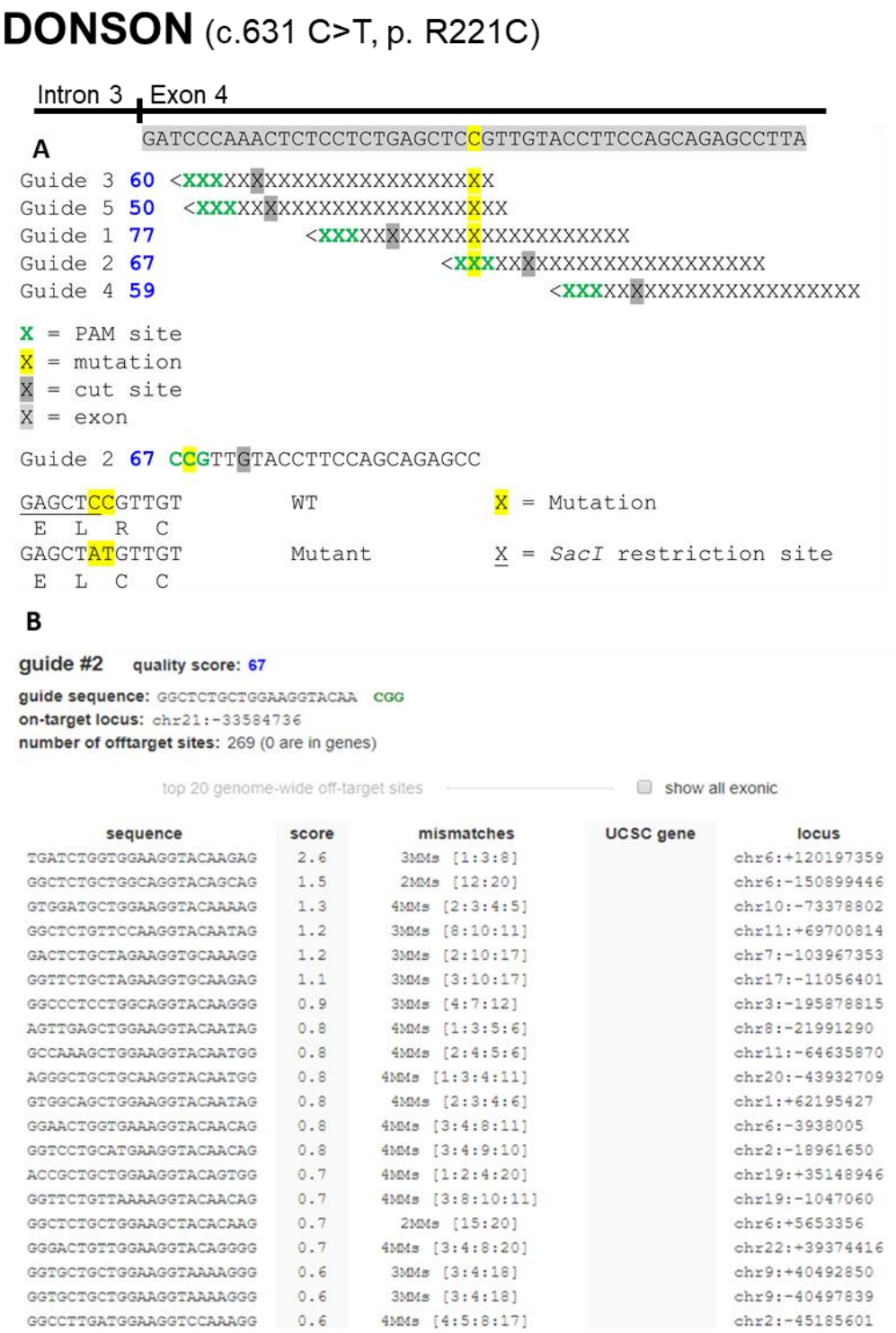


Figure 3.3 **sgRNA guide selection for p.R211C *DONSON* variant.** A) The top 5 sgRNA for the region of interest, showing PAM sites (green), variant (highlighted yellow) and Cas9 cut site (dark grey), CRISPR tool quality score indicated by blue numbers (100 highest, 0 lowest). Guide 2 was selected due to the appropriate score, and the cut site being the closest to the desired patient variant. A SacI restriction site is lost with the introduction of the desired variant and one silent variant – required for screening. B) Selected output for CRISPR design tool showing the selected sgRNA (guide 2) has 269 potential off-targets zero of which are in exons, the top 20 off-targets are shown all but one containing at least three mismatches.

Guide 2 was selected for *DONSON* p.R211C (Figure 3.3), for similar reasons as above. Guide 2 had the second highest score (67) with 269 off-targets, 0 in exons (Figure 3.3B), the cut site was also the closest of all guides for maximum efficiency. Guide two required two variants; a single synonymous substitution would lead to the deletion of a SacI restriction site for effective screening of colonies, and a second to introduce the patient variant and scramble the PAM site at the same time (Figure 3.3A). Two off-target sites contain only two interspaced mismatches (Figure 3.3B), while the other 267 off-targets contain 3-4 mismatches which considerably reduces the chances of off-target binding (Hsu et al., 2013).

3.3.2 PX458 assembly

After guides were selected, and then cloned into the PX458 vector (Addgene #48138) (Figure 3.4). PX458 was linearized at the gRNA scaffold by BbsI and NEB T4 ligase was used to join the two fragments together. To confirm the sgRNA was inserted and in the correct orientation, transformed colonies were analysed by colony PCR (cPCR) and Sanger sequencing (Figure 3.5).

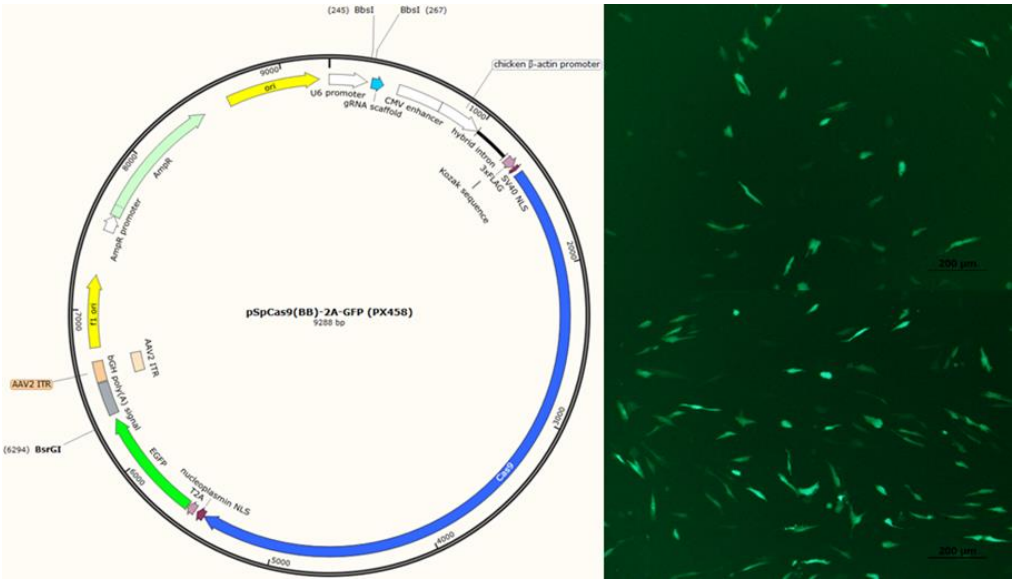


Figure 3.4 Schematic of PX458 plasmid containing Cas9-GFP complex and sgRNA scaffold. A) The PX458 plasmid contains a single sgRNA sequence under a U6 promotor to direct Cas9 to the appropriate area of the genome, a Cas9 protein for creating double-stranded breaks, and a GFP protein for optimising transfection and FACS sorting only the cells which contain the plasmid. B) Fluorescence GFP image of transfected hTERT-RPE-1 cells containing PX458 plasmid fluoresce green when imaged with Olympus IX71 inverted light microscope GFP (488) filter, allowing transfection efficiencies to be optimised by eye, as well as being appropriate to use with FACS.

The PX458 plasmid encodes a Cas9 protein fused to GFP, which is useful for sorting cells by flow cytometry to enrich for plasmid-harbouring cells, which is beneficial for cells with a low transfection efficiency. The plasmid also contains a cloning site for the sgRNA under a hU6 promoter. Having the sgRNA and Cas9 on the same plasmid increases efficiency of CRISPR as only one item needs to be introduced for the Cas9 protein to be directed to the correct site in the genome.

Following sequencing confirmation, suitable plasmid stocks were prepared following the midiprep protocol (2.1.6.5) for transfection of hTERT-RPE-1 cells. Concentrations were measured using the NanoDrop (2.1.6.6).

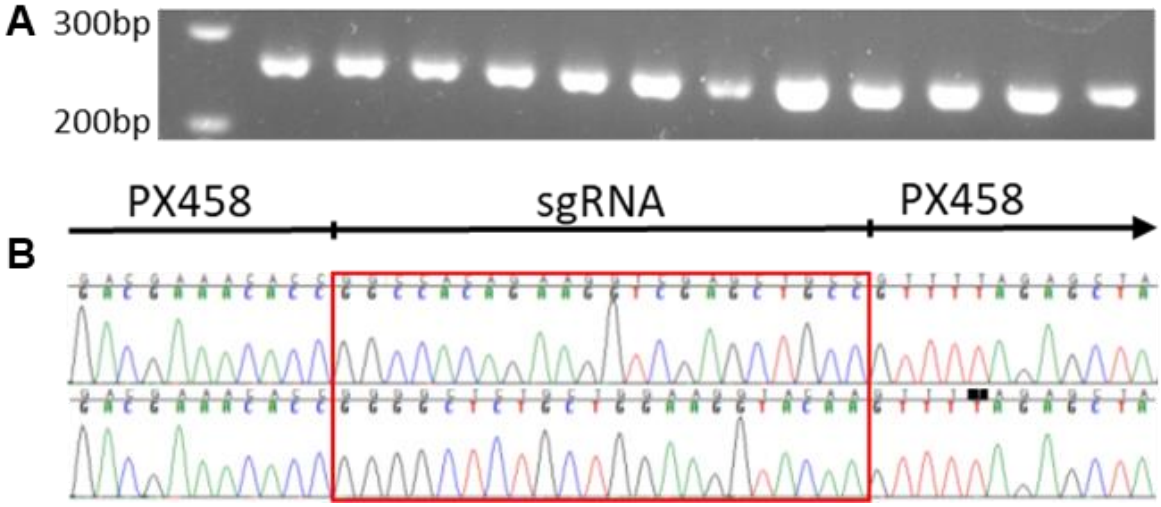


Figure 3.5 cPCR and Sanger sequencing confirmed correct sgRNA insert for both guides. A) cPCR was performed using the hU6_F plasmid specific primer and one insert specific primer (sgRNA_R). All colonies screened appeared to contain sgRNA insert. Negative control was run on a separate gel (not shown). B) Sanger sequencing confirmed correct sgRNA ORC1 (top), DONSON (bottom) in the correct orientation.

3.3.3 Repair template design

Repair templates were designed with homology arms between 80 bp and 98 bp on either side of the desired knock in variant, due to maximum length of PAGE purified oligonucleotide from IDT being 200 bp. The recommended (Yang et al., 2013) length of homology arms is at least 90 bp, some repair templates had arms slightly under this due to size constraints for PAGE Ultramer® DNA oligonucleotide (IDT) being 200 bp.

It is also recommended that the PAM site is scrambled to prevent the Cas9 protein cutting the repair template instead of the DNA, this also helps prevent any re-cutting of edited DNA by the Cas9 which could lead to indels after HDR (Paquet et al., 2016).

Both repair templates contained two variants, one to introduce the desired missense change and a second to disrupt the PAM site or disrupt a restriction digest site for easy screening.

ORC1 p.L622P repair template (Figure 3.6) contained two variants. One missense T>C variant to introduce LB005 variant resulting in a leucine (L) to proline (P) amino acid substitution and a second silent c>t variant to disrupt the PAM site 8 bp before the start of exon 12. As mentioned earlier (3.3.1) Alamut predicted this substitution would not result in any splicing changes. This silent variant also results in the loss Taq (α) 1 restriction site for screening.

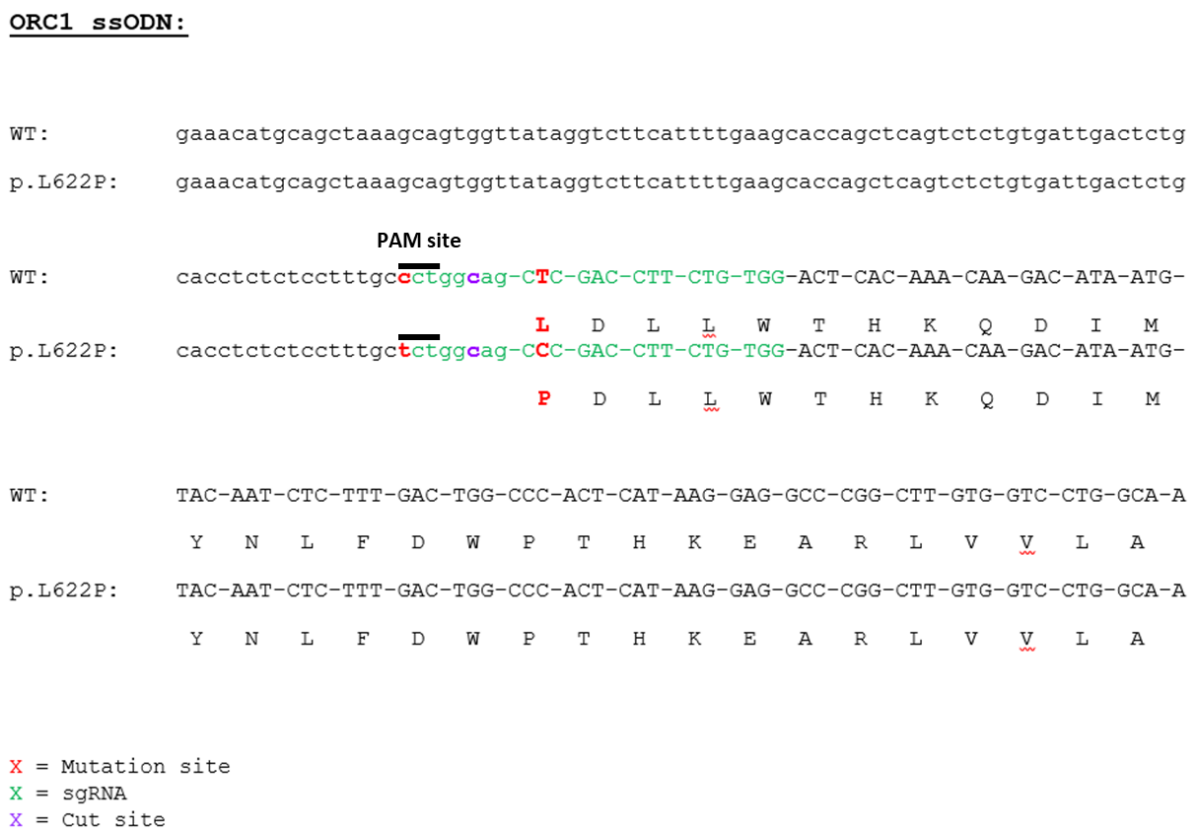


Figure 3.6 **Important features of the *ORC1* repair template.** *ORC1* ssODN contains one missense variant to introduce patient variant and one substitution mutation to disrupt the PAM site. The sgRNA is indicated by green letters and PAM site (black line) shows where the Cas9 protein will be directed to, the purple letter indicates the Cas9 cut site. The repair template contains a 98 bp 5' end arm and an 89 bp 3' end arm.

DONSON p.R221C repair template also contains two variants (Figure 3.7). The first a LB008 missense variant C>T, resulting in an arginine (R) to Cysteine (C) amino acid

substitution. The second a silent C>A resulting in no amino acid change and resulting in scrambling the PAM site. This silent variant also scrambles the SacI restriction enzyme site which allows for easy screening.

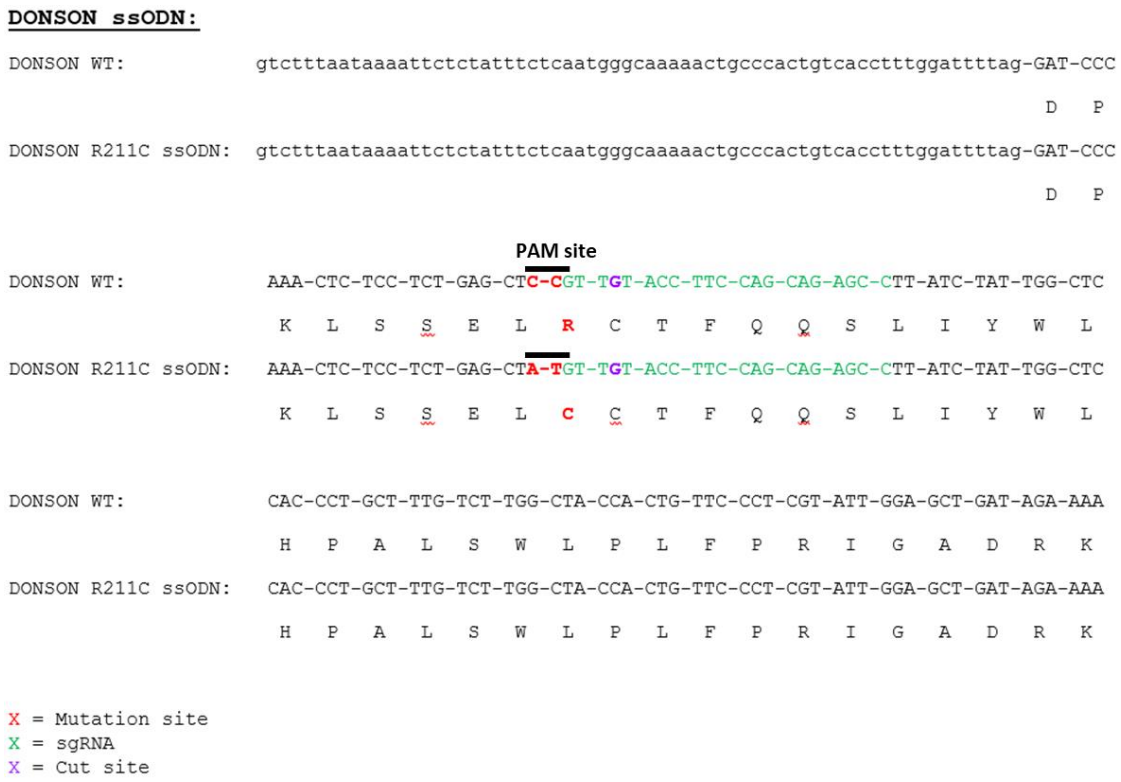


Figure 3.7 Important features of the *DONSON* repair template. *DONSON* ssODN containing 2 variants, 1 to introduce missense variant and a second to disrupt the PAM and SacI restriction digest sites. The sgRNA in green and PAM site (black line) shows where the Cas9 protein will be directed to, purple letter indicates the Cas9 cut site. The repair template contains an 88 bp 5' arm and an 89 bp 3' arm.

3.4 CRISPR protocol

Figure 3.8 depicts the general CRISPR protocol used in this project. The first step involved optimising plasmid transfections in 24 well plates as the transfection efficiency in hTERT-RPE-1 cells was relatively low, by titrating different concentrations of plasmid and examining over different time points.

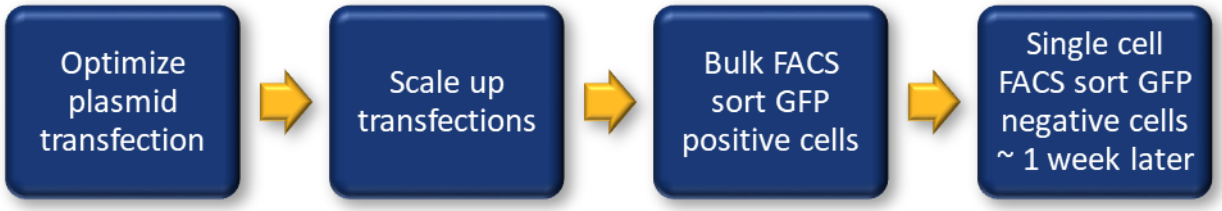


Figure 3.8 **General CRISPR protocol for transfecting and sorting cells.** Cells were transfected with 10 ng ssODN and 250-1000 ng PX458 plasmid, GFP fluorescence was observed. Optimised transfection was scaled up for bulk GFP positive FACS sort. Bulk sorted GFP positive cells were plated in 24 well plates for 6-7 days before being FACS single cell sorted, allowing time for disappearance of GFP-PX458 plasmid confirming no cells had permanently integrated the plasmid, and increased number of cells growing in 96 well plates. This was based on a protocol shared by Professor Wojciech Niedzwiedz (Weatherall Institute of Molecular Medicine, University of Oxford, personal communication).

PX458-*ORC1*-p.L622P was optimised at 1000 ng/µL 24 hours post transfection (Figure 3.9) while PX458-*DONSON*-p.R211C was optimised at 500 ng/well 24 hours after transfection (Figure 3.10). Generally FACS sorts have been performed 48 hours after transfection (Giuliano et al., 2019, 2019; Li et al., 2014; Maguire et al., 2019; Park et al., 2019), as this allows confirmation of GFP fluorescence by microscopy at 24 hours before preparing the cells for FACS at 48 hours. Many of the hTERT-RPE-1 cells had died off and were floating or were looking very unhealthy (rounding and about to detach) after 48 hours, so flow cytometry sorting at 24 hours post transfection was decided upon for optimal cell health and growth.

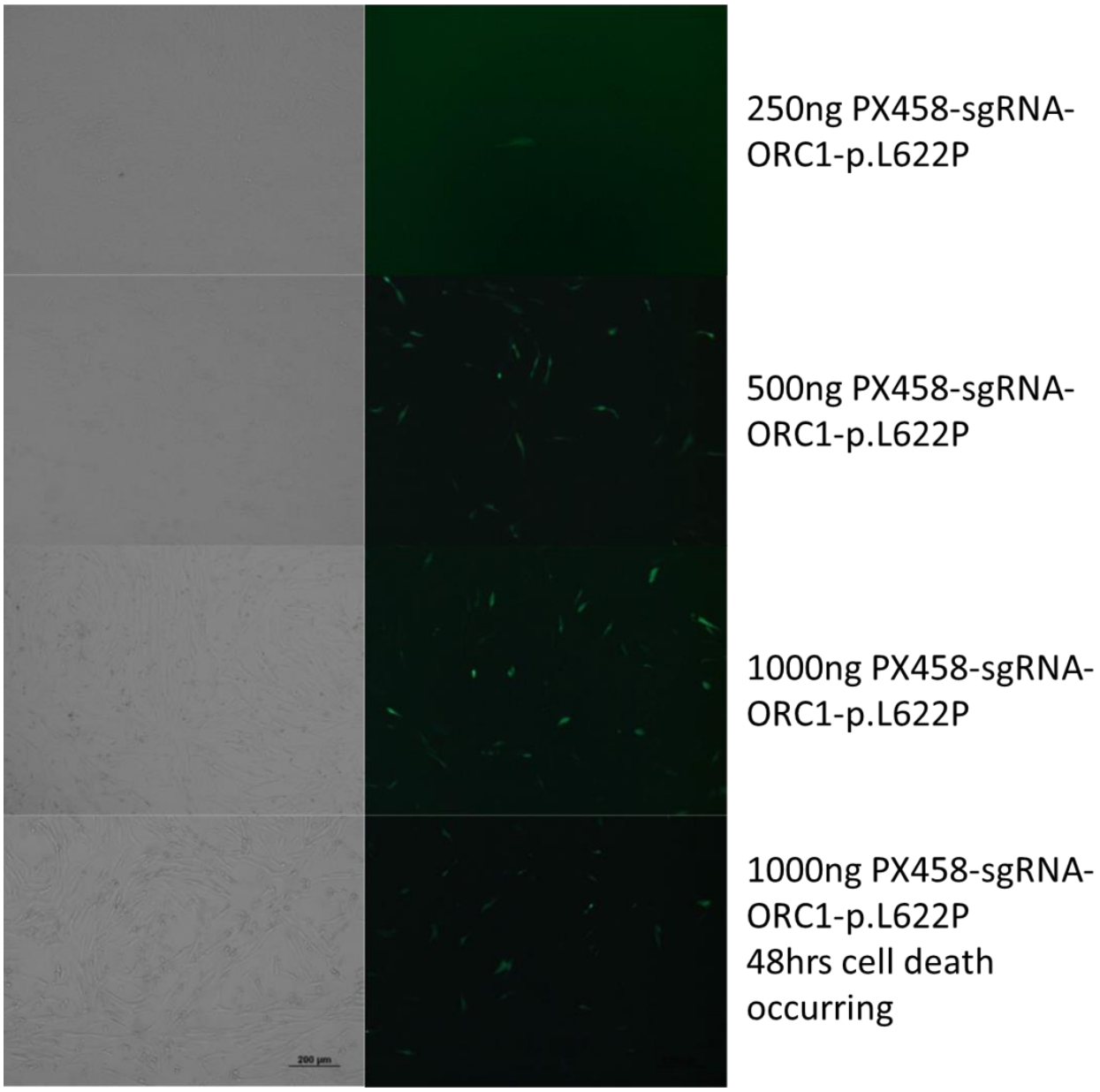


Figure 3.9 **PX458-sgRNA-*ORC1*-p.L622P transfection test.** hTERT-RPE-1 cells were transfected with 250 ng, 500 ng or 1000 ng PX458 plasmid and observed 24 & 48 hours post transfection. The optimal GFP fluorescence and cell health was observed using 1000 ng plasmid, at 24 hours post transfection (~10% transfection efficiency for 1000 ng compared to only ~2-5% for the other plasmid concentrations).

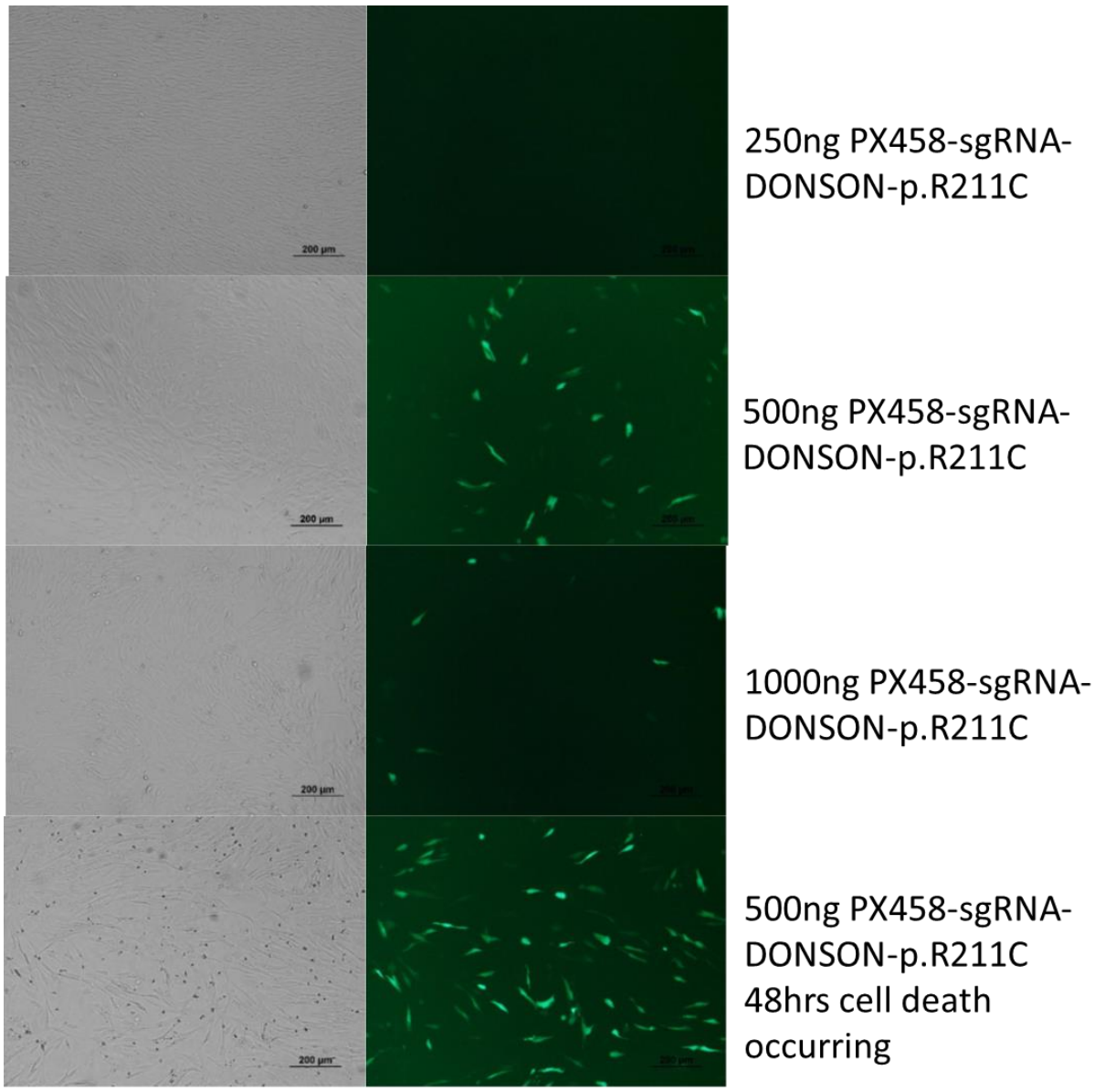


Figure 3.10 PX458-sgRNA-*DONSON*-p.R211C transfection test. Cells transfected with 500 ng PX458-*DONSON* plasmid 24 hours post transfection showed optimal GFP fluorescence (~12.5% transfection efficiency) while maintaining healthy cells. GFP fluorescence was slightly improved at 48 hours post transfection (~15% transfection efficiency) the cell health had decreased with many cells dead or dying. (Transfection efficiency for 250 ng and 1000 ng was only around 3-8%).

Once plasmid concentration using in the transfection was optimised by estimating GFP-positive cells by eye (Figure 3.9 and 3.10), the experiment was scaled up to increase cell numbers and possible chance of HDR-mediated CRISPR editing. Cells were plated in triplicate in a 6 well plate and transfected with as optimised above, with 2500 ng PX458-*DONSON* plasmid (equivalent of 500 ng in 24 well plate). Cells were collected 24 hours post transfection and prepared for FACS sorting. FACS sorting was performed by Michelle Wilson or Katie Young in the Microbiology FACS suite using the BD FACS Aria instrument.

FACS allows quantification of the transfection efficiency as all cells that pass through the laser are counted but only cells which are GFP positive and viable are sorted into a collection tube. Prior to flow laser capture, cells were stained with NIR Zombie viability dye, which stains only dead cells therefore cells negative for zombie dye were selected for. Cells which were GFP positive (containing the Cas9 protein) and zombie dye negative (viable) were bulk sorted by FACS before being plated at low confluency (2000-4000 cells/24 well) to grow for 1 week. The cells were plated at a low density after the bulk sort so they did not need to be disturbed for passaging during the recovery phase (1 week). This approach also best supported edited cells, which may grow slower than non-edited cells due to disruptions to essential DNA replication genes and reduced the chances of edited cells being out competed by potentially healthier WT cells. After 1 week, the successive rounds of cell division should mean the plasmid would no longer be present and the cells should have lost GFP expression, confirming that Cas9 had not integrated into the genome and no more editing should occur. At this point, cells were individually sorted into the inner 60 wells of a 96 well plate (Figure 3.11) as the outer wells had a tendency to dry up. The media was changed every 3-4 days for up to five weeks, to ensure no slow growing clones were missed. Once single cells had grown colonies covering >60% of the well bottom surface area (~2-3 weeks), cells were expanded to 24 well plates. DNA was extracted from cells when expanding to 24 or 6 well plates and cells were genotyped by PCR and restriction endonuclease digestion.

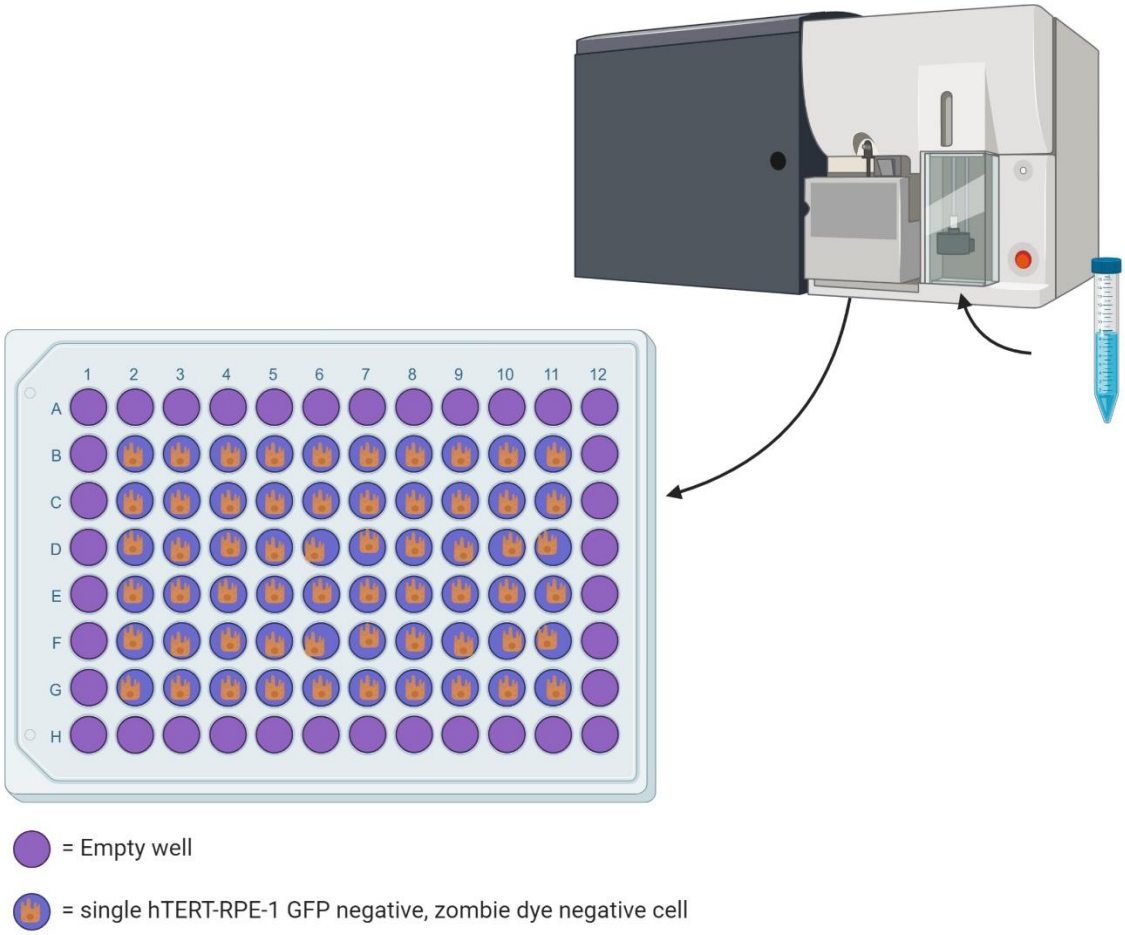


Figure 3.11 FACS BD Aria was used to single cell sort 60 GFP negative, Zombie dye negative potentially edited hTERT-RPE-1 cells in to a 96 well plate. Single cells were sorted into the inner 60 wells of a 96 well plate using FACS BD Aria instrument, as the outer wells had a tendency to dry up. (Figure created with BioRender).

The first CRISPR editing experiment was undertaken using the repair template designed for *ORC1* p.L622P. This was chosen as the Taq (α) 1 restriction endonuclease digest had already been tested on reference and patient DNA confirming the digest as an effective screening measure (Figure 3.12).

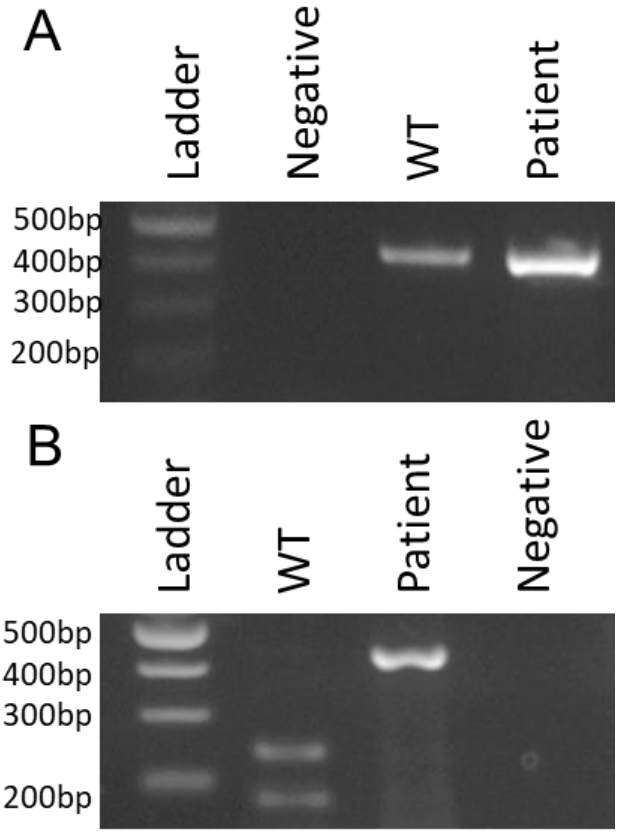


Figure 3.12 A Taq (α) 1 restriction endonuclease digest is an effective screening method for *ORC1* p.L622P edited clones. A) A 500 bp fragment of *ORC1* was amplified by PCR from reference (WT) and patient DNA. B) A Taq (α) 1 endonuclease digest was performed and confirmed as an effective screening measure for *ORC1* p.L622P edited colonies. The Taq (α) 1 restriction site is lost with the addition of the desired patient variant. Therefore only WT DNA not containing this variant is digested (giving two bands).

The first FACS sort showed a low transfection efficiency with only 0.4% GFP positive cells (Figure 3.13), suggesting scaling the experiment up to 6 well plates was not optimal. The gate determining "positive" GFP fluorescence was therefore set quite low (very close to GFP negative cells background fluorescence) to try and increase the number of cells obtained (Figure 3.13 C). All GFP positive cells were collected (8284), and resuspended at 8000 cells/mL, these were plated at 2000 or 4000 cells per 24 well. DNA was extracted from the remaining ~2284 cells and following PCR amplification the Taq (α) 1 restriction digest was performed to test for edited alleles, but no uncleaved bands, indicative of the introduced allele were observed (data not shown).

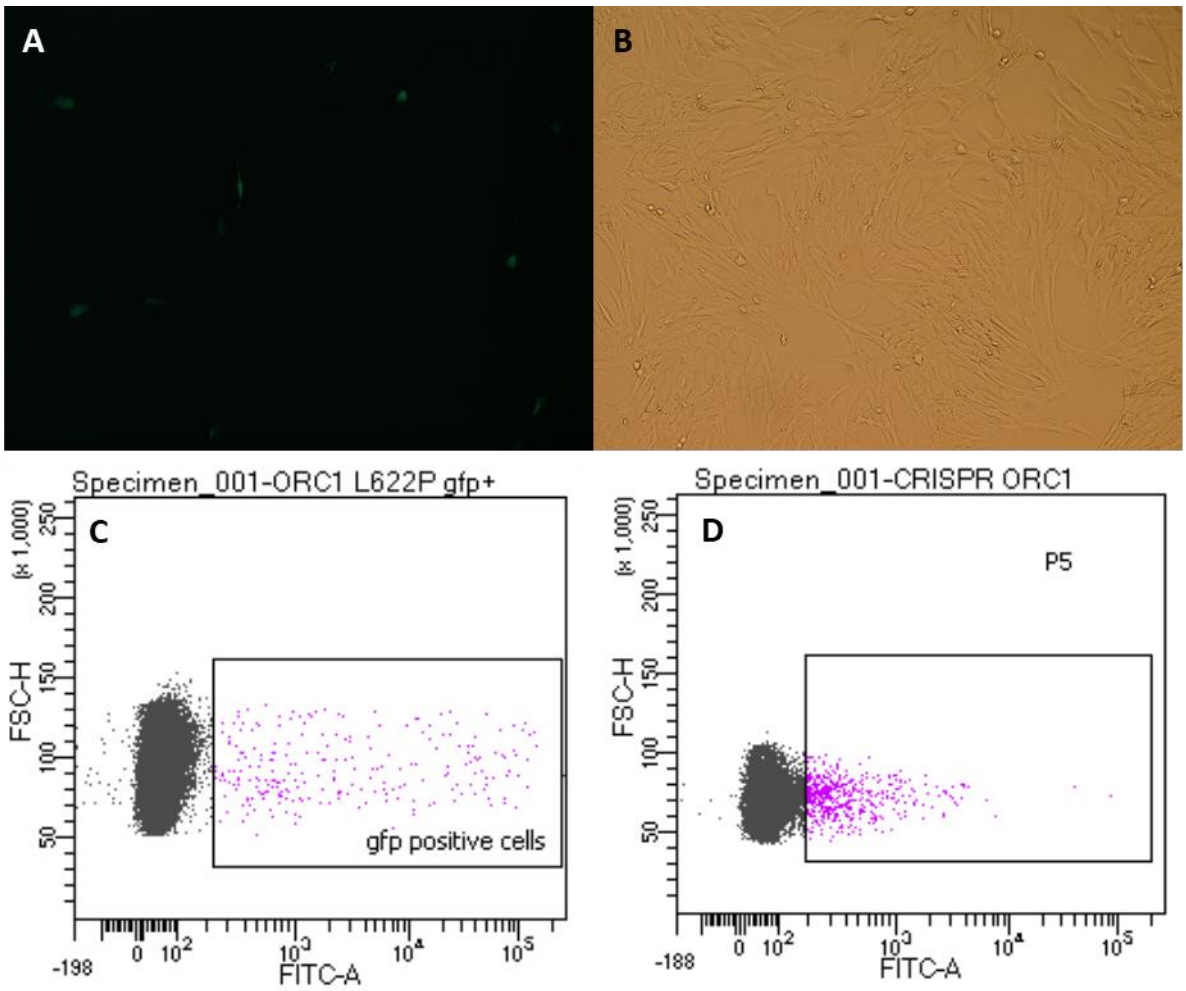


Figure 3.13 PX458-sgRNA-*ORC1*-p.L622P GFP fluorescence of transfected hTERT-RPE-1 cells. PX458-sgRNA-*ORC1*-p.L622P transfected hTERT-RPE-1 showed lower GFP expression than expected based on fluorescence observed in 24 well plates. A-B) GFP and bright field images show healthy cells with a lower than expected transfection efficiency when scaled up to 6 well plates compared to optimised transfection in 24 well plate (Figure 3.10). C-D). GFP expression was quantified during the first bulk FACS sort and confirmed cells transfected with PX458-*ORC1*-sgRNA-p.L622P expressed low levels of GFP expression. C) A total of 8284 hTERT-RPE-1 cells were gated GFP positive (0.4% of total cells), confirming low transfection level. D) 2.2% of cells sorted 6 days later were still expressing GFP.

Six days later a single cell FACS sort was performed for GFP negative cells. This was undertaken to avoid any cells that may have permanently integrated Cas9-GFP plasmid into the genome. The first single cell sort showed 2.2% of the cells were still expressing GFP, however subsequent single cell sorts were all performed 7 days after the initial bulk sort and showed very low levels (<0.1%) of GFP expression present. Based on these observations, it is likely these few remaining cells would have stopped expressing GFP in the next couple of days.

Three-hundred cells were sorted into the inner wells of 5 x 96 well plates. The media was changed every 3-4 days and wells were expanded at >60% confluency. DNA was extracted and cells frozen once ~80% confluent in a T25 flask. A 500 bp PCR amplification and Taq (α) 1 restriction digest was performed to analyse the 43 clones (Figure 3.14A). All clones were found to be WT by the restriction digest. Eleven of the clones (~25%) were Sanger sequenced. This showed no deletions or insertions present (Figure 3.14 B).

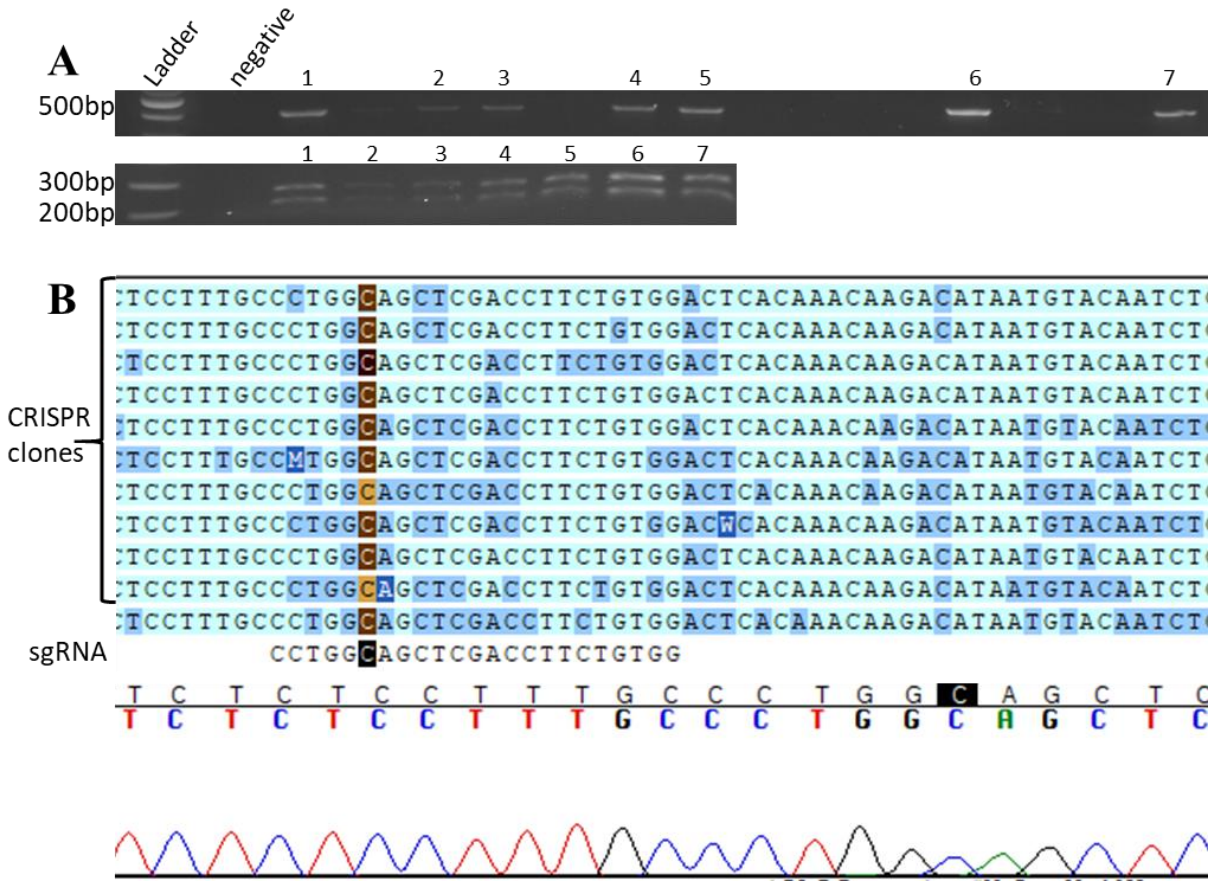


Figure 3.14 Restriction digest and Sanger sequencing confirms no editing of *ORC1*-p.L622P single cell clones. A). A 500 bp PCR was performed on crude DNA from 14 cell lines, due to the crude nature of the DNA extraction only seven of these showed successful amplification to be screened by restriction digest, 500 bp PCR was repeated on these failed samples until successful PCR amplification was achieved for all 43 clones. The seven samples showing successful 500 bp PCR were digested by Taq (α) 1 restriction digest this confirmed no editing present in these clones. This was repeated with the other 36 clones to check for the p.L622P variant, all came back WT. B) Sanger sequencing was used to screen 11 clones (~25%) to confirm no other editing was present (as indels would not likely be picked up by this screen). No editing was identified.

3.5 Allele-specific PCR to optimise repair template concentration

While FACS allows us selection of only cells expressing the GFP-Cas9 construct, the number of cells that have taken up the repair template required for HDR was unknown.

To optimise the amount of repair template (ssODN) introduced we used allele-specific PCR (AS-PCR) to identify pools of cells with the highest levels of editing. Pools of cells were transfected with the optimised plasmid concentration (Section 3.4) and varying levels of ssODN. Previous research by Cooper, 2016 in mouse stem cells found an increase in HDR editing was observed when lower concentrations of repair template (2 nM) were used. This is contrary to the general consensus on HDR editing with many new methods attempting to increase ssODN concentration at a cellular or nuclear level (Aird et al., 2018; Dupré et al., 2006; Liu et al., 2019; Savic et al., 2018).

Two forward primers matching either the WT DNA or repair template and a common reverse primer were designed using primer BLAST (Ye et al., 2012). The mutant forward primer matched the repair template therefore contained two mismatches from WT DNA at the 3′ end of the primer (-1 and -2 bp) (Figure 3.15A). The WT forward primer contained no mismatches to WT DNA. This meant there were two mismatches between the edited cells and WT reducing the chance of primer mis-binding (Figure 3.15A). Allele-specific primers were tested on WT hTERT-RPE-1 DNA and patient DNA. The M primer set did not bind to WT DNA using a high annealing temperature (Figure 3.15A). The WT primer set was more likely to bind to and amplify patient DNA (Figure 3.15B); note the faint band LB008 lane amplified by WT primers at 69.3 °C and 67.4 °C. It is likely this is due to only one mismatch being present between WT and patient DNA, as the specificity of the mutant primer pair was the most important and the WT primer pair is just being used as a positive control. This low level of mis-binding by the WT primer set was deemed acceptable.

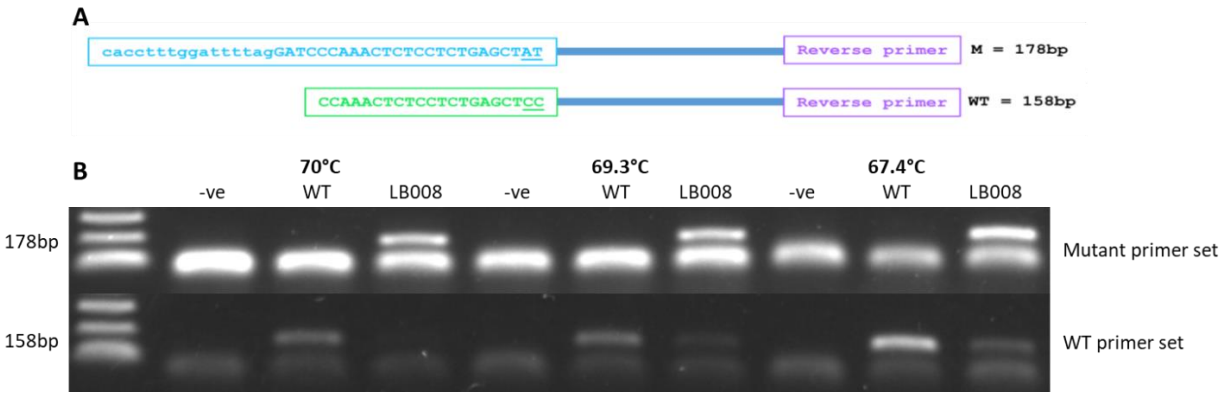


Figure 3.15 AS-PCR design and optimisation. A) Allele-specific primers for the *DONSON* p.R211C variant were designed. The mutant primer pair includes two mismatches from WT DNA at the 3' end, these two changes match the patient variant and silent PAM site variant present in the repair template. B) Primer sets were optimised using WT (hTERT-RPE-1) and patient (LB008) DNA, a negative control (-ve) was run at each extension temperature. With a high extension temperature (above 68 optimised at 70 °C) the mutant primer set does not bind to WT DNA. At an extension temperature below 70 °C the WT primer set amplified the patient DNA likely due to only one mismatch being present, as the WT primer set was being used as a control this was considered acceptable.

An attempt was also made to design AS-PCR primers for the p.L622P *ORC1* variant as well but unfortunately no primers could be designed that were robust enough to be used as a screening measure (data not shown).

Pooled cells were transfected with 500 ng PX458 plasmid and 2 nM, 3 nM, 5 nM or 10 nM of repair template. After being scaled up to 6 well plates and prepared for FACS, GFP positive cells were bulk sorted, and DNA was extracted for the AS-PCR test (Figure 3.16). This showed that higher levels of editing were present when cells were transfected with a smaller amount of repair template (2 and 3 nM) and more importantly in this experiment when GFP expression was low. Cooper, 2016 only looked at differences in the amount of repair template added, but interestingly the biggest difference observed was between the high GFP and low GFP expression rather than between repair template concentrations. Low GFP and lower repair template (2 ng and 3 ng) showed much higher levels of editing than the standard 10 ng used previously (Figure 3.16).

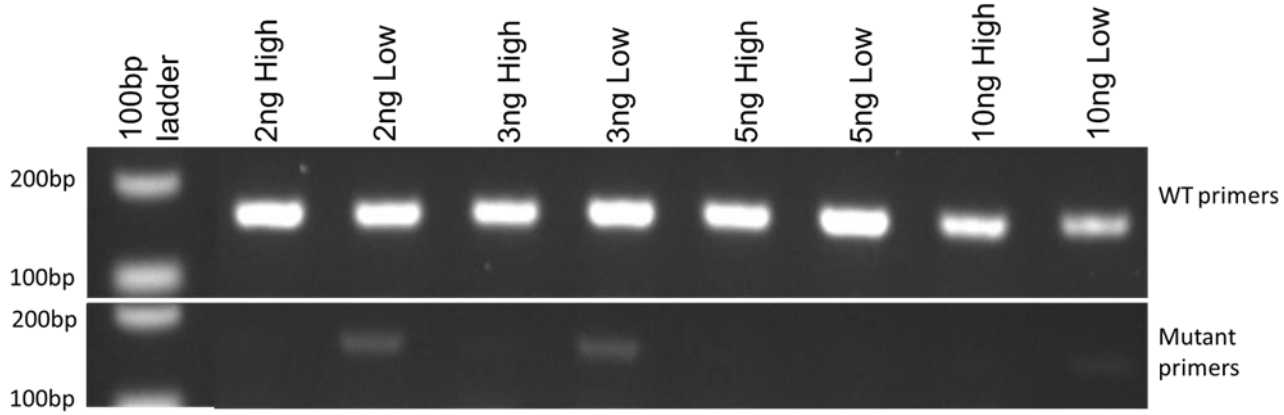


Figure 3.16 Allele-specific PCR for optimisation of ssODN concentration. Allele-specific PCR was used to optimise repair template concentrations on ~200 pooled cells 24 hours post transfection. Pooled of were cells transfected with 500 ng PX458-DONSON and different concentrations of repair template between 2-10 ng. A FACS sort was performed 24 hours post transfection and cells were sorted based on GFP expression level low ($<10^4$) or high ($>10^4$). Amplification by the mutant primer set (indicating editing) can be seen for 2 ng Low, 3 ng low and 10 ng low. Other fainter bands are present in the band but unfortunately cannot be seen here.

We hypothesized that decreased HDR editing in pooled cells with high GFP expression could be due to more Cas9 activity. High Cas9 activity could be causing more double-strand DNA breaks and therefore an increased likelihood of indels due to NHEJ being introduced before the cell has an opportunity to undertake HDR-mediated repair (Mao et al., 2008). If the repair template is not used for HDR-mediated repair the Cas9 will continue to recognise the PAM site and can therefore keep cutting until an indel is introduced leading to the Cas9 protein being unable to recognise the PAM or sgRNA site.

The proportion of cells expressing GFP decreased when higher concentrations of repair template were added, suggesting high levels of ssODN result in a lower transfection efficiency of PX458 (Figure 3.17).

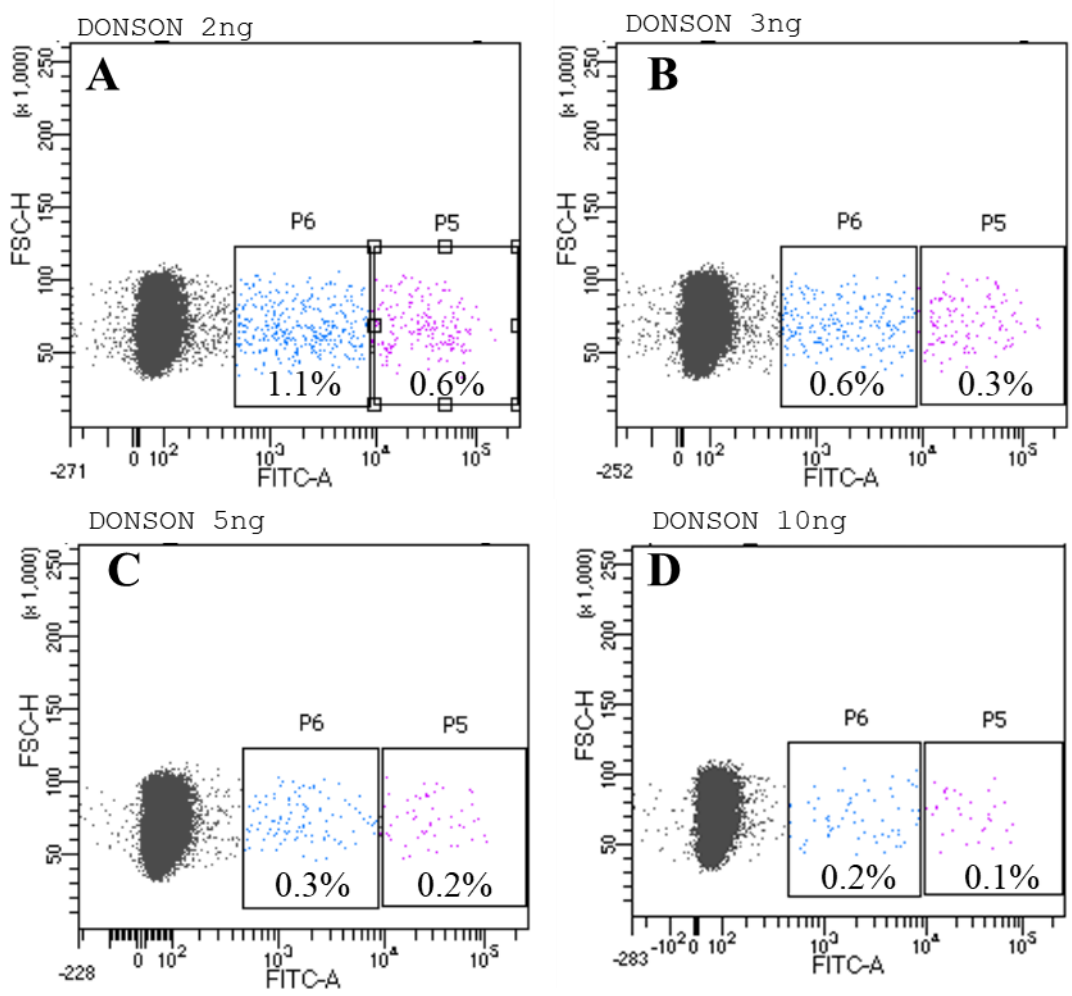


Figure 3.17 GFP expression decreased with increase ssODN concentration. Blue dots = low GFP fluorescence, purple dots = high GFP fluorescence. **A)** 2 ng *DONSON* ssODN, 1.1% of cells expressing low GFP, 0.6% of cells expressing high GFP. **B)** 3 ng *DONSON* ssODN, 0.6% of total cells expressing low GFP, 0.3% of total cells expressing high GFP. **C)** 5 ng *DONSON* ssODN, 0.3% of cells expressing low GFP, 0.2% of cells expressing high GFP. **D)** 10 ng *DONSON* ssODN 0.2% of cells expressing low GFP, 0.1% of total cells expressing high GFP.

One week after this bulk sort 317 cells from the 2 nM and 3 nM low GFP bulk pooled cells were single cell sorted into 5.3 x 96 well plates (four plates contained cells from 2 nM pooled cells and 1.3 plates contained cells from the 3 nM pooled cells). Cells in 223 wells from the original 317 cells grew and were genotyped using AS-PCR and Sanger sequencing. A SacI restriction enzyme site should have been lost with HDR editing; unfortunately the restriction enzyme was found to incompletely cut WT control DNA when following the manufacturer's protocol (data not shown), suggesting allele-discrimination would not be possible. Due to time constraints, AS-PCR was chosen to genotype the cells. AS-PCR unfortunately started to show some unspecific bands from the mutant primer set (Figure 3.18A). While there were still no bands in the negative control,

this could possibly been due to slight differences in the PCR machines as this PCR was relatively sensitive to changes in temperature and had been optimised on the most heavily used PCR machine.

Sanger sequencing was used to investigate a proportion of cells exhibiting this second fainter band from the M-primer set. When sequenced, samples with these smaller bands tended to show small indels (Figure 3.18B). These small indels were only ever present in one allele (heterozygote), as two indels at this point in the gene would be lethal to the cells (Evrony et al., 2017). These indels confirmed that the sgRNA was effective at guiding the Cas9 protein to the correct site and the Cas9 was successfully creating double-stranded breaks. Therefore it seemed the main issue in obtaining HDR editied cells was getting sufficient cells to undergo HDR-mediated repair.

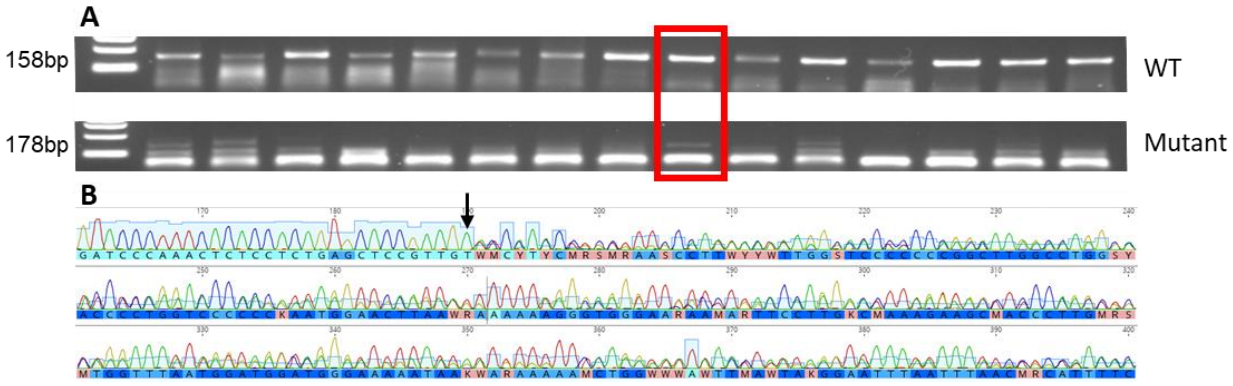


Figure 3.18 AS-PCR of single cells show amplification from WT and mutant primer pairs. A) WT amplification is shown for all 15 samples, mutant amplification is seen in four samples although not quite at the expected size. The red square indicates 1.E7 cell line which was chosen for Sanger sequencing. B) Sequence chromatogram for 1.E7 shows a 1 bp insertion directly after the cut site on one allele (indicated by arrow).

3.6 Improvement in HDR using a biotin mSA interaction

Due to the inability to produce cells containing desired patient variant other ways to increase HDR efficiency were investigated. A recent paper by Gu et al., 2018 increased HDR editing in 2 cell mouse embryos up to 95% by adding a biotin tag to their repair template and an mSA protein to their Cas9 protein. The biotin-mSA interaction tethers the ssODN in close proximity to the cut site, since the repair machinery requires the template to be nearby this proximity may be helpful in increasing HDR efficiency.

The PCS2 + mSA plasmid did not contain any florescent tag for sorting, but given the low transfection efficiency in hTERT-RPE-1 cells, a selection marker was required to identify

positively transfected cells. To overcome this Gibson assembly was used to clone the mSA fragment from the PCS2 + mSA plasmid to the C-terminus of the Cas9-GFP complex in the PX458 plasmid, creating a Cas9-GFP-mSA complex (Figure 3.19).

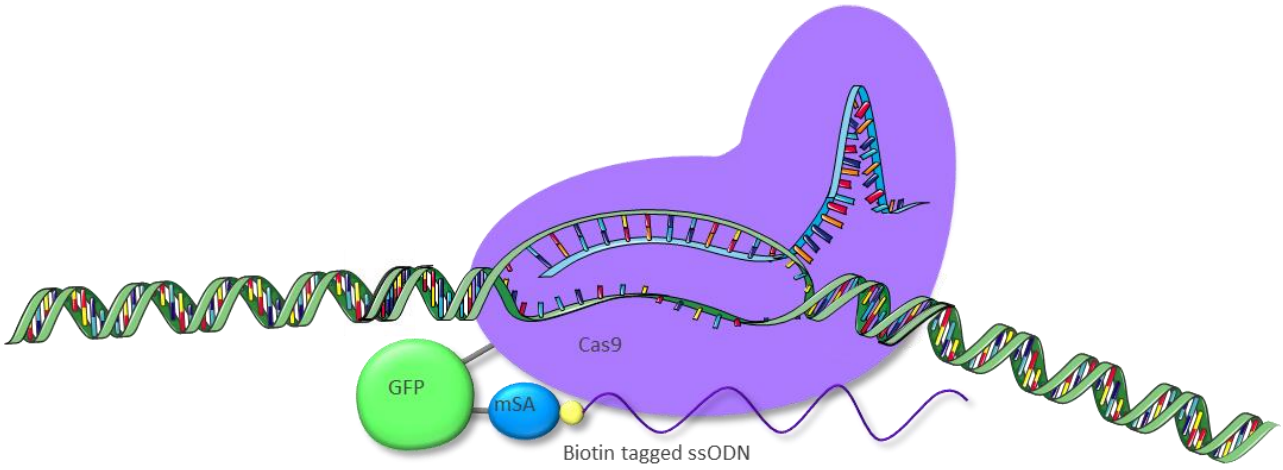


Figure 3.19 Schematic of Cas9-GFP-mSA complex plus sgRNA and ssODN interacting with DNA at desired cut site. Schematic shows how the addition of an mSA protein to the Cas9-GFP complex and a biotin tag on the repair template should increase HDR editing due to the repair template being held in close proximity to the cut site. (Figure created in PowerPoint DNA template adapted from https://smart.servier.com/).

Gibson assembly was used to create three Cas9-GFP-mSA plasmids (Figure 3.20), one empty PX458-mSA plasmid, for use as a negative control and also to clone in any future sgRNAs, PX458-mSA-*ORC1*-sgRNA-p.L622P (PX458-mSA #3.3) and PX458-mSA-*DONSON*-sgRNA-p.R211C (PX458-mSA #22.15).

As described in previous sections, plasmid concentrations for hTERT-RPE-1 transfections were optimised by eye in 24 well plates at 1000 ng per well for PX458-mSA #22.15 48 hours post transfection. The experiment was scaled up to multiple wells in a 6 well plate and a bulk FACS (GFP positive, Zombie negative) and single cell FACS sort (GFP negative, Zombie negative) were performed a week apart. As above AS-PCR was used to screen pooled cells for HDR editing. Based on results from Gu et al., 2018 an increase in HDR editing was expected, compared to the original PX458 plasmid.

In this experiment pooled cells were transfected with varying levels of repair template and bulk sorted by high and low GFP.

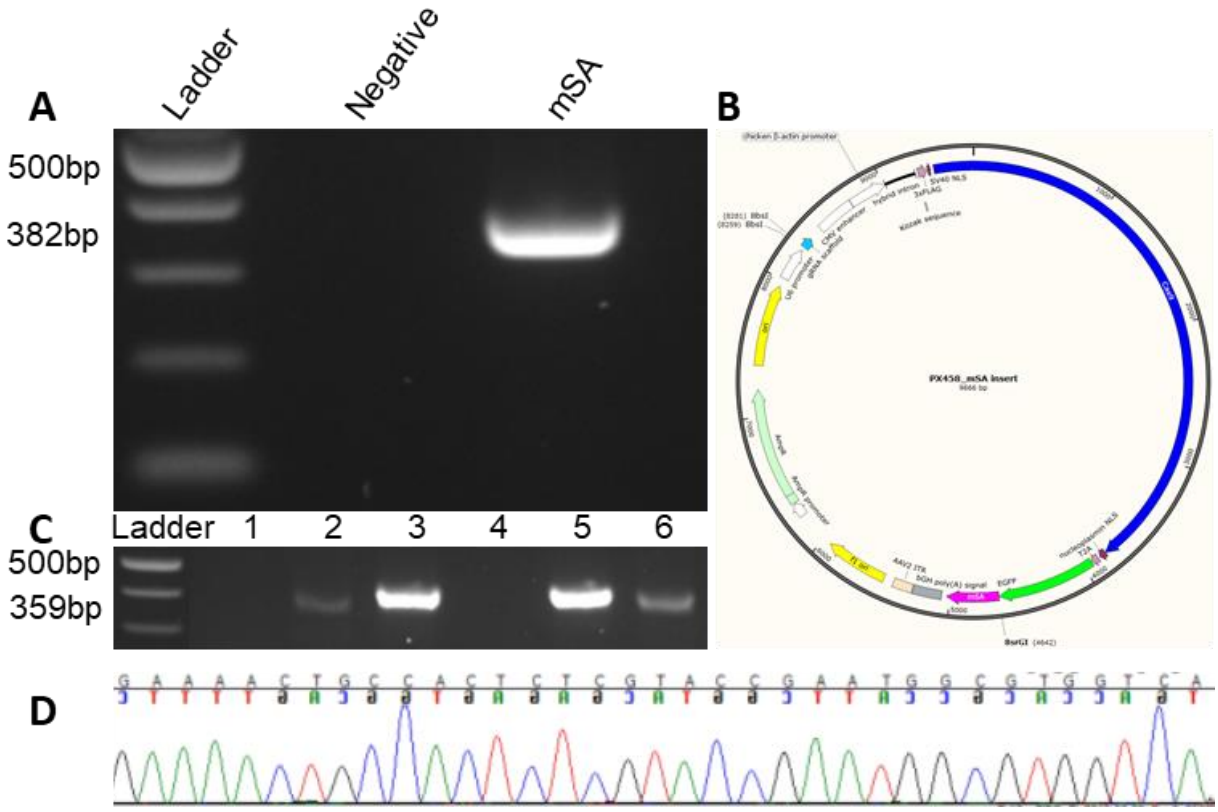


Figure 3.20 Gibson assembly was used to clone the mSA fragment into the PX458 plasmid A) The mSA fragment was amplified from the PCS2 + Cas9-mSA plasmid by Phusion PCR, using Gibson assembly primers (containing an overlap to the cut site of PX458). B) The plasmid map shows where the BsrGI restriction enzyme was used to cut PX458 (end of the Cas9-GFP complex) and where the mSA fragment sits in relation to the Cas9 and GFP in the new PX458-mSA plasmid (pink). C) cPCR was used to screen potential colonies for the correct insert, note 4 potential colonies (2, 3, 5, 6) differences in band brightness likely due to PCR amplification being done directly from *E.coli* colony resuspended in MilliQ water. D) Sanger sequencing was used to confirm the correct insert for the PX458-mSA-empty, PX458-mSA-*ORC1*-p.L622P and PX458-mSA-*DONSON*-p.R211C plasmids.

Unfortunately, AS-PCR didn't show any major differences in editing levels between the two plasmids (Figure 3.21). We hoped this experiment would replicate the original finding that HDR editing levels increased in cells transfected with 2 or 3 nM ssODN and sorted based on low GFP expression, unfortunately this was not the case. The differences in HDR editing between high GFP and low GFP was definitely not as striking as previously (Figure 3.16) but some subtle differences were observed.

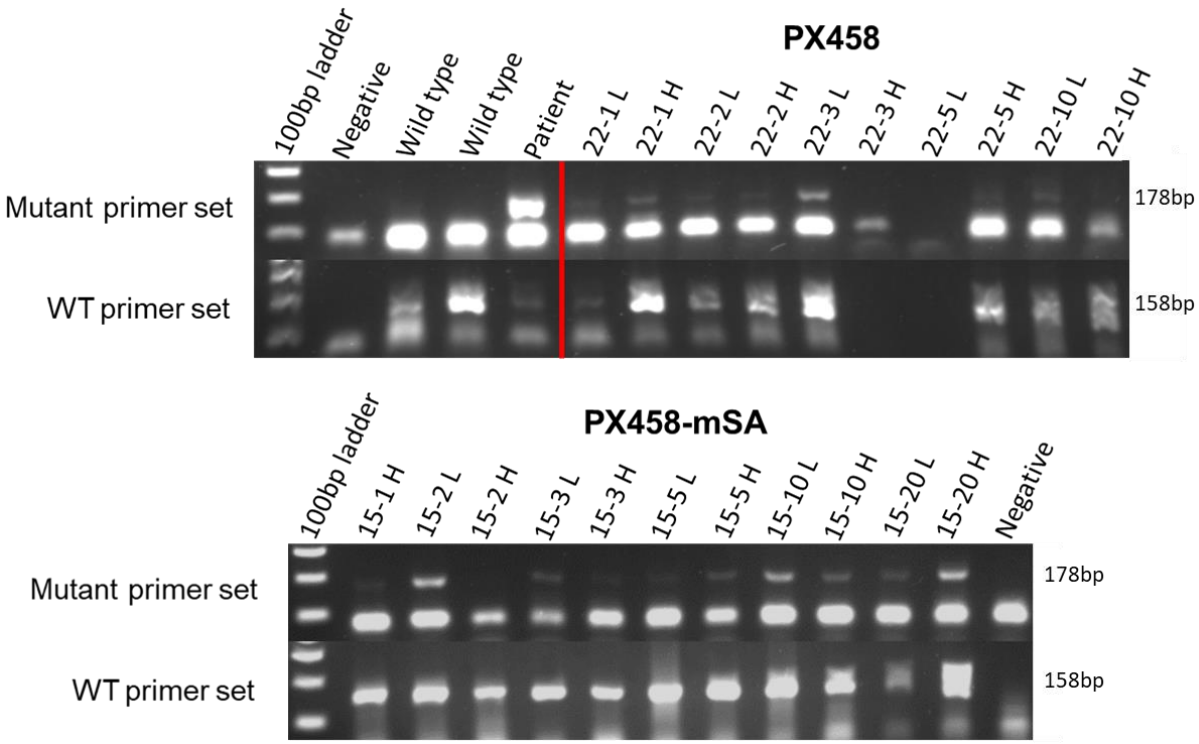


Figure 3.21 AS-PCR compares editing levels in bulk sorted cells transfected with either PX458 or PX458-mSA. Cells were transfected with 1000 ng of either PX458 or PX458-mSA, and between 1-20 nM ssODN per 24 well. Cells were bulk FACS sorted into high and low GFP conditions as laid out in previous experiment. DNA was extracted from all collected cells and editing levels were compared using AS-PCR. Mutant primer set (top of gels) shows editing levels and WT primer set (bottom of gels) is run as a control.

Previously 2 nM, 3 nM and 10 nM low GFP bulk sorted cells had significantly more HDR editing occurring than other bulk sorted cells. Here PX458 3 nM low GFP (lane 22-3L) still had the highest level of HDR editing (Figure 3.21). In the previous experiment 2 nM low a strong band suggesting a higher level of editing, however this was not replicated here, with the 2 nM low band approximately the same intensity as 2 nM high band (Figure 3.21, compare 22-2L with 22-2H). For the new PX458-mSA plasmid three bands showed relatively high levels of HDR editing; 2 nM low, 10 nM low and 20 nM high (Lanes 15-2L, 15-10L and 15-20H). Two of these fit with what was previously observed, in that lower GFP expression leads to more HDR editing (Figure 3.16). The main difference between the two plasmids used, was that the biotin-mSA interaction should retain the ssODN in close proximity with the Cas9 protein. Therefore some differences were expected in pooled cell samples containing the highest HDR editing compared to the standard PX458 approach. The high GFP, high repair template concentration showed enhanced editing. A high level of repair template would allow all Cas9 proteins in a cell to have an ssODN tethered to it therefore increasing the chance of an HDR edit. A cell with high GFP and high level of

repair template would therefore contain multiple Cas9-ssODN complexes directed to the targeted area of the genome increasing HDR potential. Unfortunately across all the pooled samples there did not seem to be more HDR editing using the mSA tag approach which was disappointing.

3.6.1 Investigating a reduction in editing levels between bulk and single cell FACS sort

The occurrence of HDR editing 48 hours post transfection in multiple concentrations as indicated by positive mutant bands in the AS-PCR, indicated that Cas9, sgRNA and ssODN were all being introduced into the cell and guided to the correct area of the genome to allow HDR editing to occur, and this was successfully occurring to some extent. After finding none of the 223 cells from the first AS-PCR CRISPR sort contained the desired variant, it was hypothesised the edited cells may not be surviving the sorting and plating process. It was also noted that when plating cells after a bulk FACS sort, many of the GFP positive CRISPR cells would die off in comparison to WT control cells. As both WT and potentially edited cells has been exposed to transfection by Lipofectamine with empty or sgRNA containing PX458/PX458-mSA, FACS preparation and FACS sorting a similar levels of death from FACS sorting associated stress on cells was expected. The only difference between the cells was that potentially edited cells contained an active Cas9 protein with the ability to be guided to the DNA to create double-stranded breaks and a repair template. Due to this it was considered that perhaps any cells containing homozygous edits were not surviving to the single cell sort stage.

The FACS sorting process is quite stressful on the cells and leads to a high level of cell death in healthy cells. Only around 40% of WT cells would survive one sort and our protocol required two FACS sorts. Previous experiments had only confirmed editing was present 48 hours post transfection, directly after bulk FACS sorting. However it was unknown how many edited cells remained with by the time the single cell FACS sort was performed.

The next experiment was designed to investigate whether the editing level was reducing due to cell death between the bulk FACS sort and the single cell FACS sort. To test this the four plasmid and ssODN concentrations with the strongest bands (highest editing levels) as seen in Figure 3.21 (22-3 low, 15-2 low, 15-10 low and 15-20 high) were chosen. These

concentrations were bulk sorted into high and low GFP as previously mentioned and half of the cells were taken for DNA extraction while the other half were plated at >8000 cells/mL in 24 well plates. Medium was replaced after 3 days and only WT (unedited) cells needed to be passaged over this time due to becoming over 60% confluent. One week after the bulk FACS sort (9 days post transfection) cells were trypsinized and DNA was extracted.

AS-PCR showed moderate levels of editing was present for the majority of samples at 48 hours post transfection, however at 7 days post FACS the majority of this editing had disappeared (Figure 3.22). It was hypothesised that HDR resulting in cells homozygous for the patient variants leads to insufficient proliferation in the absence of fully functional protein and replication stress.

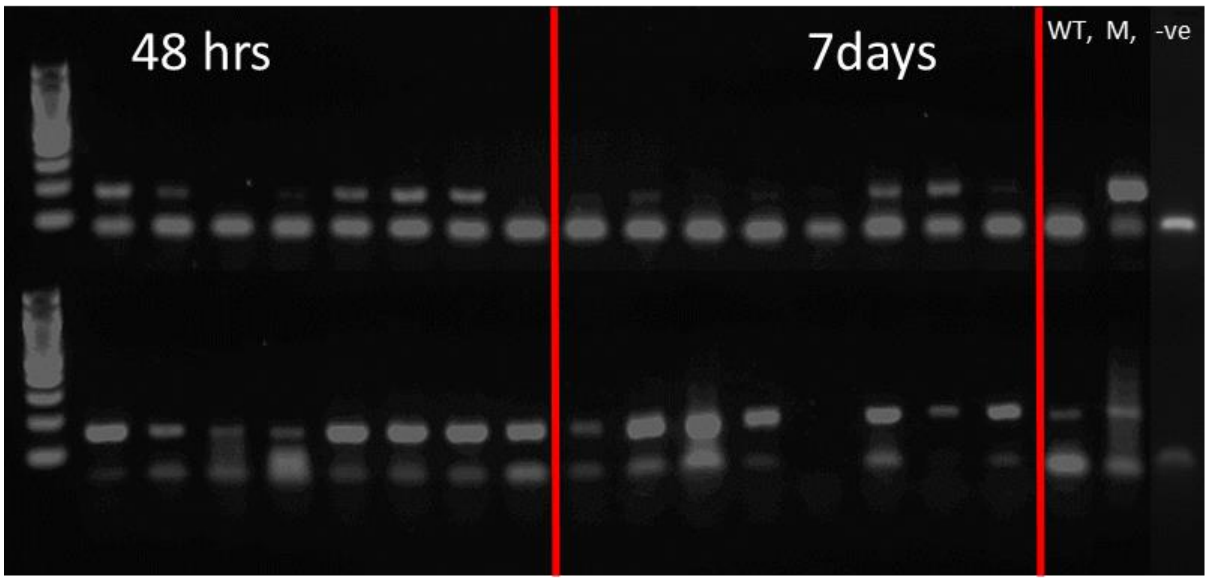


Figure 3.22 AS-PCR demonstrates a reduction in editing between bulk and single cell sorts. The pooled cells showing best HDR editing (brightest bands) from previous experiment were chosen to be investigated further. DNA was extracted from 50% of the bulk sorted cells 48 hours post transfection (after bulk FACS sort). The remaining 50% of cells were plated at ≥8000 cells/mL in wells of a 24 well plate. DNA was extracted from these cells seven days later (when a single cell FACS sort would normally occur), this showed a decrease in editing across all samples. Samples run on top half of gel were amplified with the mutant primer set. Samples run on the bottom half of the gel were amplified with the WT primer set. Three controls were run at the end of the gel (WT = wild type control DNA – only amplified by the WT primer set, M = patient DNA, -ve = negative water control).

For the final CRISPR experiment we decided to not undertake the initial bulk FACS sort. We hypothesized this would reduce the chance any edited cells would not survive by reducing the stress on the cells as they only had to survive one FACS sort. While in some

of the previous experiments an increase in HDR editing was seen when transfecting cells with lower levels of ssODN, successful editing was not observed when screening individual cells. The highest editing levels for PX458-mSA was observed at 20 nM, high GFP expression (Figure 3.21). We therefore decided to attempt a final CRISPR experiment using no bulk sort, a higher ssODN concentration and high GFP expression. The remaining ssODN was added to cells at 40 nM and 80 nM 2-4x higher than previously used. These cells were sorted directly into 24 x 96 well plates (1440 single cells). The plates were media changed every 3-4 days as in previous sorts. After 2 weeks 7 wells showed growth which were expanded and genotyped. Plates were kept for five weeks total and no more growth occurred. All cells were genotyped via restriction digest, AS-PCR and Sanger sequencing – all were confirmed WT.

As seen in previous experiments FACS sorting is stressful on the cells leading to only ~40% of WT cells surviving. It was observed that HDR editing levels decreased a week after the first FACS sort. This final experiment attempted to reduce the chance of any edited cells being out competed by WT cells, by only single cell FACS sorting the cells. Editing was present at the time of sorting however no CRISPR edited cells survived. Compared to 40% of WT cells surviving a single cell FACS sort, only 0.6% of these potentially edited cells survived. This helped confirm that any *DONSON* HDR edited cells were not surviving the single cells sorting process.

3.7 Discussion

3.7.1 CRISPR experiment limitations and improvements

CRISPR-Cas9 gene editing is a complex process, with many variables involved which could lead to the lack of HDR editing, seen in the hTERT-RPE-1 cells. Firstly, based on previous research it is known that HDR is a relatively inefficient process (Liu et al., 2019; Nambiar et al., 2019; O'Brien et al., 2019; Xu et al., 2018). Due to this, a large number of cells must be grown and genotyped to increase the chances of finding cells containing the correct patient variants produced by HDR repair. Secondly, hTERT-RPE-1 cells also have a low transfection efficiency and low FACS survival rate. This leads to a limited number of cells available to be single cell sorted, which is therefore a likely contributor to the lack of HDR edited cells. Third, it is known that RPE human diploid cells do not tend to tolerate change as well as embryonic, stem or cancer cell lines do (Haapaniemi et al., 2018). Due to

requiring these CRISPR edited cells to investigate subtle changes in DNA replication and cell cycle progression, the cell type could not be changed to an easier to edit cell line. Double-stranded breaks can cause cell death by apoptosis (Kaina, 2003; Kashiwagi et al., 2018). When the Cas9 protein creates a double-stranded break, this DNA damage is monitored by the cells. In non-cancerous cells the creation of double-stranded breaks can cause the cell to activate the apoptosis pathway, while in cell lines with mutated p53 this is less likely to happen (Gurley et al., 1998; Kaina, 2003; Menon and Povirk, 2014). Unfortunately cell lines with mutated p53 are not suitable as a model for MGS as these cell lines are less likely to remain diploid. As although they are easier to CRISPR edit (Haapaniemi et al., 2018; Ihry et al., 2018), a lack of p53 is likely to lead to more variants arising, as a result of DNA damage not leading to cell cycle arrest or apoptosis.

3.7.1.1 **FACS sorting**

As mentioned above hTERT-RPE-1 cells did not tolerate FACS sorting well. Due to the low transfection efficiency FACS sorting was required to separate GFP positive cells (containing Cas9), and separate out single clones. FACS sorting resulted in ~60% cell death in control cells. Cells which has been transfected had a higher death rate than this (as observed by eye), although whether this was due to the stress of transfection or due to Cas9 activity is uncertain. The FACS sorting process was especially detrimental for any potentially edited cells, which were likely already not growing as well as control cells, due to altering of essential DNA replication genes. As two FACS sorts were required for our CRISPR protocol, it is likely any edited cells did not survive this sorting process.

3.7.1.2 **Accidental knockouts**

A biallelic knockout of *ORC1* or *DONSON* is lethal (Dickinson et al., 2016; Reynolds et al., 2017), due to both genes being essential for DNA replication. MGS is a recessive disorder which requires biallelic knockdown variants, however, any cells which obtained two indels causing a knockout would not survive. MGS has a severe growth phenotype, it is therefore possible any cells that obtained one missense (knockdown of function) and one knockout variant may not have survived due to too little protein remaining for the cell to function correctly. This would however depend on how severe a phenotype the missense variant caused. As neither the *ORC1* nor *DONSON* variant understudy segregated with a

knockout allele it is presumed that a knockout may not be tolerated with these variants. A knockout on one allele may lead to not enough of this semi functional protein to meet the minimum threshold required for DNA replication and cell cycle progression. When added to the stress of transfecting and FACS sorting, this lack of functional protein could be contributing to the death of any edited cells.

3.7.1.3 **Delayed growth**

Another variant that could be contributing to the lack of HDR edited cells, is that edited cells may be out competed by non-edited cells during the week between the bulk and single cell FACS sort. Plating cells at a low density was done to attempt to control for this, however if HDR editing was having an effect on the cells ability to grow at the same speed as non-edited cells, then over the course of the week we will have accidentally biased the population of cells towards unedited cells. It is therefore less likely edited cells would have been sorted into single wells, as there wasn't enough time or plates to sort all of the cells from the bulk sort. Some cells were also lost when preparing for FACS, any edited cells lost here would reduce our chance of edited clones growing after the single cell sort.

3.7.1.4 **sgRNA efficiency**

Screening measures were designed to only identify HDR edited cells, rather than cells with small indels. This was due to the low proportion of cells undergoing HDR compared to NHEJ (Mao et al., 2008). Therefore, the overall editing level of the sgRNA was not measured. Because the screen was not designed to identify indels, and only a small proportion of cells were Sanger sequenced, this gave the impression that the overall editing level was low. NHEJ is the preferred method of repair by cells not in S-phase (Mao et al., 2008; Maruyama et al., 2015; Sargent et al., 1997; Symington and Gautier, 2011), it is therefore likely the overall editing level of these sgRNAs was higher than was observed, although this cannot be confirmed. As the overall editing level was not investigated, it is possible the chosen sgRNAs were not as effective at guiding the Cas9 protein to the regions of interest as presumed based on it guide score (Section 3.3.1). If this were true the lack of endonuclease activity by Cas9 could have been a contributing factor to the overall lack of HDR editing.

Ideally multiple guide RNA's would have been tested for each gene, but due to a lack of suitable guides in the region there were very few options available without changing Cas proteins/species. It is recommended that the Cas9 cut site be as close as possible (within 5 nucleotides) to the desired knock in variant (Komor et al., 2016; O'Brien et al., 2019). As indels were observed when using the *DONSON*-p.R211C-sgRNA, this gives some confidence the Cas9 protein was being directed to the correct area.

3.7.1.5 **Overview**

Before taking into account the many new methods available for CRISPR-HDR editing, which arose after the initial design of this CRISPR experiment, there are many variables at play that could have been contributing to the lack of editing observed. Due to the time constraints of this project, we decided to focus our efforts on investigating our novel MGS variants using other techniques. We had originally planned on using our CRISPR edited cells to investigate the effects these variants had at a molecular and cellular level. Due to the length of time CRISPR required if we had continued to investigate new CRISPR techniques we would not have been able to investigate any of the effects these MGS variants were causing.

3.7.2 New CRISPR methods

CRISPR methodologies to increase HDR editing efficiencies are improving all the time. A large volume of studies investigating improvements of HDR editing efficiency were published during the time course of this project. The CRISPR experiments were originally designed in March 2018, with methods and experimental protocols being redesigned and carried out until August 2019. Therefore we only had the opportunity to try one new method to increase the HDR editing efficiency, using the biotin-mSA interaction as published in Gu et al., 2018. This meant we did not have the opportunity to try any other methods published in newer studies.

3.7.2.1 **P53 knockdown**

As touched on briefly above Cas9 induces double-stranded breaks activating the p53 DNA damage pathway, inhibiting CRISPR Cas9 editing (Haapaniemi et al., 2018; Ihry et al.,

2018). Cell lines containing non-functional p53 are at an increased risk for variants and chromosomal rearrangements (Hanel and Moll, 2012), making them unsuitable for future experiments. However a temporary transient knockdown of p53 decreases the long term risk. Schiroli et al., 2019 showed transient inhibition of p53 in hematopoietic stem cells lead to increased editing without compromising genome stability. While short term inhibition of p53 may not lead to increased mutational burden it may allow cells whose genome has been damaged during the editing process itself to escape notice (Haapaniemi et al., 2018), therefore greater screening of any edited clones may be required to confirm other variants are not present.

3.7.2.2 **Cas9 nickase**

Cas9 nickase creates single-stranded nicks instead of double-stranded breaks (Gasiunas et al., 2012). Therefore two Cas9-nickase proteins and two sgRNAs are required to create a double-stranded break by making two nicks on the different strands (Jinek et al., 2012). This system can be tailored towards HDR by altering overhang length, leading to decreased NHEJ repair and increase HDR (Mali et al., 2013b, 2013a; McConnell Smith et al., 2009; Satomura et al., 2017). Because two sgRNAs and Cas9 proteins are required to create nicks on either strand in order for a double-stranded break to occur (Gasiunas et al., 2012; Jinek et al., 2012), Cas9 nickase has lower off-target effects, as both sgRNAs would need to mis-bind in close proximity to each other for double-stranded breaks to occur (Ran et al., 2013b).

3.7.2.3 **Asymmetrical repair template**

Research by Richardson et al., 2016 showed that Cas9 asymmetrically releases the 3' end of the cleaved DNA. An asymmetric repair template comprised of 36 bp (PAM-distal) and 91 bp (PAM-proximal) was tested and improvement of 57% (± 5%) was observed in HEK293FT cells (Richardson et al., 2016). When originally designing the CRISPR experiment, this paper was missed due to the large volume of CRISPR literature available at the time. After this another laboratory member decided to investigate improvement using an asymmetrical repair template, this experiment had not been performed at the time. Therefore when redesigning our experiment we did not chose to also include this.

Future experiments could be designed using the recommended 36 and 91 bp arms, to see if this leads to improvements in HDR in hTERT-RPE-1 cells.

3.7.2.4 **CRISPR base editors**

Base editors can be used to introduce C>T or A>G substitutions (Komor et al., 2016). Base editors use cytidine deaminase or adenine deaminase fused to an inactive Cas9, nickase Cas9 or nuclease deficient Cas9 (Ryu et al., 2019). They do not require a repair template and do not create double-stranded breaks (Gaudelli et al., 2017; Komor et al., 2016; Nishida et al., 2016), avoiding some of the issues associated with ds breaks such as increased cell death due to activated p53 and eliminating the chance of NHEJ occurring. Base editors have been successfully used to model human disease in mouse models (Kim et al., 2017a; Lee et al., 2018; Liu et al., 2018; Ryu et al., 2018).

While base editors tend to have a higher efficiency for introducing specific variants due to the lack of NHEJ (Gaudelli et al., 2017; Komor et al., 2016; Nishida et al., 2016), there are specific issues associated with them (Ryu et al., 2019). As with other Cas proteins a PAM site is required for recognition. The window available for the cytidine deaminase or adenine deaminase to work is therefore very small and specific (Nishida et al., 2016). Generally all C or A nucleotides within this window will be replaced with T's or G's respectively (Komor et al., 2016). Kim et al., 2017b successfully reduced the editing frame of these base editors from 5 nucleotides to 1-2 nucleotides decreasing off-target changes within the editing window. Unfortunately the placement of the correct PAM site in the desired area is required to be able to make these correct base pair changes at the desired location (Kim et al., 2017b). Therefore depending on location and composition of the desired knock in area, these base editors may not be suitable for all experiments. Due to this, using base editors in the project was not possible, however they could be useful in the future for editing other patient variants depending on their location.

3.7.2.5 **Small molecules**

Small molecule inhibitors have been shown to increase HDR efficiency by either stimulating proteins important for HDR or inhibiting proteins required for the faster and preferred NHEJ repair pathway (DiNapoli et al., 2020; Frit et al., 2014; Mao et al., 2008; Pinder et al., 2015).

KU70/80 play an essential role in NHEJ as they are recruited along with the catalytic subunit of DNA-dependent protein kinase (DNA-PKcs) to reattach the two cut pieces of DNA (Li et al., 2011; Lieber et al., 2010; Robert et al., 2015). Inhibition of KU70 or KU80 has been shown to increase HDR by inhibiting NHEJ 2-3 fold (Chu et al., 2015; Maruyama et al., 2015; Robert et al., 2015)

DNA Ligase IV is required for sealing of double-strand breaks during NHEJ (Srivastava et al., 2012). A 70% increase in HDR editing was observed in Drosophila embryos mutated for DNA Ligase IV when using Zinc-finger nuclease editing (Beumer et al., 2008; Singh et al., 2015). This lead to the discovery of the small molecule SCR7 an anti-cancer compound which interferes with the DNA binding ability of DNA Ligase IV (Srivastava et al., 2012). SCR7 has been used in mouse embryos to significantly increase CRISPR HDR editing 7-19 fold (Chu et al., 2015; Maruyama et al., 2015; Singh et al., 2015). Interestingly in rabbits SCR7 had little effect on improving HDR or inhibiting NHEJ (Song et al., 2016). Song et al., 2016 also observed severe death in human pluripotent stem cells.

Since its discovery SCR7 has been used to increase CRISPR HDR in a range of cell lines and species including a 2-3 fold increase in porcine foetal fibroblasts (Li et al., 2017), a 3 fold increase in A549 human epithelial cells, 19 fold increase in melanoma cell lines and a 13 fold increase in mouse dendritic cells (Maruyama et al., 2015)

RAD51 is a key HDR protein (Bozas et al., 2009; Rozov et al., 2019) and the small molecule RS-1 (Rad51 - stimulatory compound) has been found stabilise RAD51 association with DNA (Jayathilaka et al., 2008; Pinder et al., 2015). This small molecule had been found to increase HDR efficiency 2-6 fold, without interfering with the NHEJ pathway (Pan et al., 2016; Pinder et al., 2015; Song et al., 2016). However Zhang et al., 2017 actually found addition of RS-1 and RAD51 lead to a decrease HDR editing in HEK293 cells.

As with many of the improved CRISPR protocols the actual increase in HDR efficiency of these methods relies heavily on the cell line used. As shown with many of the small molecule enhancers or repressors the fold change in editing efficacy varies wildly between cell types. Therefore when designing improvements to CRISPR experiments it is important to take into account whether these protocols have been used in similar cell lines as this can lead to the difference between an increase in HDR editing and a decrease.

In relation to this thesis an end point had to be decided upon, as CRISPR experiments are time consuming as it takes a long time to grow cell lines from a single cell sort in order to

genotype them. As we didn't expect it to take so long to successfully edit in our MGS variants using CRISPR, we decided to end our attempts at CRISPR editing and use other techniques to explore the cellular consequences of these novel MGS variants.

Getting CRISPR set up in the laboratory as a technique to be used to study the cellular and molecular effect of suspected single base pair disease causing variants in a relatively easy to work with diploid cell line would be advantageous, as it would allow us to more easily explore how variants within a gene can cause different phenotypes. For example it would allow us to investigate how missense variants in different areas of *DONSON* can be causing both the MISSLA and MGS phenotype.

CRISPR HDR editing would also allow us to confirm that patient's variants suspected to be responsible for their phenotype is actually causing their disorder. As CRISPR editing allows us to control for background variation, resulting in easier conformation that these missense variants are responsible for the phenotype seen.

4 Investigating the effects of novel MGS variants at a molecular level

After unsuccessfully editing in the desired MGS variants via CRISPR, it was decided to investigate these novel MGS variants using other cellular techniques. Some of these experiments were undertaken as a precursor to attempting CRISPR editing as functional evidence that these variants were likely pathogenic was desired, before investing time and resources into undertaking CRISPR editing to establish isogeneic cell lines.

As well as the novel variants in *ORC1* and *DONSON*, fibroblasts from two siblings with a suspected *CDC45* splicing variant (via collaborator Dr David Chitayat, Toronto, Canada) were obtained for further investigation. CDC45's role in DNA replication fit in between ORC1 and DONSON as it is important in the preIC as well as during replication fork progression while ORC1's role is in early replication (preRC) and DONSON's known role is later in replication at the replication fork.

In this chapter, different techniques are presented to investigate these three novel MGS variants, in proteins required at different stages of DNA replication. This chapter is divided into sections concentrating on each variant.

4.1 ORC1: p.L622P – testing splicing effects by minigene assay

ORC1 is an essential DNA replication gene and its encoded protein is involved in the initiation of DNA replication during late M/early G1 phase (Klein and Gilbert, 2016). LB005 contains a homozygous missense two base pairs into Exon 13 *ORC1*, causing a p.L622P amino acid change. This amino acid residue is highly conserved (as seen in Figure 3.1) and is expected to be contributing to the phenotype. After unsuccessful CRISPR editing, and in the absence of patient derived cells we considered how else to investigate the molecular effect of this variant. As the variant is at the start of an exon near a splice site, we wished to confirm whether or not splicing was contributing to the phenotype. Before designing the CRISPR experiment Alamut Visual software (as mentioned section 3.3.1) was used to check for any potential splicing effect, however no changes in splicing efficiency were predicted. Given the limitations in splicing predictions (Bonizzoni et al.,

2006; Pervouchine, 2018; Spurdle et al., 2008), a minigene splicing assay was undertaken to further investigate this (Kishore et al., 2008).

Gateway cloning was used to introduce a 3 kb fragment of *ORC1* containing exons 12-14 plus >200 bp ajoining introns into the p.SpliceExpress vector. This fragment was PCR amplified from control DNA using gateway primers containing attB overlaps. A BP clonase Gateway cloning reaction was performed; Gateway cloning uses the attB sites on the *ORC1* insert to recombine with the attP sites present in the pSpliceExpress vector (Figure 4.1), removing the toxic *ccdB* gene from the pSpliceExpress vector. If this recombination event hasn't occurred the *E.coli* should die. Transformed *E.coli* were plated on ampicillin containing agar plates and undertook colony PCR (cPCR) to screen colonies for the correct insert. This suggested that all colonies screened contained the desired insert. Two colonies were selected for plasmid purification and Sanger sequencing. Sequencing showed that the insert specific primers were mis-binding to the plasmid and neither plasmid contained the insert. New Gibson assembly primers and insert specific cPCR primers were designed for the next Gateway cloning attempt, however colony number stayed reasonably similar and the new cPCR primers also showed mis-priming. This time five plasmids were selected for plasmid purification and Sanger sequencing and one contained the correct insert (Figure 4.1).

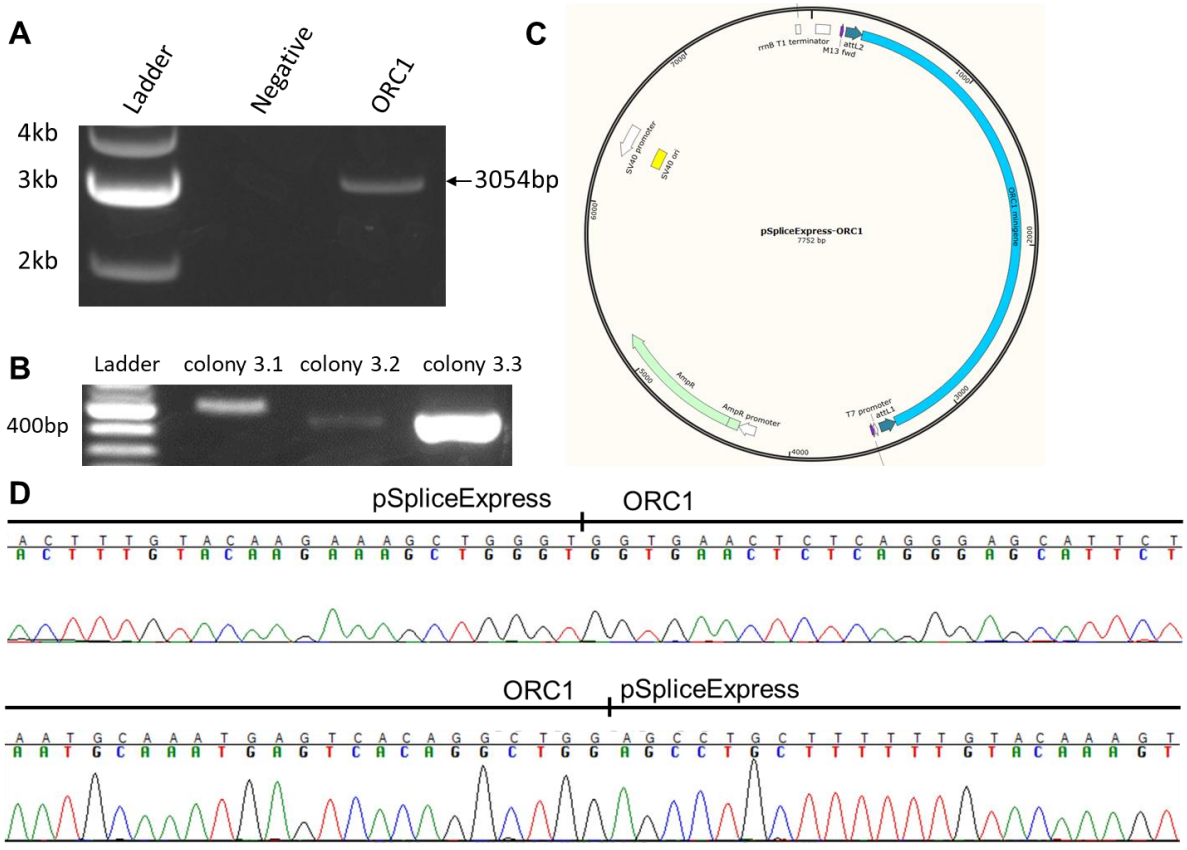


Figure 4.1 **Gateway cloning was used to introduce ~3 kb of *ORC1* into the pSpliceExpress vector.** A) PCR Phusion amplification of *ORC1* exons 12-14 including >200 bp intron DNA either side. B) cPCR confirms correct insert, colony 3.3 chosen for plasmid prep and Sanger sequencing due to strong band at correct size. C) Plasmid map of the pSpliceExpress-*ORC1* vector shows insertion of the ORC1 minigene between two Rat exons. D) Sanger sequencing of both ends of the *ORC1*-pSpliceExpresss overlap confirms the *ORC1* insert is present in the vector.

Due to inefficiency with Gateway cloning and screening, site-directed mutagenesis was used to introduce the LB005 c.1865T>C variant into the plasmid. The site-directed mutagenesis primers contained the single bp change in the middle of the 27 bp primer (Figure 4.2). The plasmid was then amplified from these primers leading to a newly synthesised plasmid containing the single bp change. This allows a single bp change to be made relatively easily. DpnI was used to digest any original plasmid before *E.coli* transformation, plasmid purification and Sanger sequencing. Both plasmids selected for Sanger sequencing showed the correct variant (Figure 4.2).

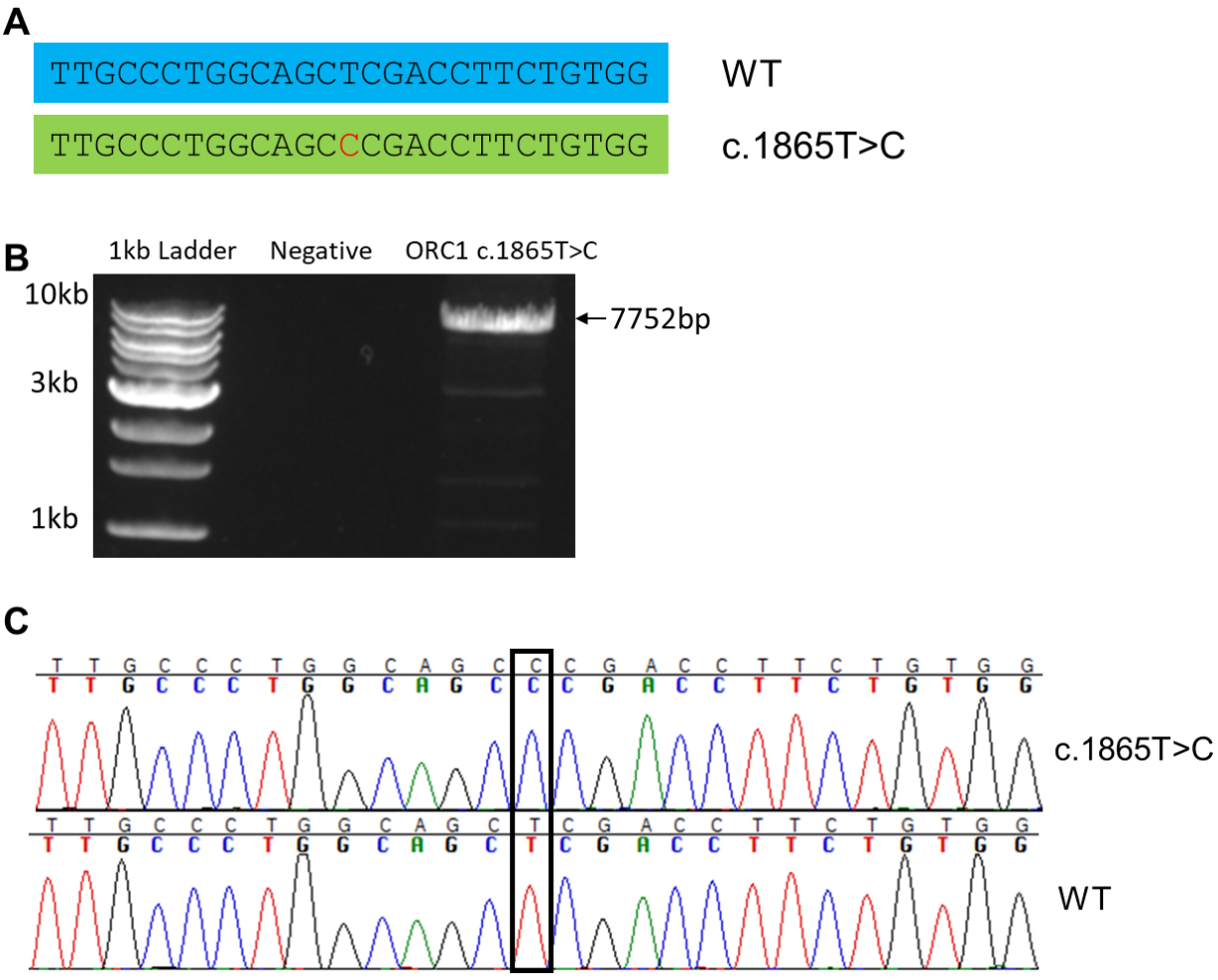


Figure 4.2 Mutagenesis introduced patient c.1865T>C variant into pSpliceExpress-*ORC1* vector. A) Mutagenesis primers contains single bp change highlighted in red in the middle of the 27 bp primers. B) A bright band at the correct size confirms successful amplification of the pSpliceExpress-ORC1 vector using the mutagenesis primers, plasmid preps were then performed. C) Sanger sequencing confirms successful mutagenesis, with new plasmid pSpliceExpress-ORC1-c.1865T>C containing the patient missense variant.

Plasmid concentrations were originally measured using the NanoDrop (2.1.6.6) but were confirmed using Qubit (2.1.6.7) before HEK293FT cells were transfected. After 24 hours RNA was extracted and cDNA was synthesised. RT-PCR was performed to investigate any splicing changes (Figure 4.3). This assay produced some unexpected bands. The 200 bp band represents plasmid to plasmid splicing, while the 578 bp band represents both plasmid exons plus all three *ORC1* exons. The other two bands were unexpected as WT *ORC1* is not expected to be splicing in this region. The Sanger sequencing results of the *ORC1* - pSpliceExpress plasmid were re-examined.

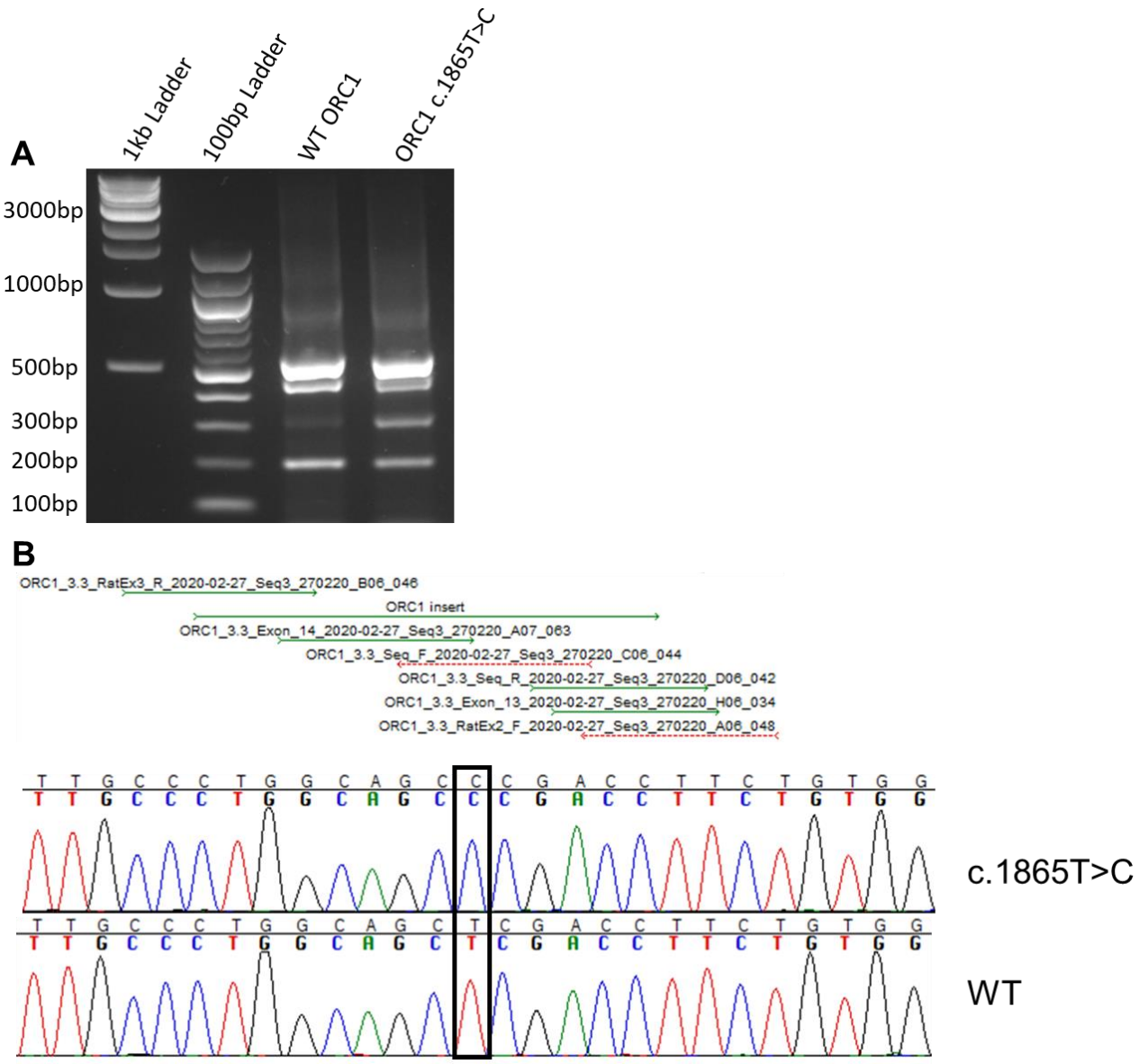


Figure 4.3 **RT-PCR and Sanger sequencing reanalysis confirmed wrong orientation of *ORC1* fragment.** A) RT-PCR gel shows unexpected results with many more bands than expected appearing from both the WT and c.1865T>C plasmids. B) Re-examination of the Sanger sequencing results confirmed the *ORC1* fragment was inserted backwards. *ORC1* sequencing alignment against the instered *ORC1* minigene fragment shows incorrect orientation. The reverse primer plasmid sequencing (green arrows) matches the origentation of the ORC1 fragments however the forwards primer sequencing (red arrows) should be in the same orientation for the promotor and plasmid exons to be situated in the correct orientation.

Reanalysis of the plasmid sequencing showed the *ORC1* fragment had been inserted in the wrong orientation which was missed when sequencing results were first analysed. Unfortunately, due to time constraints and ongoing issues cloning the *ORC1* fragment into the pSpliceExpress vector this experiment was not sucessfully repeated prior to

submission of this thesis. However in the future, it would be beneficial to confirm this *ORC1* variant does not have a splicing effect.

4.2 CDC45 splicing assay

CDC45 is part of the preinitiation complex (preIC). The preIC forms at the G1/S-interphase, after the preRC and is involved in unwinding the DNA to allow polymerase access (section 1.1.2.1). This complex is therefore essential for the initiation and progression of DNA replication. CDC45 is required for origin firing as it is the rate limiting step (Karnani and Dutta, 2011; Köhler et al., 2016). CDC45 also acts by binding single-stranded DNA and facilitating DNA displacement at the replication fork (Fenwick et al., 2016; Pacek and Walter, 2004; Parker et al., 2017). Missense and splicing variants have previously been found in MGS patients presenting with craniosynostosis and/or anorectal malformations (Fenwick et al., 2016). These studies found a marked reduction in expression of CDC45 protein expression and canonical transcripts (Cooper, 2016; Fenwick et al., 2016). CDC45 is involved slightly later in the replication initiation programme than other known MGS genes, as CDC45 is involved in both the initial replication initiation and progression of DNA replication. Previous patients with variants in *CDC45* have presented with a slightly different phenotype than other MGS patients as mentioned above, which may be due to the different role CDC45 plays in replication.

Previous patients with missense or splicing variants in CDC45 presented with craniosynostosis or MGS (Fenwick et al., 2016). The majority of the MGS patients also presented with mild to severe craniosynostosis, and the majority of the craniosynostosis patients presented with mild MGS phenotypes (slight reduction in height, hypoplastic ears, and MGS facial features such as small mouth) but no a/hypoplastic patella. CDC45-MGS patients are therefore presenting with a different phenotype than is typical for patients with variants in other genes. Along with craniosynostosis these MGS patients also presented with other secondary phenotypes such as thin eyebrows and anal abnormalities not normally associated with MGS. This difference in phenotype could be due to CDC45 working at different points in replication. Other genes involved in MGS work early and only in initiation of DNA replication (Bicknell et al., 2011b), while CDC45 works initiating DNA replication but also during S-phase at the replication fork (Pacek and Walter, 2004; Wong et al., 2011).

Here we report two severely affected foetuses, both terminated around 20 weeks gestation due to severe growth defects including severe chest and bone malformations. The parents were healthy and consanguineous (first cousins). They were both aged 25 during the first pregnancy and had no family history of inherited conditions, recurrent miscarriages, congenital abnormalities or mental retardation (Figure 4.4A).

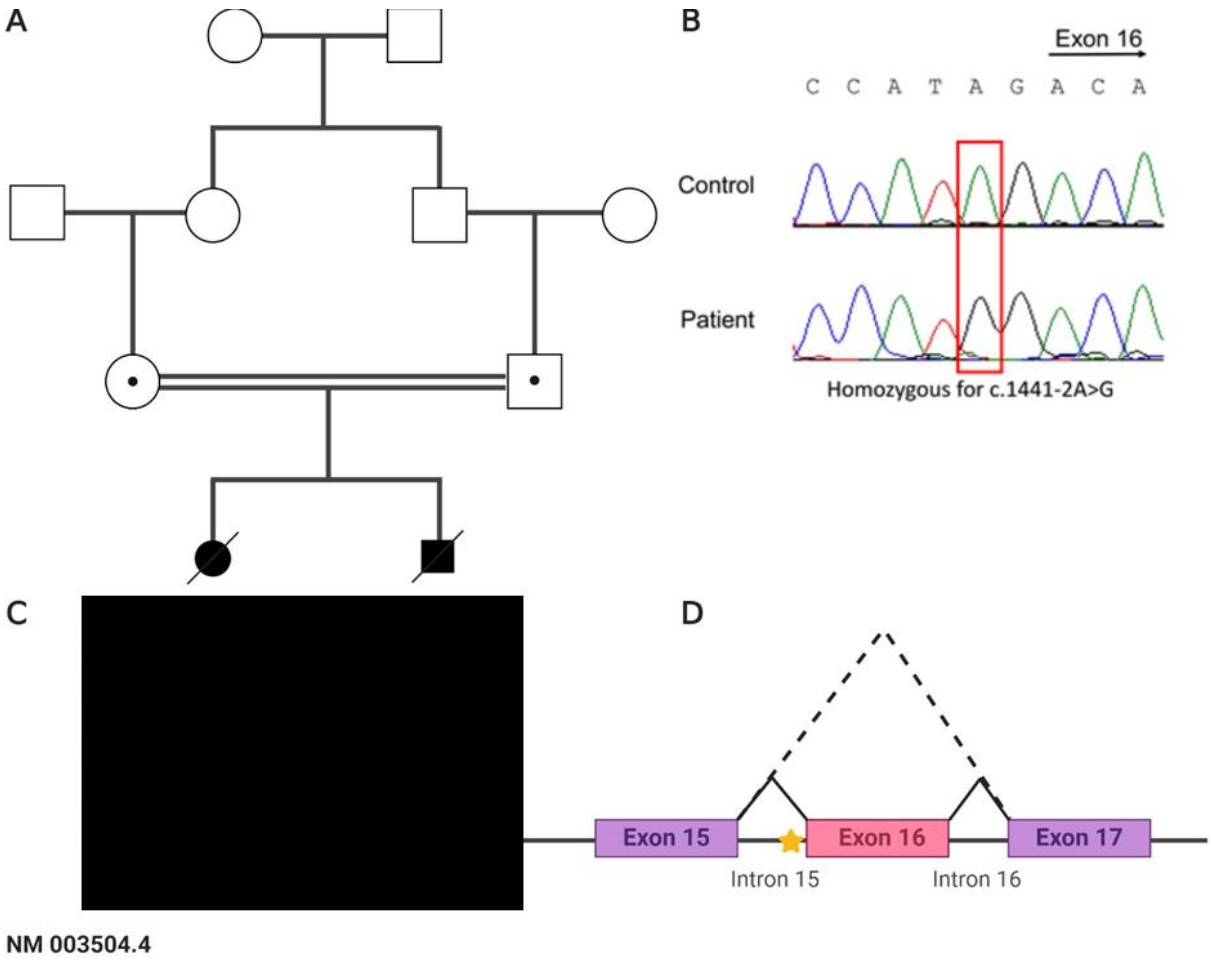


Figure 4.4 **Two severely affected siblings from healthy consanguineous parents present with a homozygous splicing variant.** A) Family pedigree shows parents of two affected siblings were first cousins with the mothers' mother and father's father being siblings. B) Sanger sequencing of patient fibroblasts confirms homozygous c.1441-2A>G suspected splicing variant (RefSeq: NM003504.4). C) Images of both siblings show small nose and mouth, absent eyelids, sloped shoulders, prominent occiput (back of head), hypoplastic external genitalia abnormal limbs including club feet with hypoplastic big toes, and hypoplastic thumbs. D) c.1441-2A>G splicing variant, results in the loss of the Exon 16 splice acceptor site. We predicted this may lead to the entirety of exon 16 being skipped out as there did not appear to be any other potential splice acceptor sites nearby. Figure created with BioRender

Foetuses suffered severe intrauterine growth restriction (~3 weeks delayed growth by 20 weeks gestation). Both foetuses also had severe malformations including absent

eyelids, cardiac abnormalities, oligohydramnios (deficiency of amniotic fluid), dolichocephaly (tall, narrow head), cerebral ventriculomegaly (increased cerebrospinal fluid in the brain), kyphotic (outward curvature of the) lumbar spine, lordotic (inward curvature of the) thoracic spine, thin ribs, absent clavicles and absent genitalia (Figure 4.4 C). This phenotype is considered more severe than what has previously been seen in MGS-CDC45 patients.

Both foetuses were found to have the same homozygous suspected splicing variant in *CDC45* c.1441-2 A>G (Figure 4.4 B). Alamut predicted the loss of the Exon 16 splice site acceptor motif when the variant is present (Figure 4.5). Such an alteration could lead to several consequences, including skipping of exon 16 due to no strong acceptor site, or the activation of a cryptic acceptor site which might change the sequence of the transcript. Of note, there were no such cryptic sites predicted nearby the canonical site (Figure 4.5).

Figure 4.5 Exon 16 splice acceptor site lost due to c.1441-2A>G variant. Alamut predicts this *CDC45* variant will lead to the loss of the Exon 16 splice acceptor site (as indicated by green boxes), as there are no other obvious splice sites nearby it is unknown where this will splice to.

4.2.1 RT-PCR was used to investigate splicing effect

Fibroblasts derived from both siblings (LB191 and LB192) were obtained from our collaborators in Canada (Professor David Chitayat – University of Toronto and Professor Jacques L. Michaud Université de Montréal). On arrival the patient fibroblasts were found to harbour a severe fungal infection, making them difficult to grow and required the addition of Fungizone (1:100 DF) to control contamination. Patient cell line LB192 was used for all experiments as this cell line was growing better than the LB191 cell line, due to less fungal contamination. The experiments which could be performed with these cells

was limited, as the addition of Fungizone can alter cell cycle progression (Zhang and Rao, 2007). Growing time was kept as short as possible for these cells to reduce the chance of this fungal contamination spreading to other cell lines in our laboratory or laboratory group. KDM14.9 control fibroblasts were also grown in the presence of Fungizone to control for any effects the addition of Fungizone may have had on the cells. Due to the fungal contamination, experiments to investigating the effect of this splicing variant on cell cycle progression using BrdU staining and FACS could not be performed, as the chance of spreading infection within the department was too big of a risk.

Sanger sequencing was used to confirm LB192 fibroblasts contained the c.1441-2A>G variant (Figure 4.4B). To do this LB192 cells were pelleted by centrifugation and high quality DNA was extracted (2.1.1.3) before being PCR amplified (2.1.2). Exonuclease I digest was used to remove additional nucleotides (2.1.2.3) before the PCR product was Sanger sequenced (2.1.4).

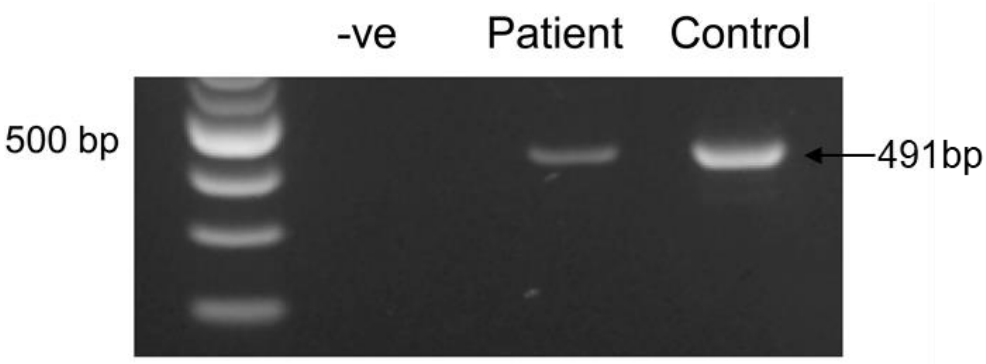


Figure 4.6 RT-PCR shows no obvious size difference due to splicing changes between patient and control fibroblasts. When RT-PCR amplification was kept under 30 cycles (shown above) the control band appears brighter than patient indicating less amplification from LB192. This was not seen with longer PCR cycles due to saturation. Other primer sets covering a larger area of *CDC45* were also used to look for any larger splicing changes, but none were observed (data not shown).

RT-PCR was performed to investigate potential splicing changes caused by the patient variant (2.1.7.4). RNA was extracted (2.1.7.1) and cDNA synthesised using oligonucleotide (dT)18 and random hexamer primers (2.1.7.2). We predicted RT-PCR may show all of exon 16 or a larger proportion of *CDC45* may be spliced out due to the loss of the exon 16 splice acceptor site. This was not observed (Figure 4.6) as both KDM14.9 and LB192 products were approximately the same size (predicted size was 491 bp). A range of primers were used to look for smaller and larger changes across the transcript by RT-PCR, but nothing was noted (data not shown), although the patient band did consistently

appear slightly dimmer than control. The RT-PCR products were extracted from the agarose gel and Sanger sequencing was performed to look for any small changes, as even a one base splicing change could lead to frameshift. Sanger sequencing showed a deletion of 6 bp at the start of exon 16 in the cDNA (Figure 4.7) causing the deletion of two amino acids but keeping the transcript in frame.

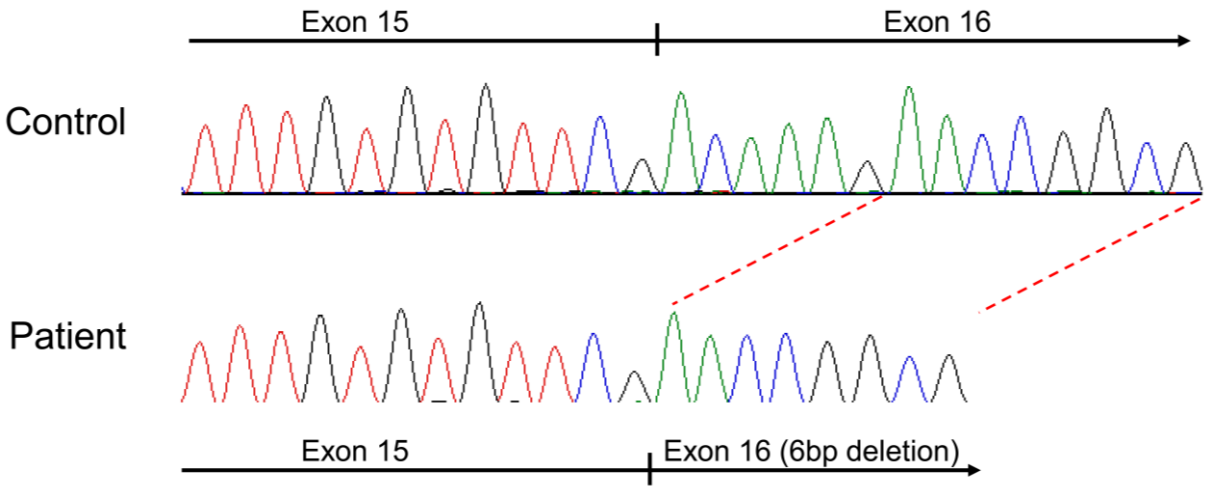


Figure 4.7 Sanger sequencing reveals 6 bp deletion at the start of exon 16 in patient cDNA. RT-PCR bands were gel extracted and Sanger sequenced to investigate small splicing changes. As observed LB192 patient cDNA showed a 6 bp in frame deletion at the start of exon 16.

4.2.2 Amino acid investigation

We examined the conservation levels to investigate the effect of this 6 bp deletion had at a nuclear and protein level. Five of the six deleted nucleotides were highly conserved between species (Figure 4.8A), as this deletion is at the splice site the high conservation of nucleotides is to be expected (Diamond et al., 1988; Wolter et al., 1988). The only nucleotide without high conservation is the third base of the first codon, this is expected as changing the last bp will not change the amino acid. Although the nucleotide conservation was high, Alamut did not rank the two deleted amino acids as highly conserved. The two amino acids were conserved between human to zebrafish and frog respectively (Figure 4.8B). This is a similar conservation to five out of nine missense variants from Fenwick et al., 2016.

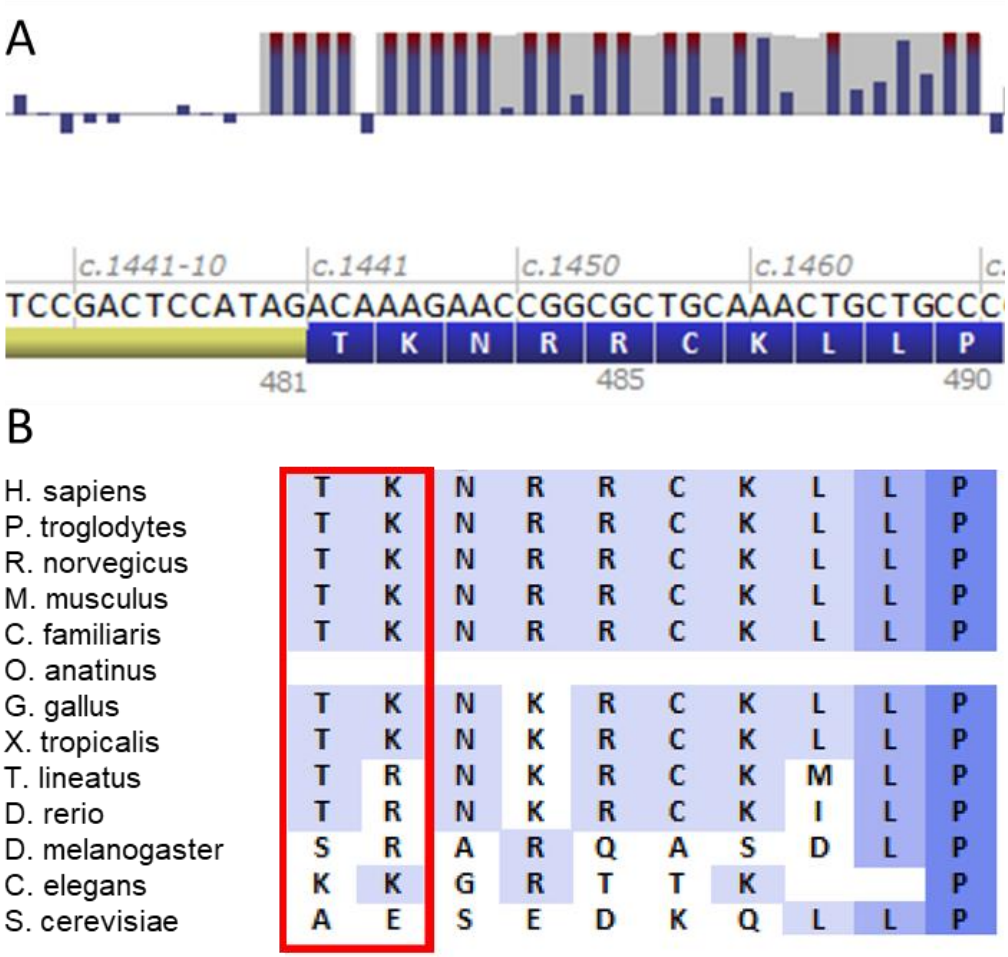


Figure 4.8 ***CDC45* exon 16 conservation.** *CDC45* c.1441-2A>G variant causes the deletion of the first 6 nucleotides of exon 16, leading to a two amino acid in frame deletion. A) Five of six deleted nucleotides show high conservation. B) Two deleted amino acids are not predicted to be highly conserved between species

Although not predicted to be particularly highly conserved, due to the severity of the patient phenotype where these amino acids sit in CDC45 in relation to the tertiary structure of the rest of the pre-initiation complex was investigated. To consider whether these amino acids could be important for interacting with other proteins in the preIC or replication fork leading to ineffective DNA replication. Protein schematics showed that the two deleted amino acids sit on the outside of CDC45 and do not seem to be important for interacting with other proteins in the preIC (Figure 4.9).

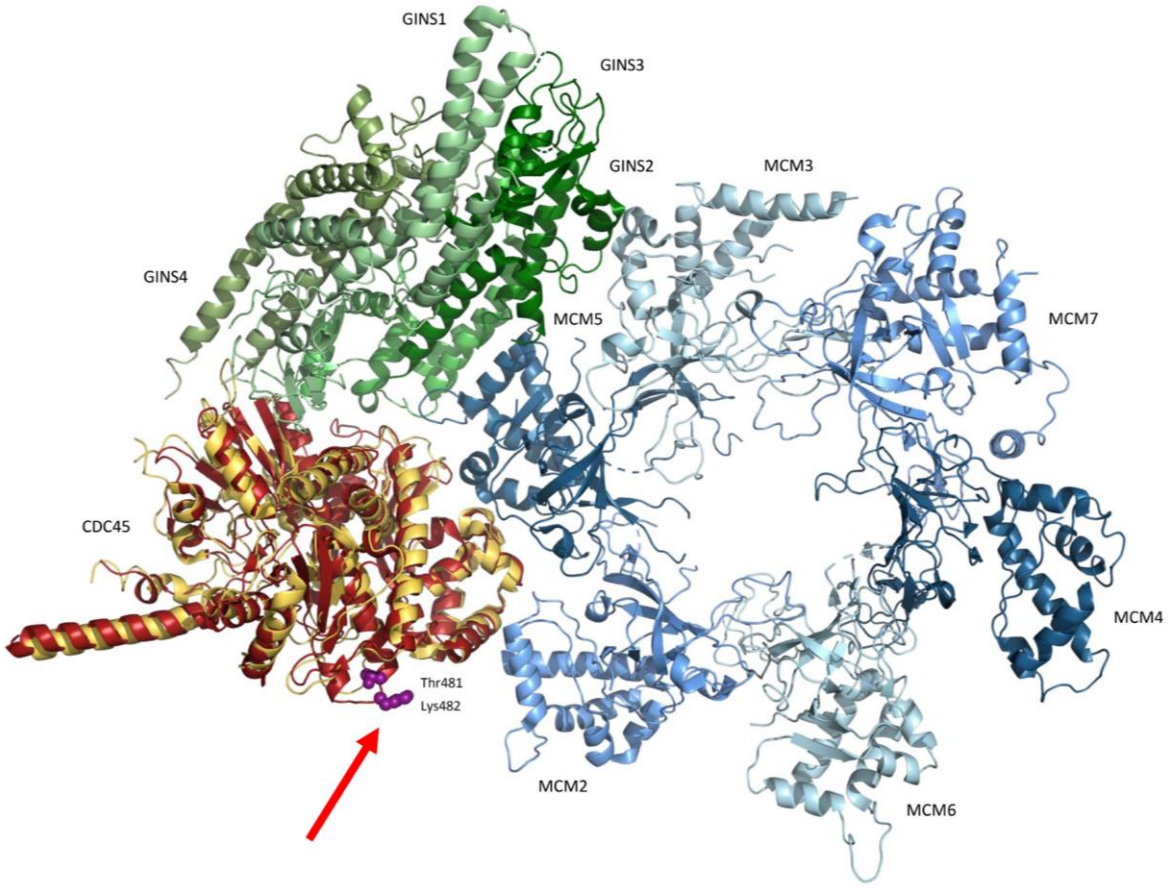


Figure 4.9 CDC45 protein interaction with other members of the pre-initiation complex. Schematic of CDC45 interacting with other members of the pre-initiation complex shows Human CDC45 (red) overlaid *S.cerevisiae* CMG complex structure (PDB ID: 3JC6), with the two deleted these amino acids (purple) which do not seem to be required for interacting with other proteins. Red arrow indicated position of two deleted amino acid residues (credit Dr Karen Knapp).

4.2.3 CDC45 transcript levels in patient-derived cells

While the splicing machinery was making use of a cryptic splice site 6 nucleotides into the exon, the efficiency of this splicing was not clear, and it was hypothesized that it was less efficient as bands produced from RT-PCR were less intense (Figure 4.5). Due to the severe phenotype seen in both siblings, we expected that this cryptic splice site may be leaky, leading to only a small proportion of mRNA being made into cDNA and the rest being degraded by nonsense mediated decay (Brogna and Wen, 2009). Therefore it was hypothesised that this ineffective splice site would likely be causing a reduction in stable mRNA expression, leading to the patient phenotype.

The reduction in CDC45 expression was confirmed by RT-qPCR using primers >200 bp apart, located across the transcript. This showed a severe knockdown (72.9%) of CDC45

expression across the gene compared to control (Figure 4.10). RNA was extracted and cDNA synthesised from control (KDM14.9) and patient (LB192) fibroblasts on three separate occasions. Three technical RT-qPCR replicates of each reaction were performed for each of the three biological replicates (appendix C.2). CDC45 expression levels were measured using the comparative CT method ($2^{-\Delta\Delta}$CT)-relative to *RPS18*. Patient values were normalised to control values set at 1.

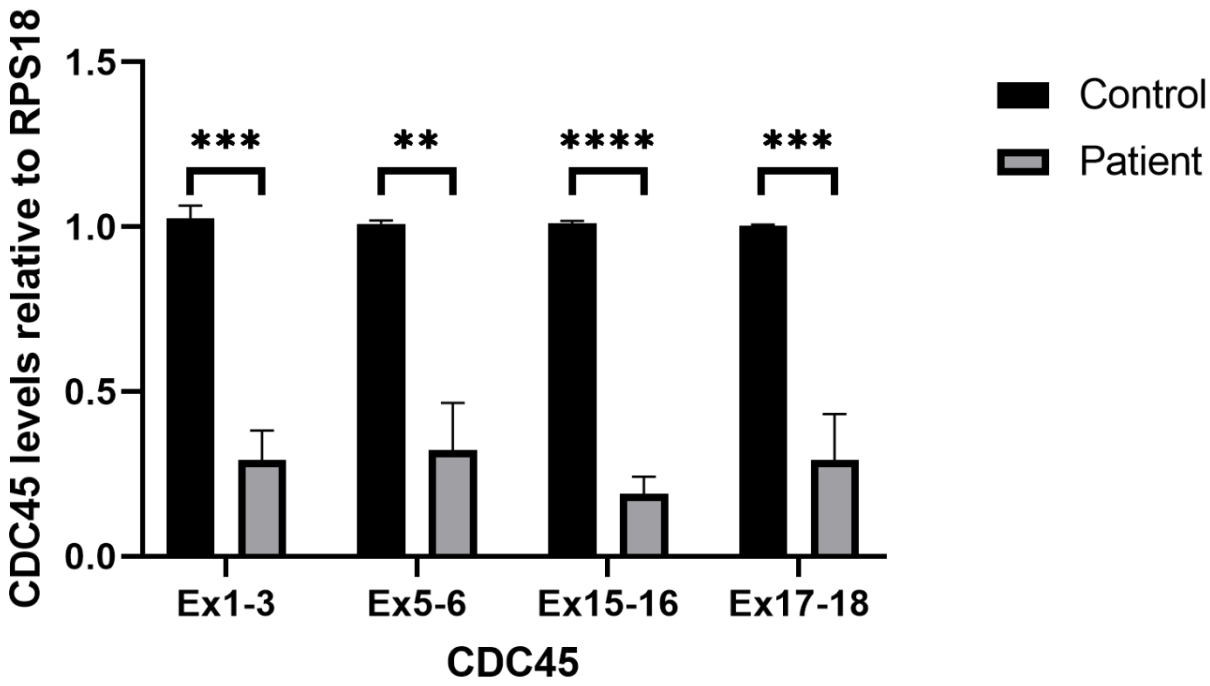


Figure 4.10 ***CDC45* splicing variant causes a significant reduction in CDC45 expression in fibroblasts.** Makeshift splice site causes 72.9% reduction in CDC45 expression across the gene. RT-qPCR $2^{-\Delta\Delta ct}$ values normalised to RPS18 are plotted from control (KDM14.9) and patient (LB192), RT-qPCR performed in triplicate with three biological replicates performed. A 72.9% reduction in CDC45 expression was seen across the gene. Multiple T-tests showed significant reduction in CDC45 expression (** <0.01, *** < 0.001, **** <0.0001)

Cooper, 2016 and Fenwick et al., 2016 saw a 93% reduction in CDC45 levels in a pair of sibling's presenting with a compound heterozygous splicing variant alongside a missense variant. These siblings present with a much milder phenotype than the siblings understudy here do and yet they have a greater reduction in CDC45 expression levels. Due to this we decided to investigate if there was an effect on protein stability due to the two amino acid deletion. As mentioned previously other previously reported CDC45 patients with splicing variants have had a similar or larger reduction in CDC45 expression but a

far less severe phenotype. Due to this we decided it was important to follow up other potential causes of the phenotype.

4.2.4 Immunoblot

A western immunoblot was performed using patient (LB192) and control (KDM14.9) fibroblasts, as well as untransfected HEK293FT cells and HEK293FT cells transfected with a flag-tagged-CDC45 plasmid encoding the canonical protein.

The endogenous signal was undetectable due to low protein expression levels in fibroblasts. The same CDC45 antibody has successfully been used to detect protein levels in LCL cell lines (Fenwick et al., 2016), which have 21 fold higher protein expression level than fibroblasts (https://www.gtexportal.org/) explaining why expression was not detected in this assay.

As the antibody gave a weak signal for transient CDC45 in HEK293FT cells (data not shown) we decided to use a transient transfection of engineered plasmids encoding either wild type or mutant protein to test relative levels by western blot analysis.

Site-directed mutagenesis was used to introduce the 6 bp cDNA deletion into a flag tagged CDC45 vector (made previously by Dr Karen Knapp). Following transformation of the mutagenesis PCR product into *E. coli*, plasmids were purified and subjected to Sanger sequencing to confirm successful mutagenesis.

HEK293FT cells were transfected (2.3.7) with the 6 bp deletion CDC45 vector or the WT flag-CDC45 vector, 24 hours before protein extraction (2.1.8). Immunoblotting was performed to look for potential differences in protein stability (Figure 4.11).

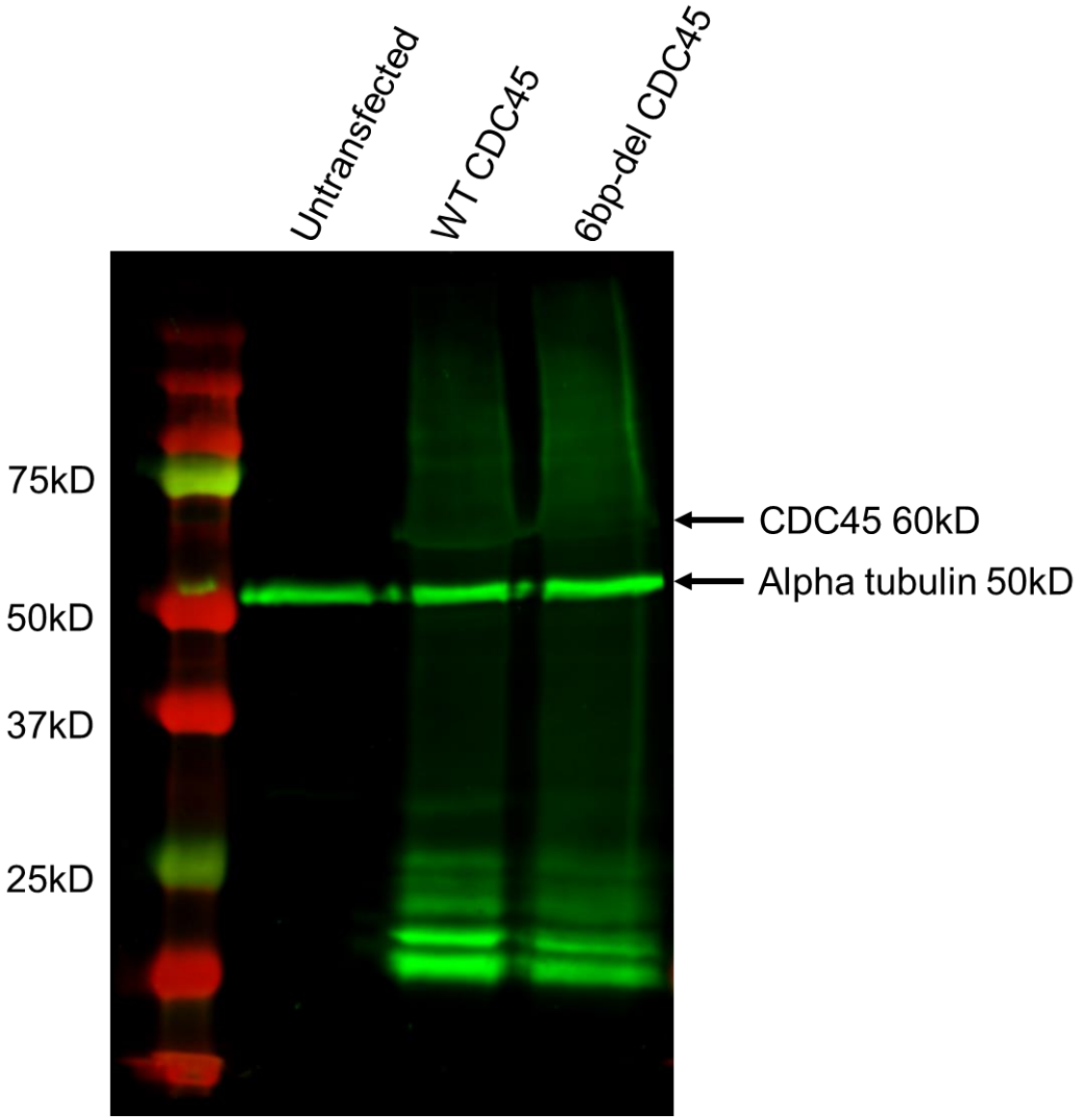


Figure 4.11 CDC45 protein levels are similar between WT and Patient. Immunoblot with anti-CDC45 shows little difference in CDC45 protein expression between HEK293FT cells transfected with WT or 6 bp-del CDC45. Alpha tubulin (50 kD) confirmed consistent loading.

Unfortunately no clean band was observed for either the WT or 6 bp deletion CDC45 (60 kD). Because of this, it wasn't possible to accurately quantify protein expression levels (Figure 4.11). Therefore this experiment should be repeated before any conclusions can be drawn. The membrane was counterstained with alpha tubulin to confirm equal loading of protein lysates

4.2.5 Discussion

While previously described CDC45-MGS patients have often presented with craniosynostosis – a slightly different phenotype to other MGS patients – the phenotype is not generally more severe. Here we reported two severely affected siblings who both inherited a homozygous splicing variant in *CDC45*. Due to the severity of the phenotype it was hypothesised that this splicing variant may lead to an almost complete reduction of gene expression. RT-PCR and RT-qPCR analysis showed a 72.9% knockdown of CDC45 function with 100% of the mRNA expressed containing a 6 bp deletion.

The first major limitation of this study was the fungal contamination present in the patient fibroblasts when they arrived. Due to this contamination we were unable to grow the fibroblasts for long periods of time, to reduce the chance of spreading the infection. This meant we were unable to investigate the effect of the mutant protein at a cellular level. Other experiments could have included EdU pulse chase assay to investigate the effect of this variant on cell cycle progression or fibre combing to examine the effect of DNA replication events and speed. This would have allowed us to gain a better understanding of the overall effect of this splicing variant at a cellular level.

Expression levels vary between cell types and developmental stages (Cardoso-Moreira et al., 2019). The second limitation of this experiment was the difference in age between the patient and the control fibroblasts and lack of other CDC45 patient fibroblasts. This study compared CDC45 expression levels between fibroblasts from a 20-week gestation foetus and a 14 year, 9-month-old boy, as this was the youngest cell line, we had access to. A younger control cell line would likely give a more accurate representation of the difference in CDC45 expression. Due to the age of the patient (20 weeks gestation) accurately matching a control cell line is realistically not possible but repeating this experiment with a younger control cell line (ideally <1 year old) may be beneficial in obtaining more meaningful results. It is expected that CDC45 would have higher expression during development (Shaikh et al., 1999). Specifically earlier in development when increased DNA replication is required to keep up with the large amount of cell division occurring during a short timeframe for brain development (Kalogeropoulou et al., 2019). Although this period had likely passed it would still be expected to see an increase in CDC45 expression levels compared to non-foetal samples. Access to other CDC45 patient fibroblasts would allow comparison of both splicing and protein

expression differences between patients presenting with a severe phenotype and patients presenting with a moderate CDC45 phenotype.

Cooper, 2016 and Fenwick et al., 2016 saw a 93% reduction in full length CDC45 expression in one of their patients with compound heterozygous splicing variants (c.203A>G, p.Gln68Arg (splicing effect) & c.333C>T, p.Asn111 = exon skipping effect). While this reduction in expression is more extreme than what was observed here, it is hard to accurately compare this knockdown in expression as they were using different cell lines and different housekeeping genes. In their study Cooper and Fenwick used patient derived LCLs to investigate knockdown in full length CDC45 expression. As mentioned earlier CDC45 expression levels vary between different cell lines. In the previously mentioned study *GAPDH* was used as a housekeeping gene, *GAPDH* although widely used as a control gene can have variable expression in LCLs and therefore is not necessarily the most accurate or stable control gene (Brouwer et al., 2006). In fibroblasts *GAPDH* is considered one of the most unstable genes in terms of expression (Panina et al., 2018), therefore *RPS18* was used for normalising RT-qPCR data. All of this can lead to differences in results making it hard to accurately compare the knockdown of CDC45 expression between these two patients. The difference in age between their CDC45 patient and control cell line is unknown, but it is likely a more accurate control than our control cell line as they were using LCLs derived from a 7 year old (Cooper, 2016; Fenwick et al., 2016).

The (c.203A>G, p.Gln68Arg & c.333C>T, p.Asn111=SE) compound heterozygous splicing variants were seen in both the patient mentioned above and his younger brother. Interestingly, although both had the same compound heterozygous splicing variants in CDC45 the older sibling presented with a relatively mild phenotype while the younger sibling had a more severe phenotype. The older sibling was born at 42 weeks gestation; length (-1.4 SD), weight (-1.5 SD). Measurements at the most recent examination (7 years old) were, height (-1.3 SD), OFC (-2.3 SD) and weight (-0.4 SD). He also presented with microtia, absent patella, thin eyebrows (associated with MGS), high arched palate and a mircopenis (associated with MGS). He did not present with developmental delay or craniosynostosis. His younger brother was born at term; length (+0.5 SD), OFC (-1.7 SD), weight (-1.2 SD). Measurements at his most recent examination (16 months) were, height (-2.5 SD), OFC (-4.7 SD) and weight (-3.7 SD). MGS phenotypes included microtia and thin eyebrows but no absent patella. Similar to his older sibling he presented with no

developmental delay, however he did present with bicoronal craniosynostosis with widely patent metopic and sagittal sutures associated with CDC45 variants. CDC45 expression levels were not investigated in this patient but presumed to be similar to the first, as both presented with the same *CDC45* splicing variants.

The difference in craniosynostosis and MGS phenotypes between these two CDC45 patients demonstrates incomplete penetrance. While these two siblings presented with a more severe knockdown of CDC45 expression compared to the two CDC45 patients presented in this study, they showed a much milder phenotype. The splicing variants in these patients from Fenwick et al., 2016 affected the 5′ end of the gene while the splicing variants in this study affect the 3′ end of the gene. In the case of our CDC45 affected siblings we see a very similar severe phenotype in both siblings, this may be due to potential protein effects as well as the splicing effect.

Although the Fenwick et al., 2016 patients showed a 93% knockdown in CDC45 expression the remaining 7% mRNA expressed encoded the full length WT CDC45 transcript. In this study, the remaining 20% mRNA expression did not encode the full length WT transcript but a transcript containing a 6 bp deletion at the start of exon 16, causing a two amino acid deletion. This could contribute to why we see such a severe phenotype in these patients.

This 6 bp deletion leads to an in frame two amino acid deletion in the protein. These amino acid residues were conserved to a similar level as 5/9 of the missense variants described in Fenwick et al., 2016, while the remaining four missense variants were more highly conserved.

Western immunoblot was used to investigate relative protein levels, as a decrease in protein levels may help explain the extreme phenotype seen in these patients. As seen above there was no obvious difference is CDC45 expression in transfected HEK293FT cells. Comparing protein expression in transfected HEK293FT cells is not the ideal way to look at protein expression but unfortunately CDC45 expression in fibroblasts was too low to give a signal, even in WT cells. While this experiment suggests that the 6 bp deletion may not influence protein stability we cannot rule out it may affect the protein in other ways. For example, this deletion may lead to a reduced ability of CDC45 to interact with other proteins involved in DNA replication.

Protein interactions impacted by this two amino acid deletion CDC45 was not something we were able to investigate during this project. We would also have liked to investigate the effect this splicing variant had on cell cycle progression, but unfortunately due to the fungal infection in the cell lines we could not investigate this.

After reporting these functional results back to our collaborators, they were not entirely convinced this 72.9% reduction in CDC45 expression would be enough to explain the severe phenotype. Due to this they wanted to re-examine the exome sequencing to look for other potential variants shared by both siblings which could be contributing to the phenotype. At the time of submission, nothing had been reported back. However as we could not analyse this sequence ourselves, we cannot rule out the potential of other variants or risk alleles which may be contributing to the phenotype.

While where this deletion in CDC45 sits in relation to other proteins in the preIC was investigated, we did not investigate how this deletion may affect the role of CDC45 plays during replication, specifically relating to elongation and leading strand displacement. This ties into the next section of our project as DONSON's role in DNA replication was only recently discovered (Evrony et al., 2017; Reynolds et al., 2017). Exactly how and which proteins DONSON interacts with in the replisome is uncertain. It is therefore possible this small deletion in CDC45 could lead to issues during DNA replication that is not understood yet.

4.3 DONSON sub cellular localisation

DONSON works during S-phase to stabilise replication forks (Figure 4.12). Previous research by Reynolds et al., 2017 found that *DONSON* missense variants causing the MISSLA phenotype (section 1.2) often resulted in the mislocalisation of DONSON-GFP protein in the cytoplasm of HeLa cells rather than in the nucleus where WT DONSON localises (Figure 4.13). HeLa cells transfected with the wild type pEGFP-DONSON vector showed only nuclear expression while the majority of the MISSLA patients showed diffused localisation in the cytoplasm (Figure 4.13). The exception to this was p.S28R & K489T variants which showed mainly nuclear expression, here western blot analysis was used to show p.K489T decreased protein stability, causing marked reduction in protein levels (Reynolds et al., 2017). Although the MISSLA phenotype is different to the MGS phenotype the inability of the DONSON protein to localise to the nucleus would be a strong

indication that these variants are pathogenic. We reasoned this would be a good first experiment to examine whether the p.R211C variant was likely pathogenic before we embarked on CRISPR engineering focusing on this variant.

At the start of this project *DONSON* was considered a novel disease gene associated with MGS, as it had only been described to be involved in MISSLA (Reynolds et al., 2017) and MIMIS (Evrony et al., 2017). Therefore we wanted to investigate the cellular localisation of pEGFP-DONSON in HeLa cells, as the inability for DONSON protein to localise to the nucleus where WT DONSON is expressed would give us more confidence that this variant is pathogenic.

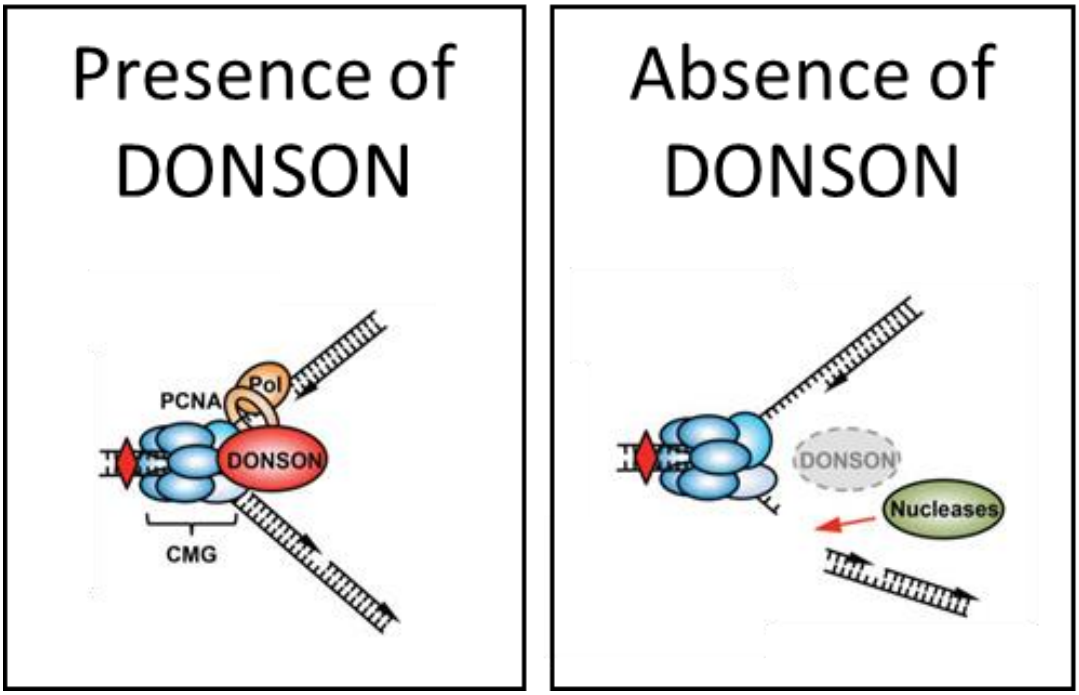


Figure 4.12 DONSON is required for replication fork stability in S-phase. In the presence of DONSON replication forks are stabilised and restarted after replication stress. In the absence of DONSON replication stress leads to double-stranded breaks. This leads to chromosomal instability and eventual cell death. Figure adapted from (Reynolds et al., 2017) Supplementary Figure 22, copyright obtained from Springer Nature - 4850530174277.

As mentioned previously DONSON works during S-phase to stabilise replication forks. This is important for chromosomal stability as without sufficient functioning DONSON these replication forks can collapse leading to chromosomal instability and cell death (Figure 4.12). Patient genotypes suggest hypomorphic variants (significantly reduced levels) as *DONSON* knockouts are embryonic lethal (Dickinson et al., 2016; Reynolds et al., 2017).

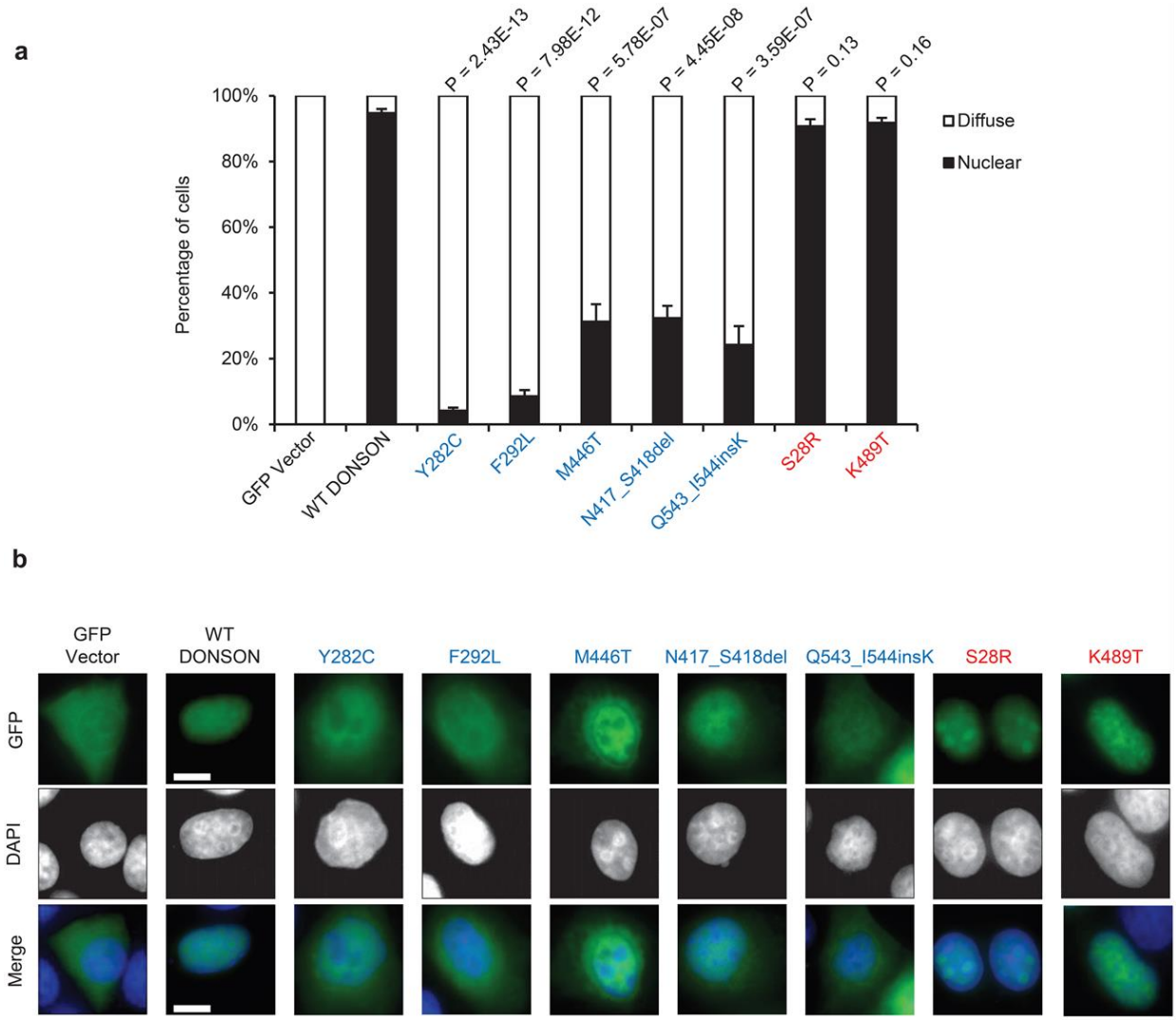


Figure 4.13 DONSON cellular mislocalisation of MISSLA variants A) Proportion of transfected HeLa cells with diffuse or cytoplasmic p.EGFP-DONSON expression fixed 24 hours post transfection (n = 5 independent experiments). B) Phenotype of pEGFP-DONSON variants as well as empty and WT controls. Figure from (Reynolds et al. 2017) Supplementary Figure 9, copyright obtained from Springer Nature 4850530174277.

4.3.1 DONSON-p.R211C Localisation

We were kindly sent the empty pEGFP vector along with WT DONSON and the seven vectors containing MISSLA patient variations from our collaborator Professor Grant Stewart, University of Birmingham.

When transfected in HeLa cells, empty-p.EGFP is diffuse in the cytoplasm while pEGFP-WT DONSON is localised to the nucleus, as this is where it works to stabilise replication forks when they stall during S-phase. WT DONSON has two main phenotypes; diffuse in the nucleus (Figure 4.14 A) or small uniform spots (Figure 4.14B) when DONSON is acting to stabilise replication forks (Figure 4.14).

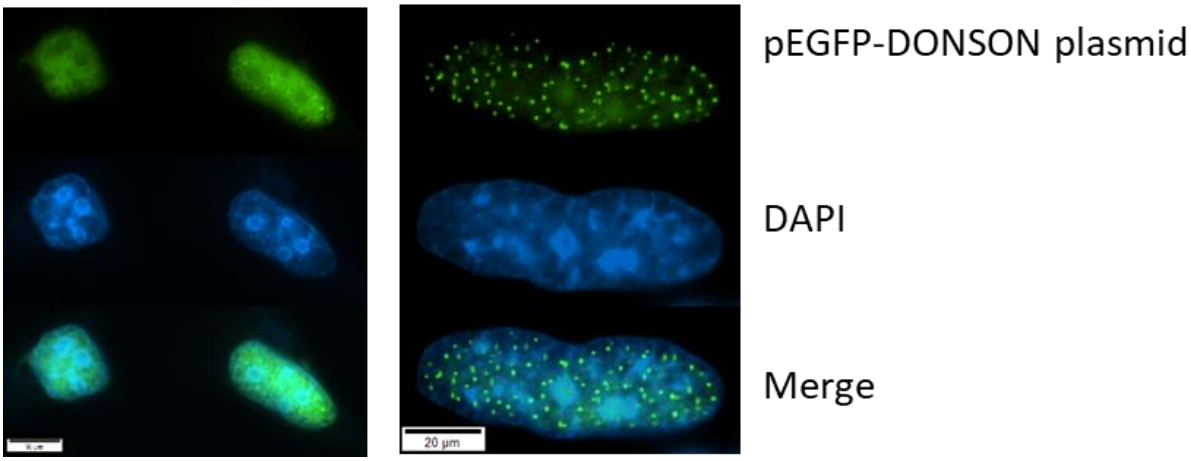


Figure 4.14 pEGFP-WT-DONSON is localised to the nucleus of HeLa cells. DONSON protein is either diffuse in the nucleus or localised to replication foci. Note there is normally no localisation in the cytoplasm.

Site-directed mutagenesis was used to introduce LB008 c.631C>T missense variant into the pEGFP-*DONSON* vector. Once Sanger sequencing confirmed successful mutagenesis (Figure 4.15), HeLa cells plated on glass coverslips were transfected with this plasmid. Cells were fixed 24 hours post transfection and stained with DAPI before being mounted and viewed under the Olympus BX53 upright microscope.

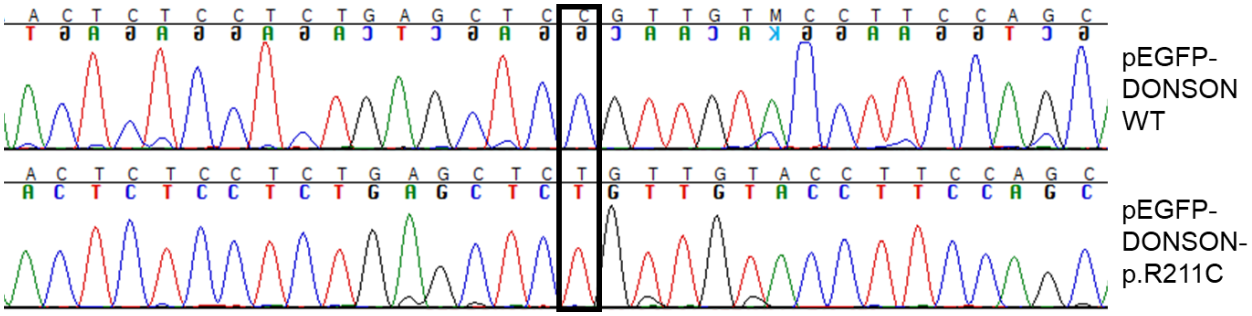


Figure 4.15 Sanger sequencing confirms successful mutagenesis of pEGFP-*DONSON* vector.

Preliminary transfection tests with MISSLA variants showed similar results (Reynolds et al., 2017) (Figure 4.13) with majority of cells showing cytoplasmic expression (data not shown). MISSLA variant p.F292L variant was used as a positive control for all experiments.

Only a few cells transfected with the pEGFP-DONSON-p.R211C plasmid showed cytoplasmic expression (Figure 4.16). Immediately the most noticeable p.R211C phenotype was the number of cells containing bright GFP aggregates (Figure 4.17). While some of these aggregates were occasionally seen in WT (Figure 4.14/4.18) there were generally less aggregates and less cells expressing these aggregates. Such condensations were rarely seen with WT-DONSON and other MISSLA patient plasmids, it is likely these were ignored or presumed to be DONSON acting at replication forks in previous experiments. This GFP-aggregate phenotype was not seen at all with the p.EGFP-empty vector, suggesting this observed aggregate phenotype was due to DONSON and not the GFP protein.

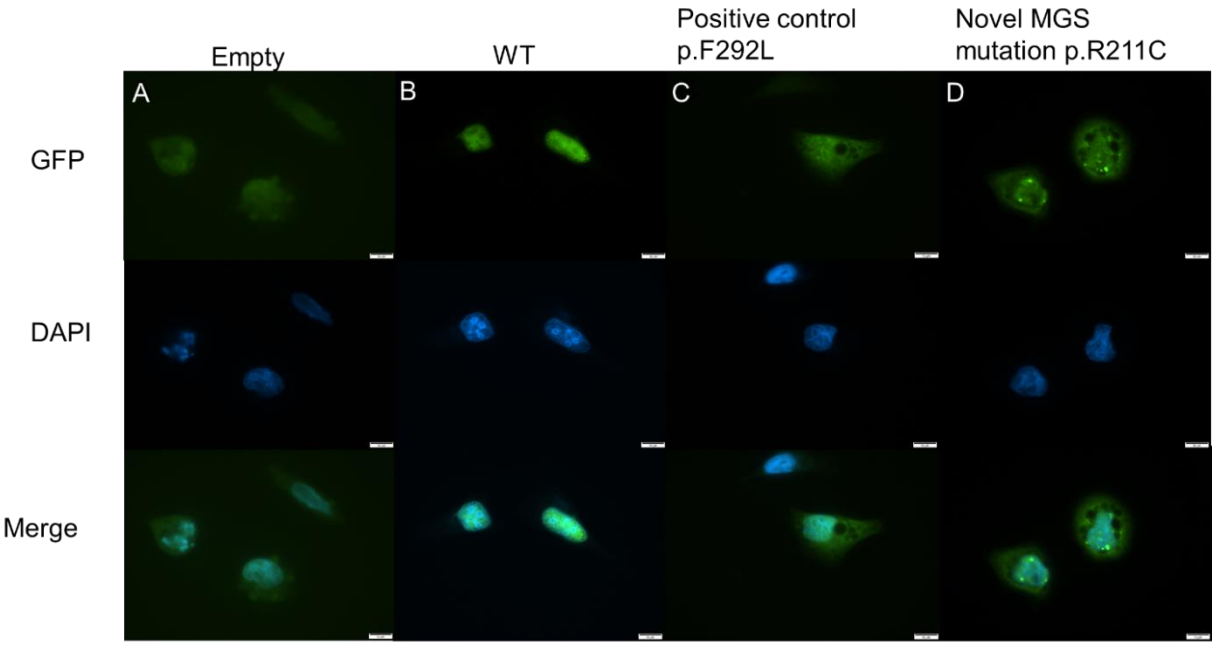


Figure 4.16 Cellular localisation of pEGFP empty or pEGFP-DONSON plasmids transiently expressed in HeLa cells fixed 24 hours post transfection. Novel p.R211C sometimes showed cytoplasmic expression. A) pEGFP-empty shows only cytoplasmic expression. B) WT-DONSON shows only nuclear expression. C) pEGFP-DONSON-p.F292L positive control MISSLA variant from Reynolds et al., 2017 shows mostly cytoplasmic expression. D) Some pEGFP-DONSON-p.R211C HeLa cells show cytoplasmic expression but not the majority.

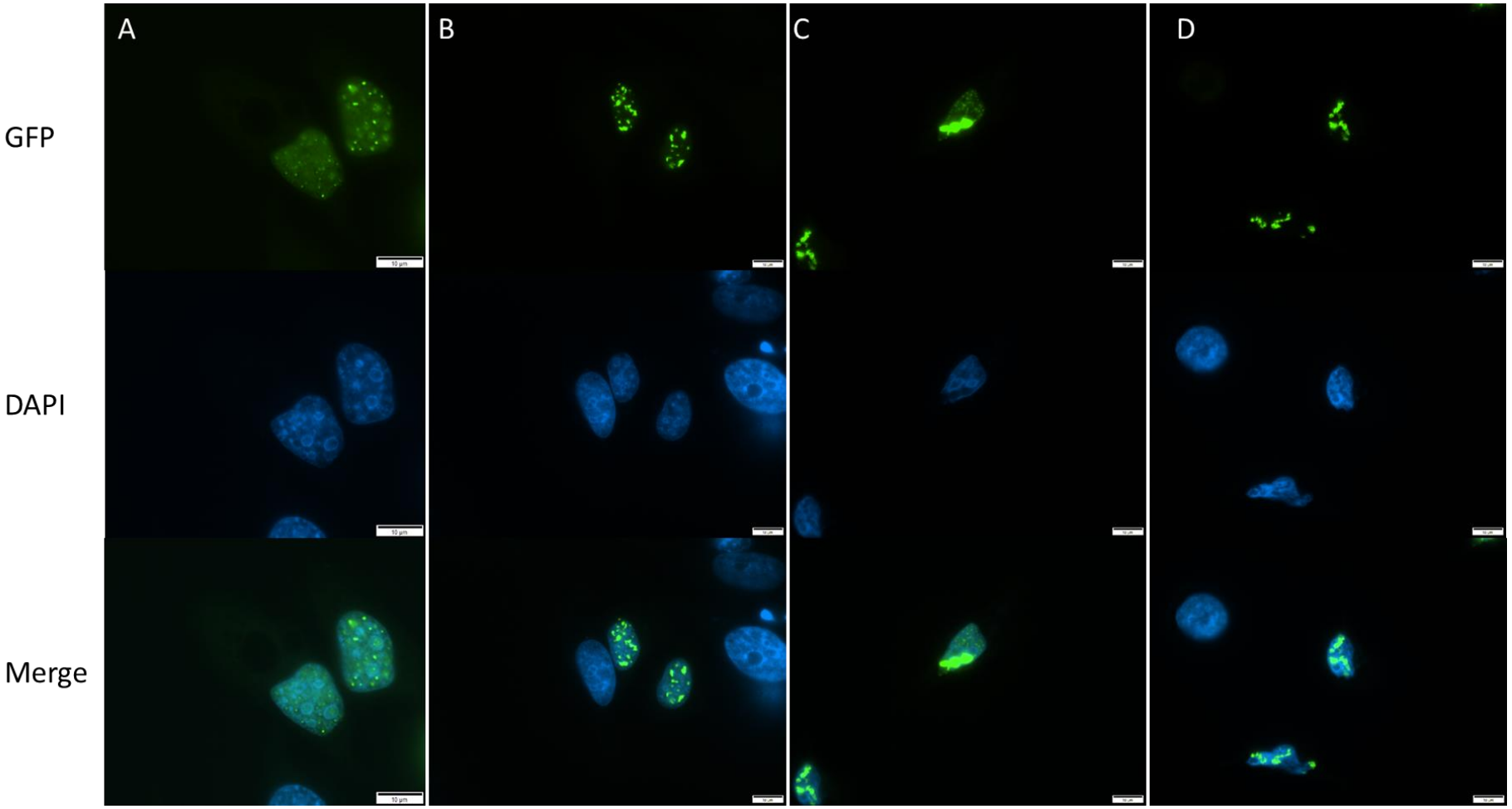


Figure 4.17 p.R211C aggregate phenotype is different from the speckled phenotype seen in some WT cells when DONSON is active at replication forks in S-phase. GFP and DAPI fluorescent images were taken of HeLa cells transfected with p.R211C DONSON and images merged. A) p.R211C transfected cells occasionally showed normal looking WT spotty phenotype. B-D) Majority of p.R211C DONSON transfected cells showed a more severe aggregate phenotype.

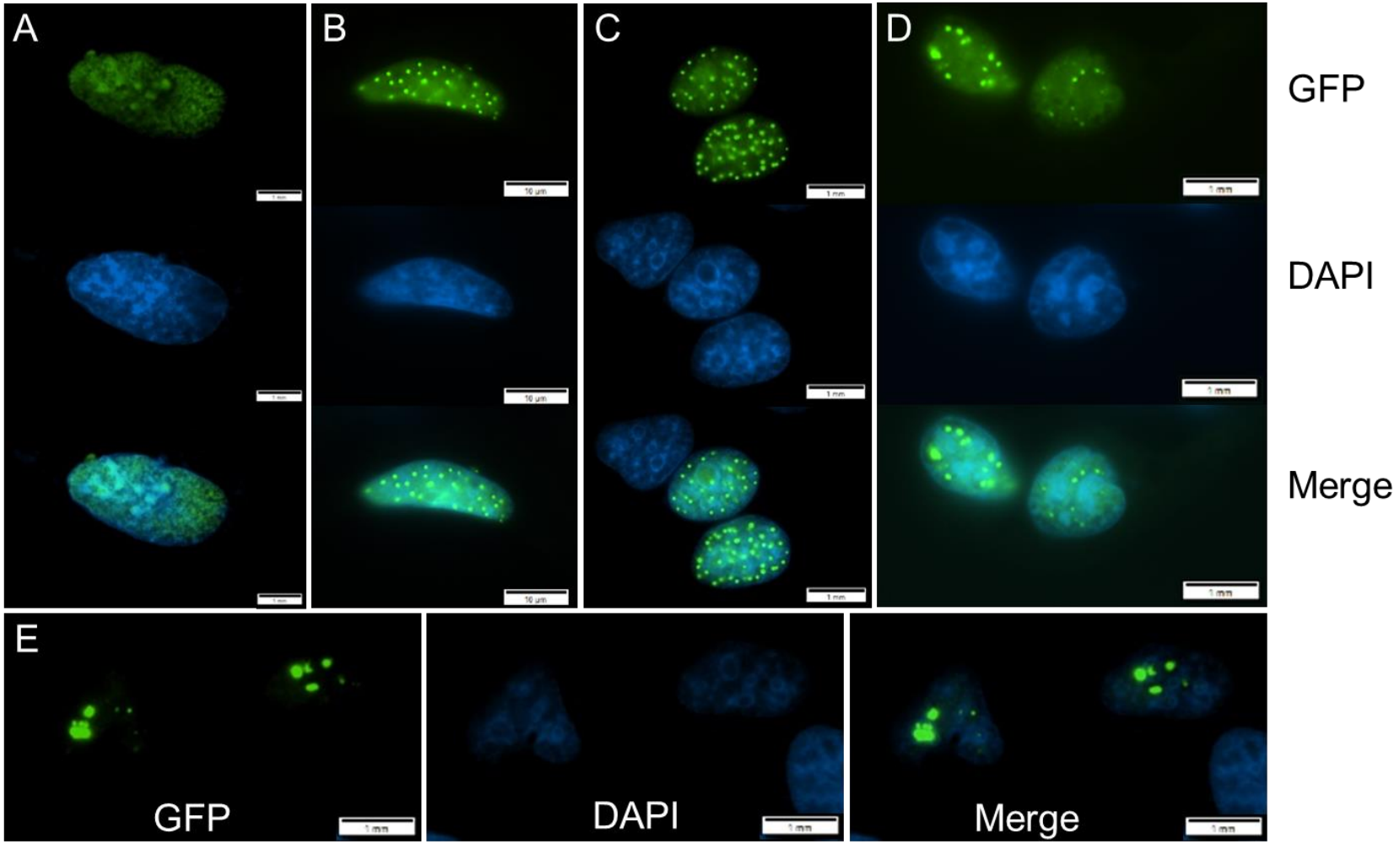


Figure 4.18 WT-DONSON transfected in HeLa cells show diffuse to slightly aggregate phenotype. A) Diffuse nuclear expression is seen in majority of WT-DONSON cells. B and C) uniform spotty phenotype is seen in WT-DONSON cells during S-phase when DONSON is required for stabilising replication forks. D) Slightly aggregate phenotype seen in some S-phase WT-DONSON cells. E) A small proportion of WT cells showed aggregate phenotype generally less severe than seen with p.R211C.

While occasionally we saw an aggregate phenotype in WT cells it was generally less severe than what is seen in p.R211C and far less cells showed this phenotype. The majority of WT cells which did not have diffuse nuclear expression had a uniform spotty phenotype (Figure 4.18 B-D).

This difference in phenotype observed between p.R211C and WT gave us confidence this *DONSON* variant was worth investigating further by establishing isogenic cell lines using CRISPR as described in the previous chapter (3.0).

4.3.2 Quantifying DONSON localisation

Next we quantified the p.R211C pEGFP-DONSON expression based on nuclear (WT), cytoplasmic (empty) or this new aggregate phenotype. By this point we had a total of four DONSON-MGS patients with five missense changes and one splicing variant between them (Table 4.1, Figure 4.19).

Table 4.1 MGS-*DONSON* patient variants, inheritance and population frequency (Table previously published in Knapp et al., 2019).

Patient	Variant 1	gnomAD MAF	Variant 2	gnomAD MAF	Segregation
P1	c.631C>T, p.Arg211Cys	NA	c.631C>T, p.Arg211Cys	NA	Both parents heterozygous.
P2	c.494T>C, p.Phe165Ser	0.00001768	c.607–36G>A (splicing)	0.00005540	Mother heterozygous for c.706–36G>A, homozygous reference for c.494T>C, p.Phe165Ser.
P3	c.1634C>T, p.Pro545Leu	0.00001061	c.809A>G, p.Tyr270Cys	0.00001599	No parents available.
P4	c.670C>T, p.Pro224Ser	0.00006369	c.809A>G, p.Tyr270Cys	0.00001599	No parents available.

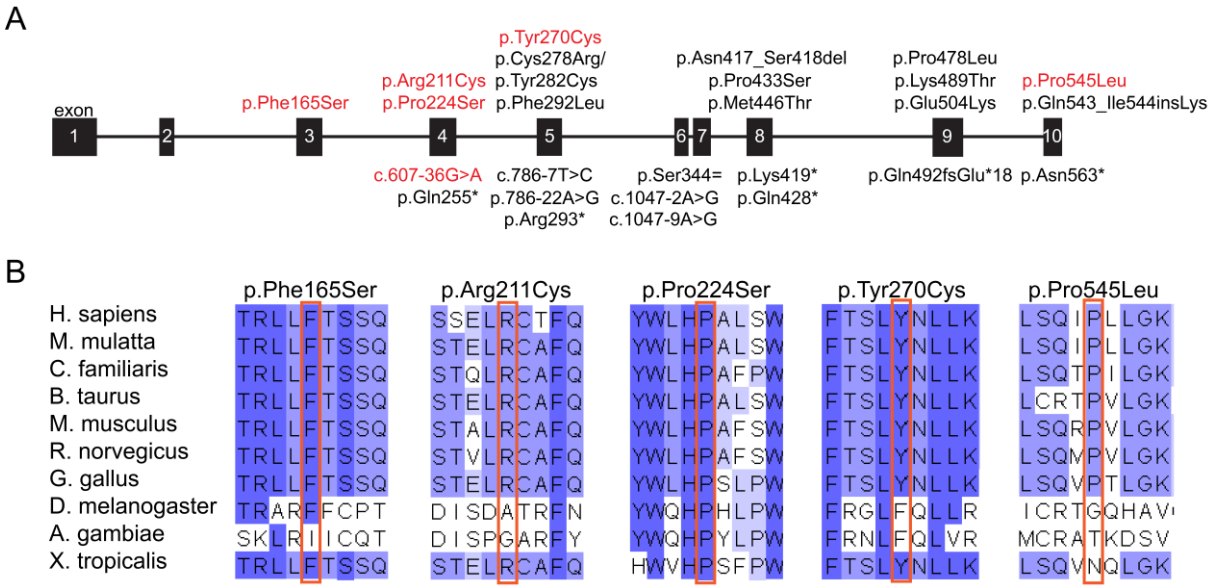


Figure 4.19 Four of our MGS patient cohort (~25%) contain 5 missense and one splicing variant in *DONSON*. A) MISSLA and MIMIS variants shown in black tend to cluster around the C-terminus of the protein while all but one MGS-*DONSON* variant are found at the N-terminus. B) All 5 missense variants are found in relativity highly conserved regions of the protein. (Figure previously published in Knapp et al., 2019).

Dr Karen Knapp used site-directed mutagenesis to introduce the other four missense variants into the pEGFP-DONSON vector. Five out of the six *DONSON*-MGS variants clustered at the N-terminus of the protein while MISSLA and MIMS variants were located more towards the centre and C-terminus of the protein. All DONSON-MGS missense variants were in relatively highly conserved regions of the protein (Figure 4.19A).

HeLa cells were transfected with the five MGS-DONSON variant plasmids, two MISSLA variant positive controls (p.F292L and p.Iso543_Pro544insLys), empty pEGFP or WT-DONSON. A total of ~1200 cells were counted for each plasmid across five transfection replicates (Figure 4.20)

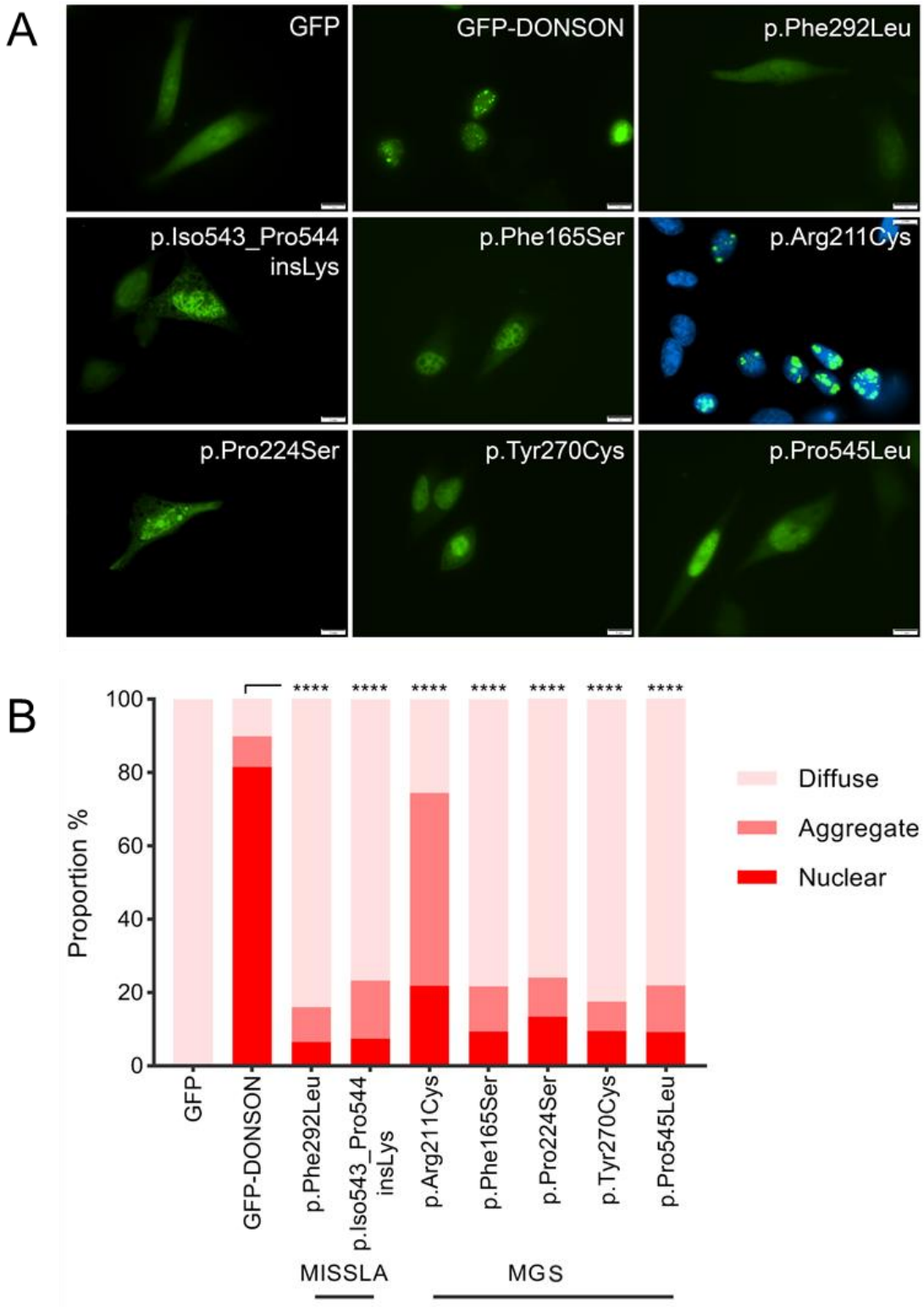


Figure 4.20 DONSON subcellular localisation comparing DONSON-MGS variants to WT and MISSLA variant controls. A) GFP Representative images of GFP, WT-DONSON, the five DONSON-MGS variants (p.Phe165Ser, p.Arg211Cys, p.Pro224Ser, p.Tyr270Cys and p.Pro545Leu) and two MISSLA control variants (p.Phe292Leu and p.Iso543_Pro544insLys) transfected in HeLa cells. p.Arg211Cys is co-stained with DAPI to show aggregate nuclear phenotype different to other variants which show diffuse cytoplasmic expression. B) Quantification of DONSON subcellular localisation of nine previously mentioned plasmids (a) (n=1200 cells/plasmid over five biological replicates, one-way ANOVA allowing for multiple comparisons). Figure previously published in Knapp et al., 2019.

Table 4.2 Subcellular expression percentages for MGS-DONSON variants and controls.

Plasmid	Nuclear localisation	Aggregate localisation	Cytoplasmic localisation
Empty	0.08	0	99.9
WT	80.9	8.9	10.2
p.F292L	5.4	10.1	84.5
Ins-K	6.6	15.5	77.9
p.R211C	23.3	50.3	26.4
p.F165S	9.1	13.1	77.8
p.P224S	9.1	10.7	80.1
p.Y270C	7.5	8.0	84.4
p.P545L	7.9	13.8	78.3

All MGS-DONSON patients showed similar results to what was seen in Reynolds et al., 2017, with the majority of cells showing a cytoplasmic expression (Figure 4.20). Due to the cytoplasmic expression, it is thought that perhaps the variants in DONSON are resulting in the majority of the DONSON protein no longer being able to successfully localise to the nucleus to act stabilising replication forks, leading to the patient phenotype. The exception to this was the p.R211C plasmid, mentioned above. The majority of these cells showed an aggregate phenotype seen rarely in WT and other DONSON variants (Figure 4.20/4.21). WT DONSON shows a speckled phenotype, reflecting replication foci when DONSON is active stabilising replication forks. This is generally seen as relatively uniform small spots in the nucleus (Figure 4.14) (Reynolds et al., 2017). The aggregate phenotype seen is much larger clumps of GFP-DONSON. What could be causing these aggregates is unknown, as it doesn't appear to be nucleolar or P bodies (Ling et al., 2014; Luo et al., 2018; Panico and Forti, 2013). The p.R211C variant generally showed a less extreme cytoplasmic expression than other variants and had a higher proportion of nuclear localised cells than any other pathogenic variant (Figure 4.21).

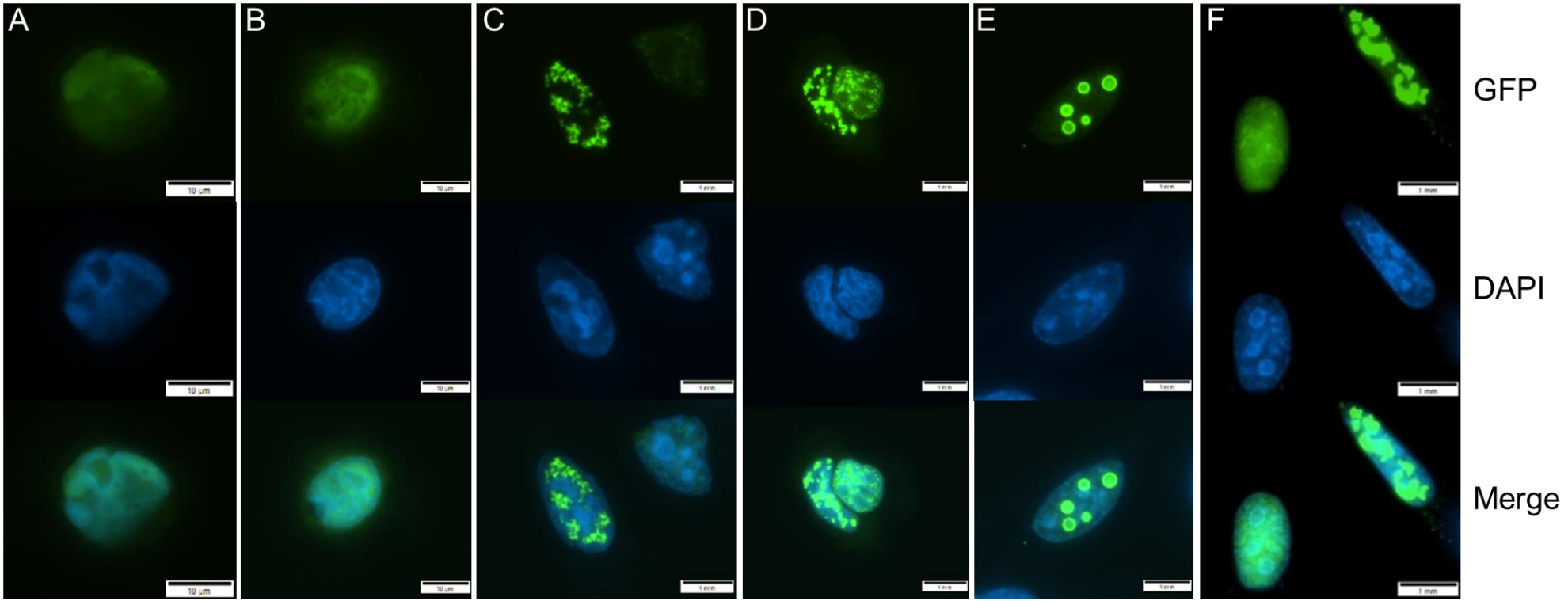


Figure 4.21 p.R211C mostly showed aggregate expression. A-B) cytoplasmic expressing p.R211C cells, only 26.4% of p.R211C cells showed cytoplasmic expression. C-F) Different aggregate phenotypes seen in p.R211C transfected cells.

4.3.3 DONSON in Mitosis

Further immunocytochemistry experiments were performed after ending the CRISPR experiments early. This gave us the opportunity to follow up the aggregate GFP-DONSON phenotype in more detail. This aggregate phenotype was seen in 50.3% of p.R211C cells but only 8.0-15.5% of other variants. As shown in Figure 4.18 there was some cross over between normal spotty and abnormal aggregate phenotype.

First a mitosis marker (histone H3) was used to investigate whether these aggregates or DONSON protein in general was localising to the nucleus during mitosis. We chose to stain with a mitosis marker as we had previously seen a few p.R211C transfected cells showing very bright aggregate GFP expression just after they appeared to have undergone mitosis. As DONSON has only been described to act during S-phase (Reynolds et al., 2017), we thought this would be interesting to follow up. If these DONSON aggregates localised at mitosis, this would indicate DONSON was active earlier in replication than previously thought. After staining we observed 100% of these aggregates did not localise with mitosis at all (Figure 4.22).

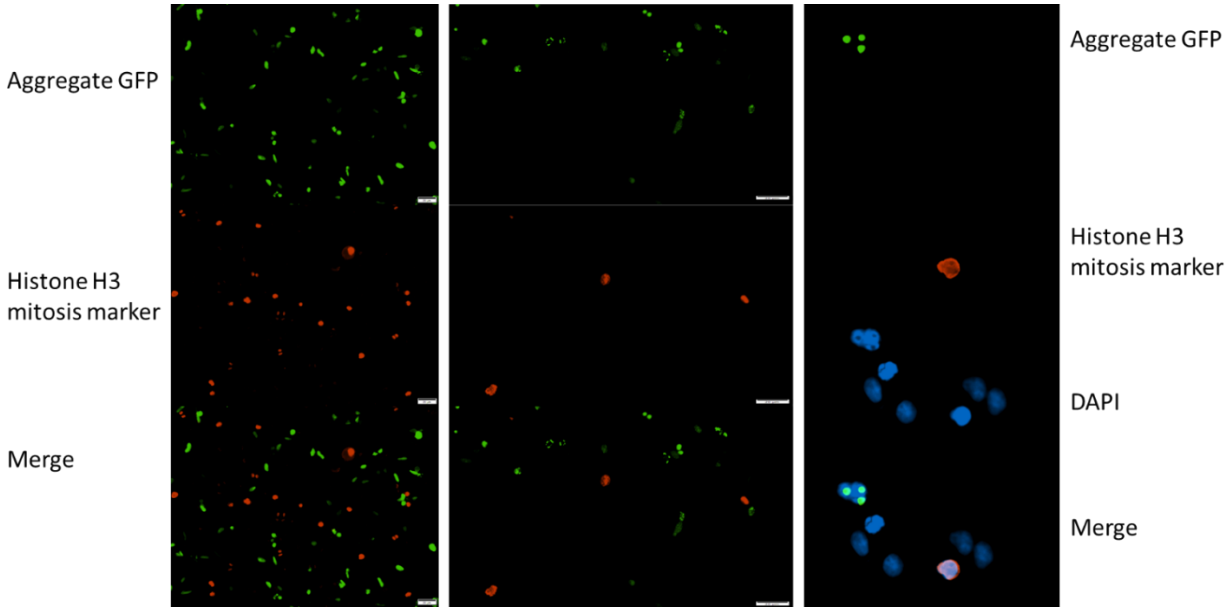


Figure 4.22 DONSON aggregates localisation in mitosis. DONSON protein aggregates did not appear to localise with the histone H3 mitosis marker in p.R211C cells. HeLa cells plated on glass coverslips were fixed 24 hours post transfection, stained with histone H3 mitosis marker and DAPI before being visualised using the Olympus BX53 microscope.

Empty, WT-DONSON and p.R211C transfected HeLa cells were all stained with mitosis marker (histone H3). There was no obvious localisation of WT or p.R211C-DONSON with

the mitosis marker at all, indicating DONSON is not active during mitosis. This fits with published data on the role of DONSON during replication fork stability during S-phase (Reynolds et al., 2017).

4.3.4 DONSON DNA damage

Next transfected HeLa cells were stained with Y-H2AX (DNA damage marker) (Ichijima et al., 2005; Sone et al., 2014). As DONSON is required for stabilising replication forks during S-phase, MISSLA-DONSON variants lead to increased DNA damage, chromosome instability and cell death. As our MGS-DONSON patients presented with a different phenotype to these MISSLA patients we investigated the effects of the DONSON-p.R211C variant on DNA damage (Figure 4.23).

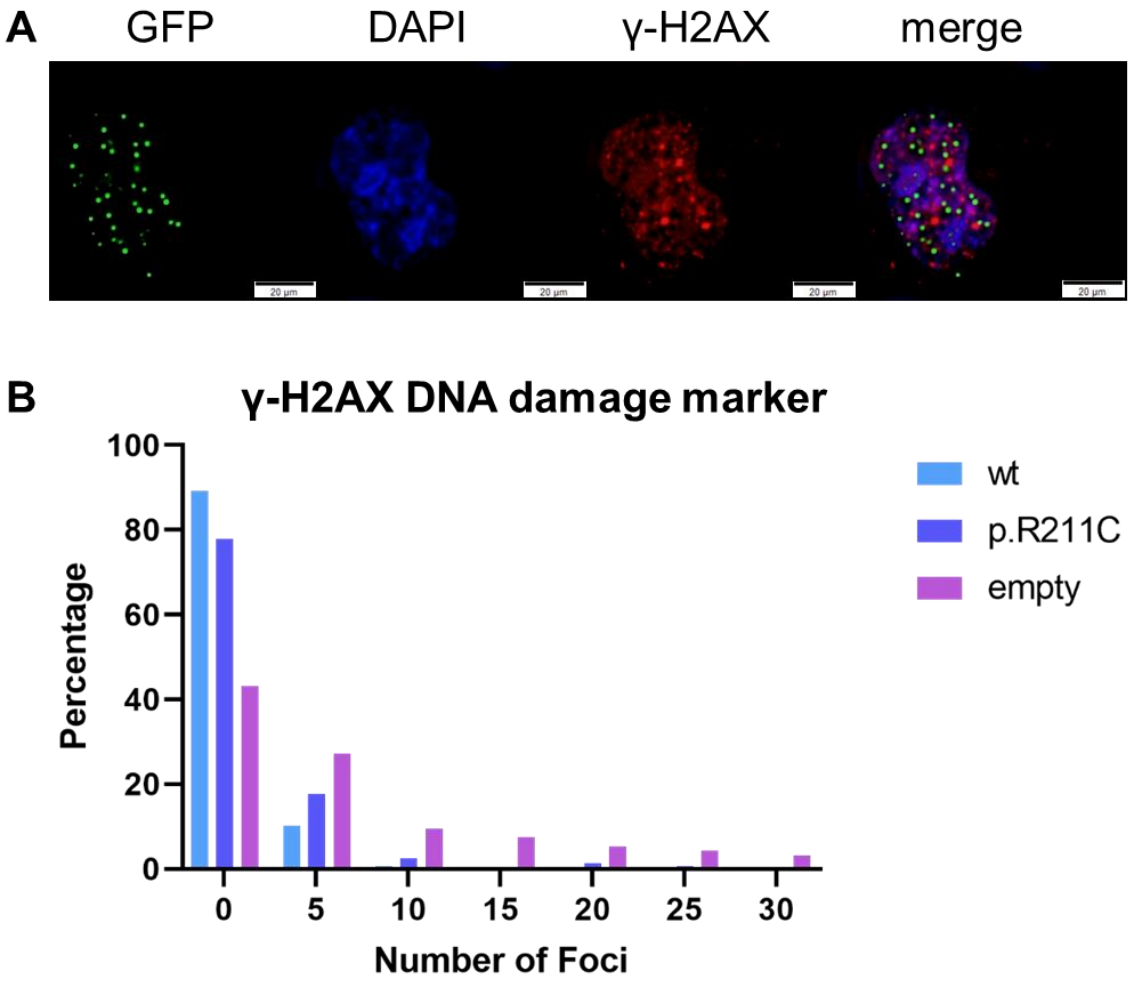


Figure 4.23 DNA damage marker is reduced in cells transfected with WT or p.R211C-DONSON. HeLa cells transfected with control or DONSON-p.R211C plasmids were fixed 24 hours after transfection, and stained with γ-H2AX DNA damage marker. A) Panel shows representative staining of γ-H2AX antibody. B) DNA damage foci were counted by eye and ranked into the above categories (N=160, one replicate).

Due to time constrains only one replicate (N=160 cells) were counted. Therefore, any results are only indicative of what may be happening, but more replicates and larger cell counts are required for a more accurate representation.

The most DNA damage was present in HeLa cells transfected with empty GFP (Figure 4.23). This decreased dramatically in cells transfected with WT-DONSON with the majority (89.2%) of cells containing only 0-4 DNA damage foci, 10.2% containing 5-9 foci and only 0.6% of cells with more than 10 foci. Cells transfected with the p.R211C plasmid also showed a decrease in DNA damage with 78% cells containing <4 foci, 17.6% with 5-9 foci and 4.4% containing over 10 DNA damage foci, but this was not as extreme as seen in WT.

This may suggest that transient WT and p.R211C DONSON may be working to protect against DNA damage in HeLa cells.

4.3.5 Discussion

Mislocalisation of GFP-DONSON protein confirmed these DONSON variants were likely pathogenic. The difference in cellular expression between other MISSLA and MGS variants and the p.R211C aggregate phenotype was unexpected. There are no obvious phenotypic differences between LB008 (homozygous for p.R211C) and other patients with other DONSON-MGS variants. This aggregate cellular phenotype had not previously been reported. As only a small proportion of WT-DONSON cells (and other variants) showed an aggregate phenotype, it is likely this phenotype was ignored as it may have been thought to be a variation on the normal speckled WT-DONSON phenotype.

50.3% of p.R211C cells showed an aggregate phenotype while MISSLA and MGS-DONSON variants tested, showed at least 77.8% cytoplasmic expression. As there is no noticeable difference in phenotype between LB008 and the other DONSON-MGS patients this may suggest the aggregate DONSON protein is completely non-functional. However this is counteracted with a higher proportion (23.3%) of WT expression compared to only (7.5-9.1%) with the other DONSON-MGS variants. Both cytoplasmic and aggregate DONSON expression suggests protein mis-folding or inability for the protein to get into the nucleus. As MISSLA is caused by DONSON dysfunctioning at replication forks leading to double-stranded breaks, chromosome instability and cell death we would expect MGS-DONSON

variants leading to the inability of DONSON localising to the nucleus to cause similar problems. Limited DNA damage data indicates that p.R211C is almost as good as WT DONSON at reducing DNA damage in transfected HeLa cells. This and phenotypic data suggests that these DONSON variants are not leading to increased DNA damage or at least not as severe DNA damage as is seen with MISSLA variants (Reynolds et al., 2017). Although this work needs to be expanded to better understand the effect of MGS-DONSON variants on double-stranded breaks, this is briefly followed up in the next section.

GFP-DONSON does not localise to the nucleus during mitosis. Although this is not entirely unexpected, we found that DONSON, specifically aggregate DONSON was not present during mitosis. We chose to followed this up based on an observation early in the DONSON localisation experiment where one cells showing bright GFP aggregates looked to have just undergone mitosis. While we did not find DONSON to localise with mitosis, if we had it would have been a strong indication that DONSON has a much earlier role in DNA replication than previously thought.

Previous research has shown that GFP can self-aggregate based on pH and concentration transfected (Krasowska et al., 2010). In this experiment we controlled for aggregates caused by GFP with control plasmids, empty and WT-DONSON. The empty-GFP plasmid showed no cells with GFP aggregates, while our WT plasmid showed some cells ranked as aggregate (8.9%), however the majority of these were likely WT spotty cells which showed a slightly more aggregate phenotype and were therefore classified as aggregate based on our ranking criteria decided upon before counting all cells.

Aggregates causes by GFP are also more likely to be found in dividing cells (Kopito, 2000; Stiess et al., 2010), therefore the mitotic staining confirmed these GFP aggregates are a result of the DONSON protein, not GFP.

As all other MGS genes work early in DNA replication (preRC or preIC) we hypothesised that DONSON may play another currently undiscovered role earlier in DNA replication. Following on from this and based on previous research (Reynolds et al., 2017) we expect the inability of DONSON to localise to the nucleus would lead to replication fork instability, chromosomal instability and cell death. This is something we follow up more in the next section.

4.4 DONSON fibre combing

As MGS is a replication disorder, we wanted to utilise fibre combing as an assay to investigate the proportion of replication events occurring in MGS-DONSON patients. This dataset could also give us an insight into different consequences of MGS vs MISSLA *DONSON* variants. Reynolds et al., 2017 found MISSLA patients had less progressing replication forks and more stalled forks than control cells.

As the majority of our previous work had been on LB008 (p.R211C) we wanted to follow up what effects (if any) this p.R211C DONSON variant had on replication events. Unfortunately, we could not obtain cells from the p.R211C individual. We did however have access to cells from LB128. This patient has two missense variants (c.670C>T, p.Pro224Ser and c.809A>G, p.Tyr270Cys) similar to LB008, and has the very typical triad of MGS phenotypes; global reduction in size, microtia, and bilateral absent patella. LB128 was born at 32 weeks gestation, measuring 37 cm length (-1.97 SD), 28 cm occipital frontal circumference (OFC) (-0.94 SD), 1.08 kg (-1.85 SD). At his most recent examination (8 years 0 months) he was 98.4 cm tall (-5.26 SD), OFC 47 cm (-4.45 SD) and 12.3 kg (-7.27 SD). Facial dysmorphism included submucosal cleft and bifid uvula. Developmentally he presented with motor and speech delay, hypotonia and normal cognition. Skeletal abnormalities included bilateral congenital knee dislocations, hyperextension and dislocation of both knees at birth. Other abnormalities included conductive hearing loss (external auditory canal stenosis), spontaneously resolved nystagmus, mild left optic nerve hypoplasia, bilateral inguinal hernia and bilateral single incomplete palmar creases.

While both the height and head circumference are more extreme than seen in LB008 height 115 cm (-3.48 SD), OFC 48 cm (-3.97 SD), both patients presented with a relatively similar phenotype (Figure 4.24); global reduction in size, microtia and absent patella. LB008 also presented with delayed development, high arched palate, micrognathia, mild hearing loss and hypopigmentation.

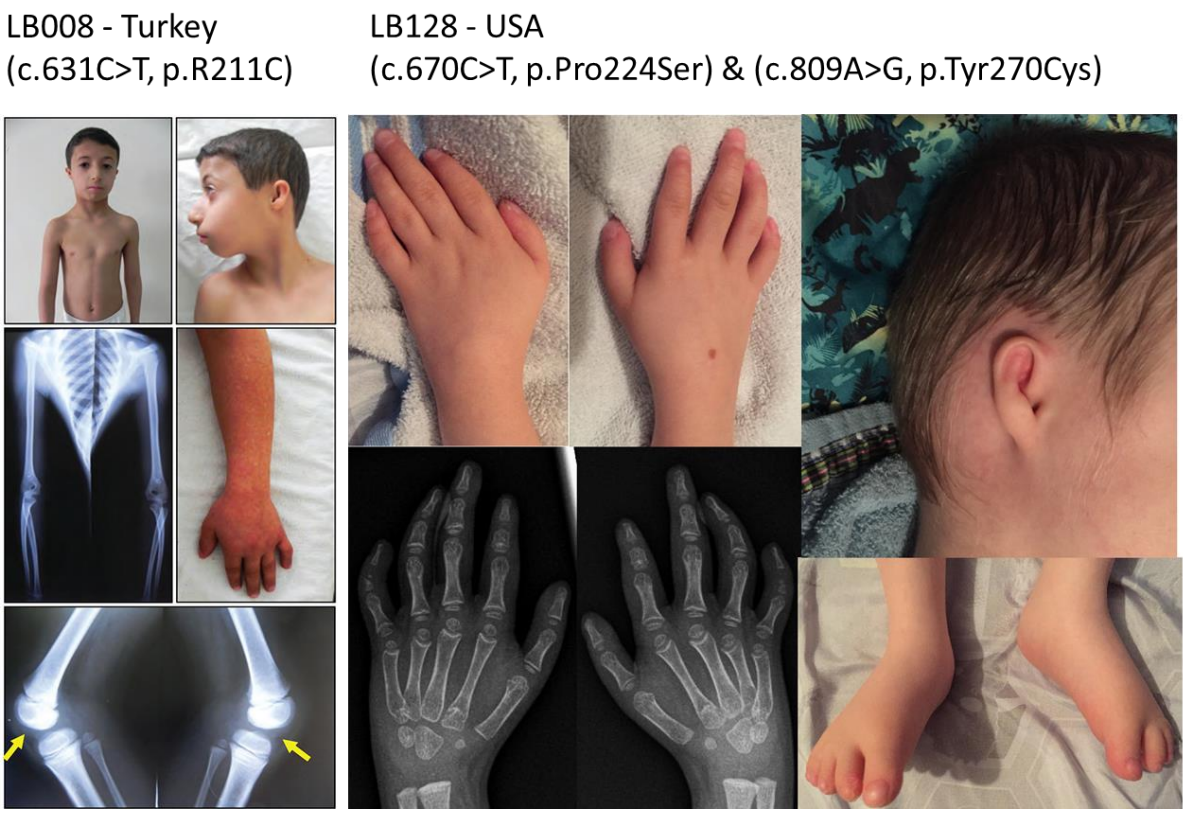


Figure 4.24 LB008 homozygous *DONSON* variant (p.R211C) and LB128 compound heterozygous *DONSON* variant (p.P224S & p.T270C) presents with very typical MGS symptoms. Note the small misshaped ears, absent kneecaps, tapering fingers and shortened toes. Figure adapted from (Karaca et al., 2019 & Knapp et al., 2019) copyright obtained from John Wiley and Sons – 4850530640988 and under the Creative Commons Attribution 4.0 International License.

4.4.1 Quantifying Replication events by DNA fibre combing

Fibre combing or molecular combing can be used as a quantitative measure of origins of replication in the genome (Herrick and Bensimon, 1999), as well as a number of other replication events such as progressing replication forks (Schwab and Niedzwiedz, 2011). Fibre combing uses surface tension (water-air) to uniformly stretch the DNA along a glass slide (Kaykov et al., 2016; Michalet et al., 1997), before florescent probes are added, to stain actively replicating DNA.

Patient and control cells were incubated with two nucleotide analogues CldU and IdU for 20 minutes each. The cells were then lysed and the DNA was uniformly spread along a glass slide using surface tension to stretch out the DNA. Staining was performed with rat and mouse BrdU antibodies which, despite their name, specifically recognise CldU and IdU respectively (Quinet et al., 2017). Fibre combing can be used to quantify multiple replication events such as progressing forks, stalled forks, new origins of replication,

bidirectional forks and termination/collision of forks based on the red/green pattern (Figure 4.25). The speed of progressing forks and bidirectional forks can also be measured. To do this the length of tracts were measured using specialised software, ImageJ (Quinet et al., 2017). Track lengths were then converted from μm to kb using the common conversion 2.59 kb/1 μm (Daigaku et al., 2010; Jackson and Pombo, 1998; Quinet et al., 2017). Length in kb was then divided by 20 minutes to give speed in kb/minute.

Further information on fibre combing protocol is available in section 2.4.2.

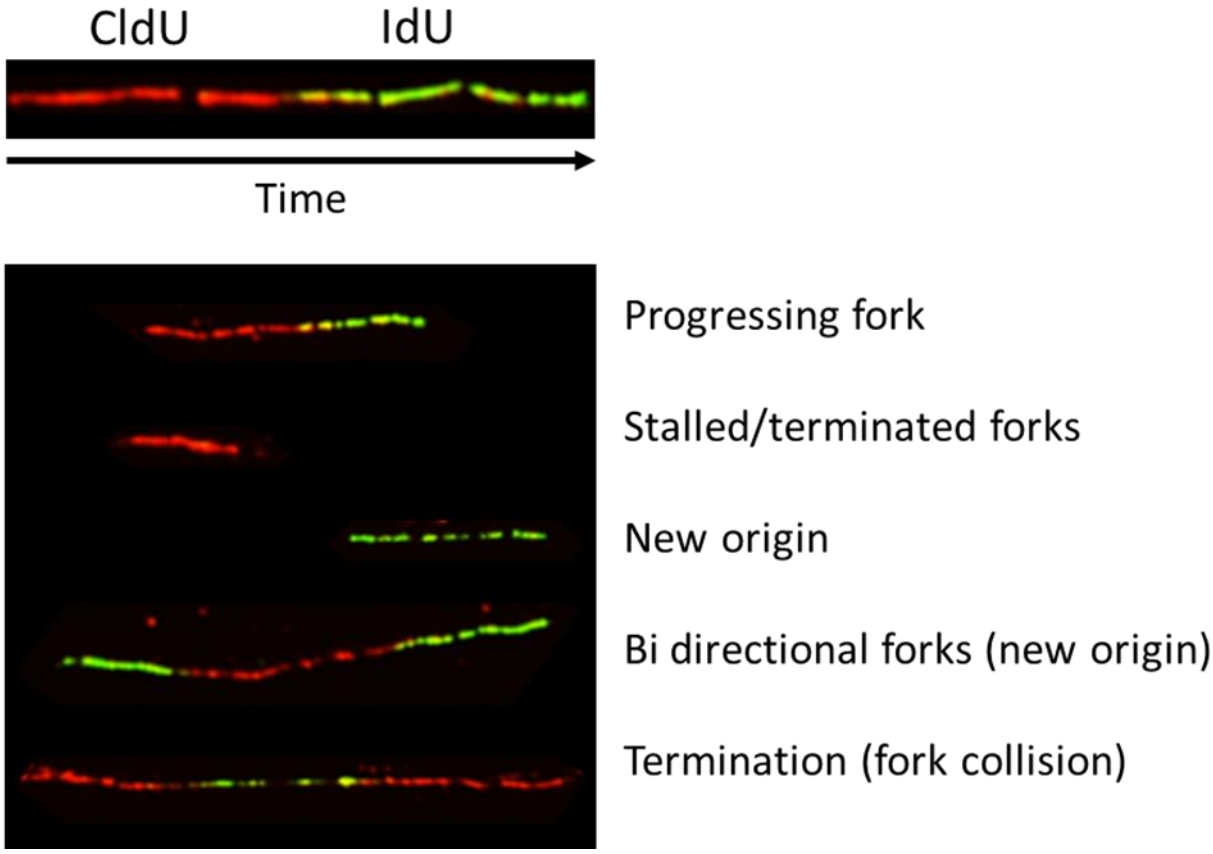


Figure 4.25 Five replication events can be quantified using fibre combing. Cells were pulsed for 20 minutes each with CldU and IdU before being stained with BrdU antibodies, images taken on Olympus BX53 upright microscope and tracts analysed using ImageJ software. Five distinct replication events can be quantified by fibre combing, including replication fork progression, termination, bi directional initiation, new origins and stalled forks.

As presented in Table 4.3 and Figure 4.26, although not statistically significant there were some minor differences in replication events between MGS vs WT cell lines noted. The percentage of stalled replication forks was slightly higher for LB128 (22.5%) than the control (11.2%), this is similar but less severe than what was seen in MISSLA patients

(Reynolds et al., 2017). There were also slightly fewer new origins of replication initiated in patient cells (19.1% WT, 12.1% LB128), particularly bi-directional replication forks (10.8% WT, 4.8% LB128). The number of progressing and terminating/colliding replication forks was the same between both groups (Figure 4.24). In MISSLA patients there were less progressing replication forks compare to WT, which is another point of difference. Based on this experiment it seems the MGS-DONSON phenotype has a less severe effect on the DNA replication than is seen in MISSLA cell lines.

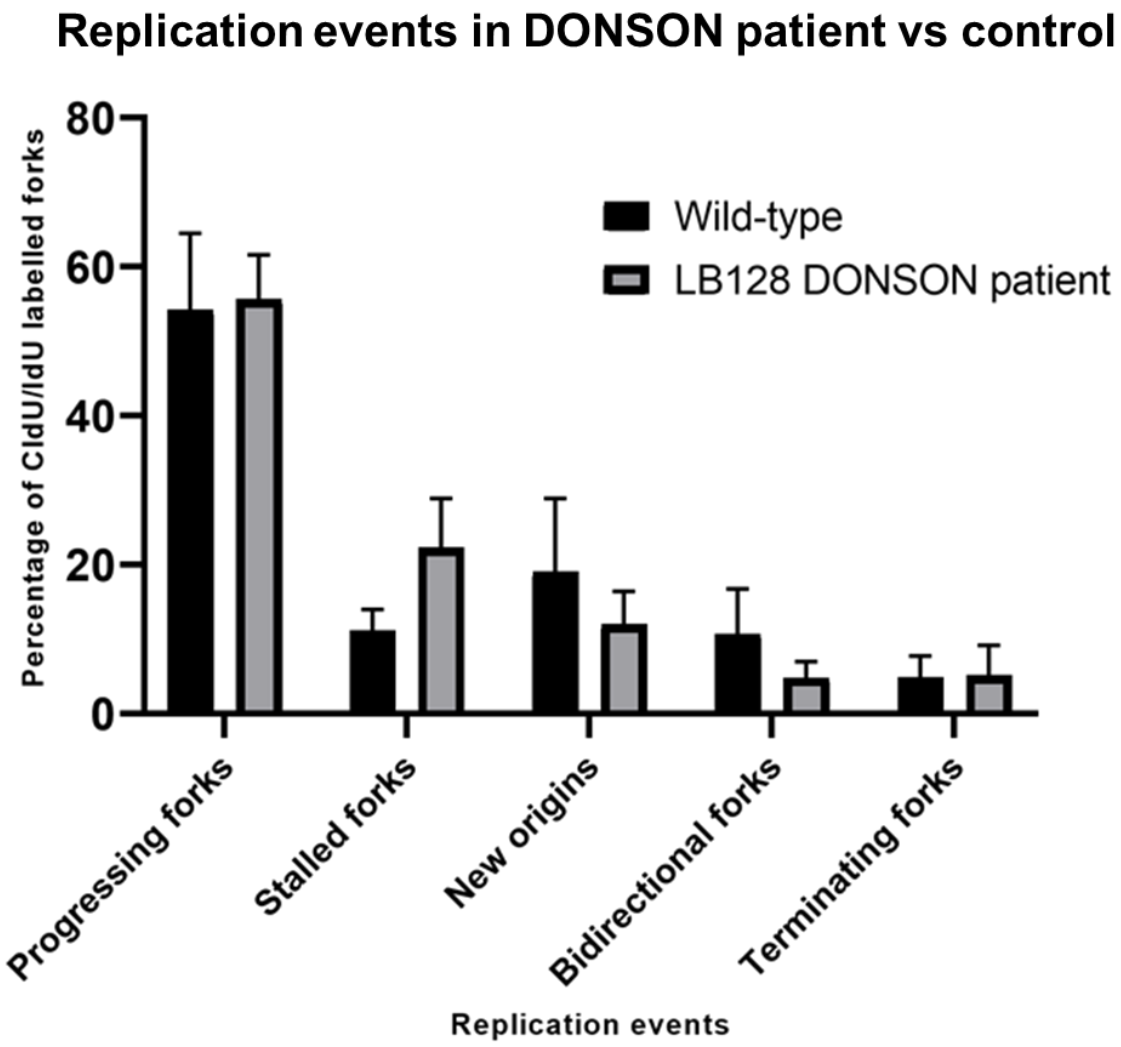


Figure 4.26 Percentage of fibre combing events in control and patient cells. Percentage of replication events in WT control cells and DONSON patient fibroblasts were quantified (N=601 total across three replicates). Number of new, stalled, bidirectional, terminating and progressing forks were counted and measured using image J. (Percentage of replication events plotted, error bar = SD, Multiple-T tests were performed and results are not significant)

Table 4.3 Percentage of replication events in control or patient fibroblasts.

	WT	LB128
Progressing forks	54.1%	55.6%
Stalled forks	11.2%	22.5%
New forks	19.1%	12.2%
Bi directional forks	10.8%	4.8%
Collision/termination of forks	4.8%	5.0%

4.4.2 Replication velocity

Tracts were measured and converted to kb/minute as described in section 4.4.1. Approximately 200 tracts were counted for each replicate as recommended by Quinet et al., 2017; Técher et al., 2013. A total of 601 tracts were counted across all three replicates for each cell line.

Replication fork speed localised around 1 kb per minute in WT cells, while there tended to be much more variation in the replication fork speed in LB128 cells (Figure 4.27).

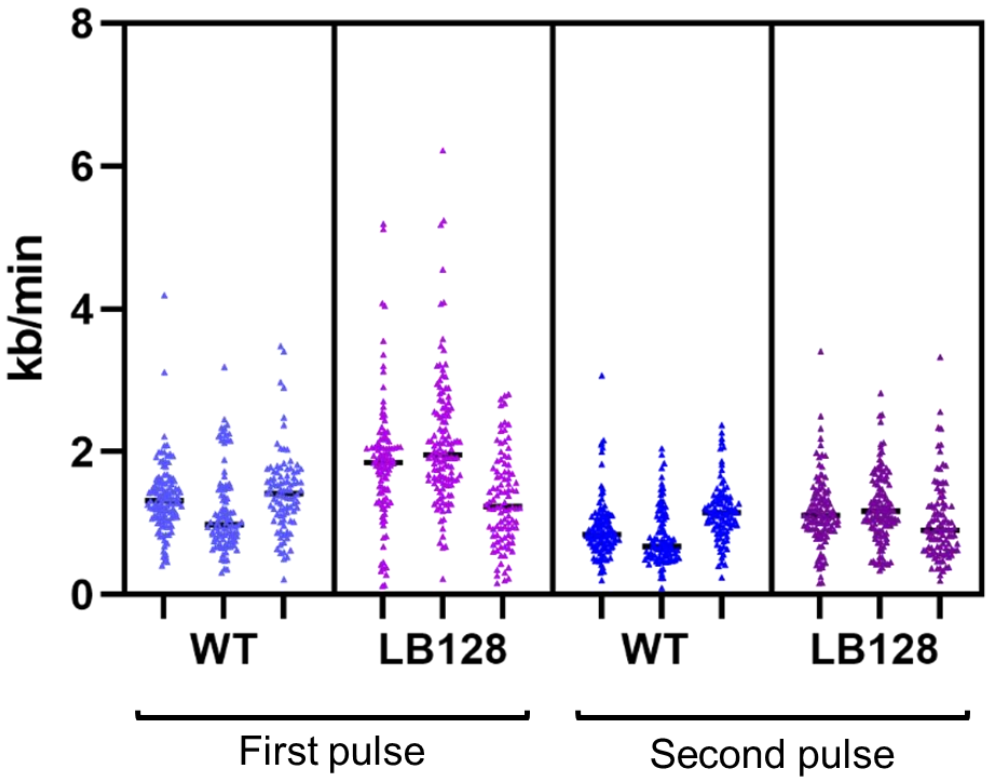


Figure 4.27 MGS-DONSON variants lead to changes in replication fork velocity. Red and green tracts from progressing replication forks were measured using ImageJ software, converted to kb/minute and plotted using Prism GraphPad. Mean of triplicates: WT first pulse = 1.32 kb/min, LB128 first pulse = 1.76 kb/min, WT second pulse = 0.97 kb/min, LB128 second pulse = 1.40 kb/min.

The range of mean replication fork speed in WT was 0.81-1.44 kb per minute while in LB128 it was 1.01-2.13 kb per minute. Overall, this suggests that LB128 cells seem to be replicating their DNA faster than WT cells. While this is not consistent with the reduced growth phenotype, it is similar to what was seen when Reynolds et al., 2017 knocked down DONSON expression using siRNA. Due to the large range of replication fork velocity present in LB128 cells we considered if this overall increase in replication forks was due to replication fork asymmetry. This is where one side of the replication fork replicates faster than the other side, this can lead to replication fork collapse and DNA damage as a result of double-stranded breaks. Replication fork asymmetry could also help explain the large amount of variation seen in replication fork speeds (Figure 4.27).

We briefly followed up replication fork asymmetry by reanalysing fibre combing images to increase the number of bidirectional forks we had to analyse. Due to time constraints we were unable to count the recommended 100-200 tracts so this data is just a brief estimate at what may be going on. This needs to be repeated to increase the number of tracts counted from each replicate.

As bidirectional tracts are not as common as progressing tracts (Figure 4.26) there were fewer tracts available to analyse. Ideally this experiment would be repeated to improve the number of tracts counted across the replicates. Although the number of bidirectional tracts counted was low there didn't appear to be any obvious differences in replication fork symmetry (Figure 4.28). If such a conclusion holds, this could explain why we do not observe cellular or clinical phenotypes related to chromosomal instability in our MGS patients. Conversely, this still does not explain why variants in *DONSON* can cause MGS or MISSLA.

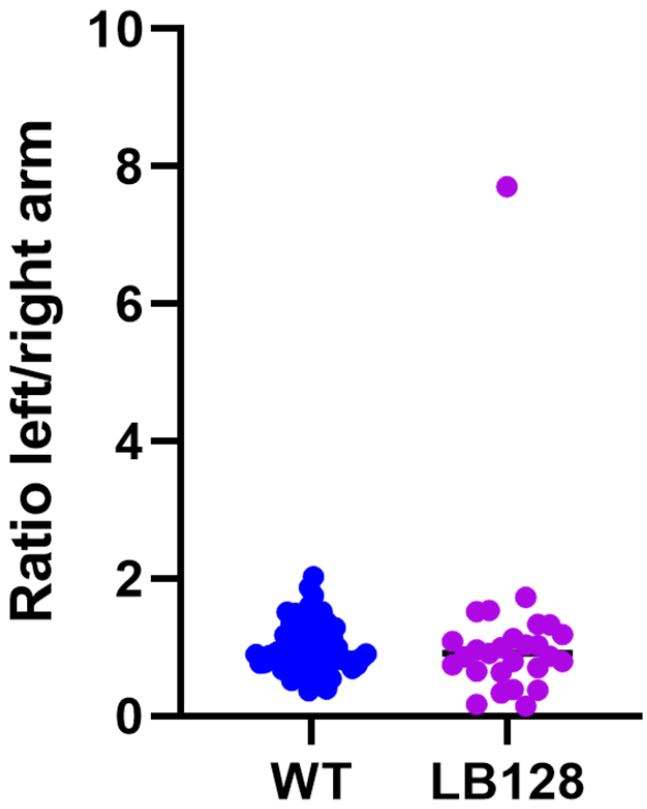


Figure 4.28 Replication fork symmetry. Replication fork symmetry appears similar between WT and LB128. Only one replicate performed due to time constraints.

Potentially an increase in DNA replication speed could make up for delays elsewhere in replication explaining why a normal growth curve is seen in MGS patients after 1 year of age. Conversely this difference in replication speed could be an artefact and normal growth in DONSON-MGS patients may be due to a less severe effect on the DONSON protein which is tolerated better by the cells.

In other MGS patients with variants in genes involved in early DNA initiation, we would not expect to see differences in replication speed between patient and WT. But as DONSON's only known role is during replication it is interesting that we see increased replication speed but not due to replication fork asymmetry.

Other MGS variants are thought to cause delays at the initiation of DNA replication not during replication so it would be interesting to repeat this experiment with other MGS cell lines to investigate differences in replication speed. Perhaps MGS growth is normal as G1 phase may be delayed but is made up for in S-phase. To investigate this further this fibre combing experiment should be repeated with all DONSON-MGS patients as well as control cell lines from MISSLA, MIMIS and other MGS patients.

4.4.3 Discussion

In this section we investigated differences in DNA replication initiation, progression and termination events in a patient-derived cell line harbouring biallelic *DONSON* variants. We saw a slight decrease in new replication origns and a slight increase in stalled forks but this was not significant or as extreme as is seen in MISSLA patients (Reynolds et al., 2017). More surprisingly we saw an increase in replication fork progression which does not correlate with the reduction in size seen in MGS patients. This was followed up by analysing replication fork symmetry, both to compare to the previous literature (Reynolds et al., 2017), and to investigate whether this increase in replication fork speed may be due to bidirectional replication forks progressing in a varying speeds leading to asymetrial replication forks – which are likely to stall. This also was not observed. As neither the change in replication events or increase in speed is particularly large perhaps this slight increase in replication speed compensates for the slight reduction in new forks and increase in stalled forks.

Differences in the proportion and type of replication events in control and LB128 fibroblasts suggests this MGS-DONSON variant may cause different effects at the molecular level than MISSLA-DONSON variants. As MISSLA and MGS patients present with different phenotypes it is expected we would see differences at the molecular level, however how DONSON functions at a protein level to cause these differences is unknown.

Brain tissue is primarily affected in MISSLA and MIMIS patients (Evrony et al., 2017; Reynolds et al., 2017). The inability of DONSON to stabilise replication forks in tissues which undergo a large amount of cell division during development (i.e brain tissue) is obvious in these patients presenting with severe microcephaly. As our MGS-DONSON patients do not present with severe microcephaly we hypothesised that DONSON may play another role earlier in DNA replication which was contributing to the MGS phenotype.

Based on evidence in this section and previous research (Abdelrahman et al., 2019; Evrony et al., 2017; Karaca et al., 2019; Reynolds et al., 2017; Schulz et al., 2018), it seems these DONSON-MGS variants may be causing a less extreme MISSLA phenotype. MGS-DONSON variants cluster at the N-terminus of the protein, and while MGS-DONSON variants causing a less extreme phenotype explains the microcephaly differences it does not explain why MGS patients present with a global reduction in size, (or a more severe

growth phenotype) than what is seen in MISSLA patients. How these differences in phenotype are caused is still unknown as we do not know the exact function of DONSON during replication.

One hypothesis explaining the growth phenotype in MGS patients is that defective DNA replication proteins function relatively normally under standard conditions. However at a critical point in development when extra pressure is put on DNA replication due to maximal proliferation needed to keep up with standard development, the overall number of cells is reduced due to the inability to replicate ones DNA fast enough (Klingseisen and Jackson, 2011). This could help explain the subtle differences seen between control and LB128 especially in relation to the proportion of replication events

Based on the research in this chapter, it is unknown how variants in DONSON are causing issues with nuclear localisation as DONSON does not appear to cause issues associated with replication fork stability in these patients, and yet does seem to cause issues with DNA replication initiation (a reduction in new origins). Previous research showed MISSLA was caused by the inability of DONSON to stabilise replication forks (Reynolds et al., 2017), due to mis-localisation or reduced protein levels. In this chapter we showed that all five MGS-DONSON variants showed similar levels of mis-localisation either due to cytoplasmic or aggregate expression, however there is no evidence that this mis-localisation of DONSON is leading to increase DNA damage or chromosomal instability associated with MISSLA. How exactly DONSON is still functioning at a level to stabilise these replication forks during S-phase is unknown. This may suggest that DONSON plays another currently undefined role earlier in DNA replication, and the small proportions of DONSON which is localising to the nucleus is successfully stabilising replication forks therefore not leading to the MISSLA phenotype. However as MGS is caused by defects in proteins essential for the initiation of DNA replication leading to an overall reduction in size it is thought that DONSON may play another role here which is not being tolerated either by the reduced protein levels due to mis-localisation or the position of missense variants in a domain important for this function.

As DONSON plays such an important role in replication stability it would be beneficial to follow up double-stranded DNA damage and chromosomal instability in MGS patients. Although not directly investigated during this project, based on phenotypic data and limited replication fork dada we do not seem to see any markers of chromosomal instability. How DONSON is functioning normally (enough) to not lead to ds breaks and

yet still leading to the MGS phenotype is unknown and something that future studies should follow up.

5 Discussion and future directions

MGS is caused by the inability to replicate one's DNA in an appropriate time frame. In this study we investigated three novel MGS variants in three DNA replication genes. These genes caused varying levels of severity in terms of the MGS phenotype. These three variants are in genes known to act at three different stages of DNA replication: pre-replication complex (late M-phase/ early G1 phase), pre-initiation complex (G1/S interphase) and replication fork progression (S-phase).

5.1 Novel AAA+ domain ORC1 variants

ORC1 functions the earliest in DNA replication. It is active during late M/early G1 phase, as it is an important component of the ORC complex, the first complex to bind to origin of replication sites in the genome to initiate DNA replication. ORC1 has previously been reported to be involved in MGS. In previously published literature it was thought the MGS phenotype was caused by variants in the BAH domain (Kuo et al., 2012). The BAH domain is important recognising H4K20me2 present on the chromatin at origin of recognition start sites and loading the ORC complex onto the chromatin (Kuo et al., 2012; Zhang et al., 2015). All previously reported ORC1-MGS patients had at least one variant in the BAH domain, therefore it was thought this may be the crucial domain, which leads to the often more severe height phenotype seen in ORC1 patients (Kuo et al., 2012; de Munnik et al., 2012b).

Although we did not manage to explore this gene in as much depth as we would have liked to during this study, we can still be reasonably confident this p.L622P variant is causing the MGS phenotype seen in this patient. LB005 c.1965T>C variant causes a p.L622P amino acid change in a highly conserved area of the protein. The AAA+ domain is important for ATP hydrolysis, as well as a range of other cellular activities essential for DNA replication including remodelling and loading subunits involved in DNA replication (Herbig et al., 1999; Neuwald et al., 1999). LB005 is homozygous for a missense variant in the AAA+ domain. Previous literature has described other MGS patients to have heterozygous and homozygous missense and truncating variants in the AAA+ domain of ORC4, CDC6 and MCM5 (Bicknell et al., 2011b; Bongers et al., 2001; Guernsey et al., 2011; Kerzendorfer et al., 2013; Vetro et al., 2017). Due to this and the important role the AAA domain plays in

DNA replication (Herbig et al., 1999; Neuwald et al., 1999), it is likely this homozygous c.1965T>C variant is causing the MGS phenotype seen in this patient.

This variant is also not present in gnomAD giving some evidence that it is not generally tolerated, however as this patient is of Saudi Arabian decent this population is under represented gnomAD and therefore may actually be present at a slightly higher proportion in this population. As the *ORC1* splicing assay, was unable to be completed during this project, it should still be carried out to ensure our assumptions are correct and this c.1965T>C variant does not cause a splicing effect.

Issues in the initiation of DNA replication by any member of the ORC complex will likely lead to fewer licenced origin of replication sites, leading to a delay in G1 phase progression or a reduction in back up licenced origins (Blow et al., 2011). In healthy cells there are always more origin of replication sites licenced than needed (dormant origins), which act as a reservoir ready for activation during replication to ensure timely completion of S-phase (Ge et al., 2007). As only a small proportion of these origins are actively replicating at one time (McIntosh and Blow, 2012), there are plenty of back up sites which can be used if replication forks stall or terminate early due to replication stress (Woodward et al., 2006). During replication stress, dormant origins will fire leading to DNA replication being completed in a timely fashion (Blow and Ge, 2009; Blow et al., 2011; Ge et al., 2007; Woodward et al., 2006, 2006). ORC1-MGS variant likely lead to less efficient binding by ORC1 at the chromatin, leading to less dormant origins. During normal replication a large proportion of these dormant origins will not be required (Ibarra et al., 2008; McIntosh and Blow, 2012), however these dormant origins will be required in times of replication stress or increased proliferation during development. In these MGS patients we predict there are not enough of these origins present and so, when replication forks stall, they must wait to be restarted or there may end up being fewer active progressing replication forks altogether, leading to increased duration of S-phase DNA replication duration.

ORC1 which does not function as it should could also lead to delays during G1 phase as it may take longer for the ORC complex to bind and recruit other proteins required for the preRC. Future research should investigate whether AAA+ domain ORC1 variants lead to fewer origins of replication during S-phase or delayed G1 phase progression.

Bicknell et al., 2011a and Kerzendorfer et al., 2013 observed delayed S-phase entry and slower S-phase progression in patient cells harbouring ORC1 variants in the BAH domain

compared to control. S-phase progression was investigated using pulse BrdU labelling and pan-nuclear γ-H2AX immunostaining. To investigate delayed S-phase entry due to prolonged G1 phase, cells were serum starved to synchronise cells in G0, G1 phase was initiated by the addition of serum and BrdU was added to monitor S-phase entry by FACS. Experiments similar to this could be used to investigate the effect of the AAA+ domain variants on ORC1's ability to function in cell cycle progression. Other experiments such as fibre combing could also be used to investigate whether these delays in S-phase are due to a lack of new origins or slower replication fork speed during S-phase

Future study of these non-BAH domain ORC1-MGS variants should include investigating the chromatin loading ability of these, proportion of replication events during S-phase (as above) and the cells ability to progress through the cell cycle (as above).

Previous research by Bicknell et al., 2011a showed a reduction in protein levels of both unbound and chromatin bound ORC1, as well as a reduction in other preRC proteins ORC2 and MCM2. This was investigated by immunoblot of soluble and insoluble protein as well as chromatin-enriched fractions generated by micrococcal nuclease digestion (Bicknell et al., 2011a).

The licencing ability of mutant ORC1 was also investigated by monitoring the replication of Epstein-Barr virus episomes (Bicknell et al., 2011a; Kerzendorfer et al., 2013). Epstein-Barr virus is a 165 kb double-stranded DNA virus which requires a functional ORC complex to replicate (Dhar et al., 2001), therefore variants in *ORC1* affecting the assembly of the ORC complex and its ability to load chromatin will result in less amplification than WT.

As many of these techniques require cell lines containing the patient variants, it would therefore be extremely beneficial to have CRISPR editing working as a technique to achieve this in our laboratory, as patient cell lines were not available for these *ORC1* variants (as discussed in Section 3.7).

5.2 Novel *CDC45* splicing variants lead to a severe MGS phenotype

The next of our three novel variants to work in DNA replication is CDC45, this protein is important for preIC firing and displacing the leading strand during replication fork progression (Karnani and Dutta, 2011; Köhler et al., 2016; Pacek and Walter, 2004; Wong

et al., 2011). *CDC45* is therefore different to other published MGS genes which tend to work earlier in DNA replication (preRC). Previously described CDC45 patients also generally present with a slightly different phenotype to other MGS patients (Fenwick et al., 2016). It was thought this difference in phenotype may be due to CDC45 working later in replication than the other MGS genes.

RECQL4 is a replication gene which is involved in a range of disorders and associated phenotypes. Variants in *RECQL4* cause Rothmund-Thomson syndrome, Ballier–Gerold syndrome and RAPADILINO syndrome. These syndromes cause a wide range of disorders, with Rothmund-Thomson syndrome patients presenting with short stature (delayed growth) and increased risk of cancer (overgrowth). While precise geneotype-phenotype correlations are not fully understood (Kellermayer, 2006), this range of phenotypes is thought to be due to RECQL4 interacting with many different proteins and pathways during DNA replication and cell cycle progression (Capp et al., 2010; Khakhar et al., 2003; Kitao et al., 1998; Matsuno et al., 2006; Sangrithi et al., 2005; Siitonen et al., 2009).

As the *CDC45* variant (c. 1441-2 A>G) studied here lead to a more severe phenotype, it is hypothesised that the severe knockdown alongside the lack of any CDC45 canonical transcript being present in the cell has led to CDC45 functional issues at multiple point in replication, therefore leading to the more severe MGS and secondary phenotypes.

The CDC45 patients studied, presented with a more severe phenotype than typically seen. Due to this we hypothesised these patients may have a more severe knock down of CDC45 than previously described. While our collaborators hypothesised these siblings may have an almost complete loss of CDC45, it is unlikely an embryo with a complete knockout of CDC45 would be viable, as CDC45 plays such an integral role in DNA replication. The 72.9% reduction in CDC45 expression is less severe than what has been seen in previous CDC45 patients presenting with a less severe phenotype (Cooper, 2016; Fenwick et al., 2016). Due to this we expected to see differences in protein stability between the WT-CDC45-FLAG plasmid and the 6bp_del-CDC45-FLAG plasmid. Unfortunately it was hard to tell from our western blot whether or not there was a reduction in protein stability.

Unlike in our CDC45 patients previous *CDC45* splicing variants resulted in some full length canonical transcript remaining (Fenwick et al., 2016). In LB192 the remaining 27.1% CDC45 protein expressed contained a two amino acid deletion, although this two amino acid deletion did not appear to cause problems with protein stability, it could be leading

to other issues in DNA replication which we were not able to cover during this project. Issues could include CDC45's ability to displace the leading strand during replication or interact with other components of the replication fork during S-phase. We can also not rule out the possibility these patients contain another variant(s) in a different gene(s) that could be contributing to the phenotype.

Future study of this *CDC45* splicing variant is required to better understand the full effect of this novel variant. Future work could include investigating the effect of this two amino acid deletion on the ability of CDC45 to interact with other members of the preIC and replication complex, as well how this splicing variant affects the cells ability to progress through the cell cycle.

The difference in phenotype between the *CDC45* splicing variant in this study and in the *CDC45* splicing variant in Fenwick et al., 2016 is striking. Between two splicing variants in *CDC45* we saw phenotypes ranging from mild MGS to craniosynostosis to the very severe phenotype seen in LB191 and LB192. These major differences in phenotype could be better investigated in animal models of development, such as *Danio rerio* or *Mus musculus*. It is possible other proteins during development are somehow compensating for the reduction in CDC45 expression in certain tissues during development, this may not happen all of the time therefore leading to a less severe phenotype in some patients but not others.

Using an animal model would also allow us to investigate how proteins such as DONSON may interact with other members of the preIC and replication fork during development. As DONSON's role in DNA replication was only recently discovered, exactly how and which proteins DONSON interacts with is uncertain. As both *CDC45* and *DONSON* cause MGS as well as other diseases future studies should investigate their role and possible interactions during development. *D. rerio* would allow us to attempt to rescue CDC45-MGS or DONSON-MGS phenotype using other proteins suspected to be involved in the pathway. An animal model could also be used to further investigate when in development the MGS phenotype arises giving us some idea of the time point that knockdown of DNA replication is not tolerated in an organism

5.3 Novel MGS gene *DONSON*'s role in MGS

While variants in *CDC45* causes a difference in phenotype, thought to be due to it working later in replication, we do not see a difference in phenotype between our MGS-DONSON and other MGS patients. In terms of protein function this is unexpected, as DONSON's only known role is during S-phase, later in DNA replication than other MGS genes are known to act. All other MGS variants play integral roles in early DNA replication, to us this suggests that DONSON may have other undiscovered roles earlier in DNA replication. Due to no differences in phenotype between DONSON-MGS patients and other MGS patients, this suggests that all variants cause similar problems in DNA replication, leading to the same overall effect on the cells and individual as a whole.

The majority of MGS-DONSON variants clustered away from other MISSLA and MIMIS-DONSON variants. This could suggest the N-terminal of DONSON may be important for interacting with other proteins/pathways during replication.

Four of the five MGS-DONSON variants showed diffuse cytoplasmic expression. Suggesting that MGS-DONSON variants cause similar issues with DONSON localising to the nucleus as MISSLA variants do. While indicative that these variants are pathogenic, this does not explain the difference in phenotype between these different syndromes. It is expected that issues with DONSON localising to the nucleus would also impair DONSON's ability to stabilise replication forks. Based on our limited fibre combing data, we did not see this as there was no difference in replication fork symmetry. Data in this project suggest that these DONSON-MGS variants may be causing a more subtle effect on DONSON than what is seen in MISSLA patients.

As DONSON's role in replication has only recently been discovered, there is still much we do not understand about how it functions and what proteins it interacts with. Therefore DONSON's molecular role in MGS requires more extensive research to better understand the role it is playing in MGS and DNA replication.

5.4 Future directions

Next I will describe some suggestions for future studies. These include, following up replication forks asymmetry and confirming this is not the cause of increased replication fork speed seen in LB128.

As done in previous research (Reynolds et al., 2017) we could also add replication stress to these cells in between pulses to investigate how this effects the cells ability to restart DNA replication. Adding replication stress to DONSON-MGS variants would likely exacerbate any differences seen between patient and control, as DONSON would be required to stabilise these stalled replication forks for a period of time. Therefore we would be able to see if our MGS-variants are better or worse at stabilising replication forks and restating replication than MISSLA-DONSON variants are. As we don't seem to see any chromosomal instability in our MGS patients we would expect MGS-DONSON variants are better at stabilising replication forks than MISSLA-DONSON variant are.

Fibre combing could also be used to investigate the proportion of replication events in other MGS patients, *ORC1* vs *CDC45* vs *DONSON*, as this could give us an insight into how MGS variants in different DNA replication genes effect DNA replication events and speed. We would expect to see slight differences in replication events depending on what gene the variant is in. For example with our *CDC45* variant we would expect to see less new origins firing as CDC45 is a rate limiting step in replication as it is recycled between origins (Musiałek and Rybaczek, 2015). While in *ORC1* variants we may expect to see normal replication under normal conditions as these variants are thought to primarily effect G1 phase (Stiff et al., 2013). If we exposed these *ORC1* variants to replication stress by the addition of hydroxyurea (which inhibits the incorporation of nucleotides (Vesela et al., 2017)), we may see fewer newer origins as we hypothesise there may be less dormant origins due to ineffective chromatin loading by ORC in G1 phase. This could potentially lead to a better understanding of why MGS patients with variants in *ORC1* and *ORC4* present with the most severe height phenotype.

One aspect of the project we still do not understand is what the GFP-aggregates present in p.R211C transfected HeLa cells are and what they could be localising with. We confirmed that these aggregates are not localising with mitosis and they don't seem to lead to an increase in DNA damage but more experiments need to be performed to confirm this.

As mentioned previously the effect of MGS-DONSON variants on DNA damage is something that requires further investigation. Not enough cells or replicates were performed (due to time constraints) with p.R211C plasmid to definitively conclude the effect of this variant on DNA damage. The experiment should be repeated with all five MGS-DONSON variants as well as controls and MISSLA variants to investigate the

differences in DNA damage between MGS-DONSON variants and MISSLA-DONSON variants.

There is no solved structure of DONSON, therefore it is difficult to understand how variants in different areas may be affecting the protein as DONSON shows no similarities to other proteins and no clear domains are predicted (Reynolds et al., 2017). One option for exploring what areas of the protein are important could be to delete different sections of the gene and observe the effect on DONSON function. As the majority of our MGS-DONSON variants cluster towards the N-terminus of the protein we may find that small deletions or missense changes in this area lead to a MGS effect whereas deletions or missense changes at the C-terminal lead to a MISSLA or MIMIS phenotype.

This project describes differences in phenotypic and molecular consequences of DONSON variants in MGS and MISSLA patients. Many of these experiments lead us to believe DONSON may play another role earlier in DNA replication than previously thought. Due to the lack of knowledge about how DONSON directly interacts with proteins of the replisome during replication it is important to follow on from this project to better understand DONSON's role in DNA replication. In depth investigation of DONSON's interactions with components of the preIC could be performed via pulldown. Based on experimental data, we hypothesize that we may see DONSON localising with some components of the preIC although what role it may be playing in early DNA replication is unknown.

5.5 DNA replication

Overall, these three novel MGS variants show how variants at multiple stages of DNA replication can lead to disease. Although all three variants cause effects on DNA replication in different ways, they all lead to the same overall phenotype of a global reduction in size. While there is still much to be uncovered about these MGS genes and variants this project has given us a good baseline on where to move forward from here.

Replication disorders cause a range of disorders and phenotypes from delayed growth to overgrowth (Jackson et al., 2014; Klingseisen and Jackson, 2011; Mazouzi et al., 2014; O'Driscoll, 2017). There are many genes involved in many stages of DNA replication which lead to disease, due to the essential role these encoded proteins play in the cell (O'Driscoll,

2017). DNA replication is a complex process, involving many more proteins and pathways than are discussed in this thesis. Primordial dwarfisms are a group of disorders which group together as they are all cause restricted growth throughout development and are caused by genes working in similar pathways (Khetarpal et al., 2016; Müller et al., 2012).

As shown in this thesis and previous literature single genes can cause a wide range of phenotypes due to the diverse role many of these proteins play in the cell. Many proteins for example CDC45 and RECQL4 act in multiple stages of the cell cycle and therefore lead to a range of phenotypes and disorders. This is likely due to these proteins interacting with multiple other proteins and pathways. Therefore depending on where the variant is the phenotype may differ due to different pathways being effected.

Replication disorders give us a unique opportunity to study the effects of single proteins on replication and cell cycle progression as a whole. While this process is complex, the ability to study the effects of these replication disorders helps better our general understanding of the different roles and pathways these proteins are involved in.

6 Conclusions

This project investigated the effect of three novel MGS variants on DNA replication using a variety of techniques. While we were unsuccessful at using CRISPR techniques to knock in these variants in near-diploid cell lines we gained some valuable insight as to what doesn't work and how it might be going wrong. This along with new techniques being published on a regular basis may allow us to successful use CRISPR to knock in our patient variants in the future. Although in depth analysis of all three variants could not be performed within the scope of the project, this project allowed us to gain valuable insight into how these variants effect DNA replication at different stages, and lead to disease. While there is still much to explore, this project has allowed us to better understand how all three variants lead to issues in DNA replication and has a baseline to design further experiment around.

References

Abdelrahman, H.A., John, A., Ali, B.R., and Al-Gazali, L. (2019). Further Delineation of the Microcephaly-Micromelia Syndrome Associated with Loss-of-Function Variants in DONSON. Mol. Syndromol. *10*, 171–176.

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. Nat. Commun. *9*, 1911.

Aird, E.J., Lovendahl, K.N., St. Martin, A., Harris, R.S., and Gordon, W.R. (2018). Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. Commun. Biol. *1*, 54.

Aladjem, M.I. (2007). Replication in context: dynamic regulation of DNA replication patterns in metazoans. Nat. Rev. Genet. *8*, 588–600.

Anderson, E.M., Haupt, A., Schiel, J.A., Chou, E., Machado, H.B., Strezoska, Ž., Lenger, S., McClelland, S., Birmingham, A., Vermeulen, A., et al. (2015). Systematic analysis of CRISPR–Cas9 mismatch tolerance reveals low levels of off-target activity. J. Biotechnol. *211*, 56–65.

Arnott, S., Chandrasekaran, R., Hall, I.H., Puigjaner, L.C., Walker, J.K., and Wang, M. (1983). DNA Secondary Structures: Helices, Wrinkles, and Junctions. Cold Spring Harb. Symp. Quant. Biol. *47*, 53–65.

Ballabeni, A., Zamponi, R., Moore, J.K., Helin, K., and Kirschner, M.W. (2013). Geminin deploys multiple mechanisms to regulate Cdt1 before cell division thus ensuring the proper execution of DNA replication. Proc. Natl. Acad. Sci. *110*, E2848.

Bandura, J.L., Beall, E.L., Bell, M., Silver, H.R., Botchan, M.R., and Calvi, B.R. (2005). humpty dumpty Is Required for Developmental DNA Amplification and Cell Proliferation in Drosophila. Curr. Biol. *15*, 755–759.

Baple, E.L., Chambers, H., Cross, H.E., Fawcett, H., Nakazawa, Y., Chioza, B.A., Harlalka, G.V., Mansour, S., Sreekantan-Nair, A., Patton, M.A., et al. (2014). Hypomorphic PCNA mutation underlies a human DNA repair disorder. J. Clin. Invest. *124*, 3137–3146.

Bartholdy, B., Mukhopadhyay, R., Lajugie, J., Aladjem, M.I., and Bouhassira, E.E. (2015). Allele-specific analysis of DNA replication origins in mammalian cells. Nat. Commun. *6*, 7051.

Beghini, A., Castorina, P., Roversi, G., Modiano, P., and Larizza, L. (2003). RNA processing defects of the helicase gene RECQL4 in a compound heterozygous Rothmund–Thomson patient. Am. J. Med. Genet. A. *120A*, 395–399.

Bell, S.P. (2002). The origin recognition complex: from simple origins to complex functions. Genes Dev. *16*, 659–672.

Bell, S.P., and Labib, K. (2016). Chromosome Duplication in Saccharomyces cerevisiae. Genetics *203*, 1027–1067.

Bell, S.P., and Stillman, B. (1992). ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature *357*, 128.

Bellido, F., Pineda, M., Aiza, G., Valdés-Mas, R., Navarro, M., Puente, D.A., Pons, T., González, S., Iglesias, S., Darder, E., et al. (2016). POLE and POLD1 mutations in 529 kindred with familial colorectal cancer and/or polyposis: review of reported cases and recommendations for genetic testing and surveillance. Genet. Med. *18*, 325–332.

Benchling [Biology Software]) (2019). Benchling (Biology Software).

Benmerzouga, I., Concepción-Acevedo, J., Kim, H.-S., Vandoros, A.V., Cross, G.A.M., Klingbeil, M.M., and Li, B. (2012). Trypanosoma brucei Orc1 is essential for nuclear DNA replication and affects both VSG silencing and VSG switching. Mol. Microbiol. *87*, 196–210.

Beumer, K.J., Trautman, J.K., Bozas, A., Liu, J.-L., Rutter, J., Gall, J.G., and Carroll, D. (2008). Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. Proc. Natl. Acad. Sci. *105*, 19821.

Bicknell, L.S., Walker, S., Klingseisen, A., Stiff, T., Leitch, A., Kerzendorfer, C., Martin, C.-A., Yeyati, P., Al Sanna, N., Bober, M., et al. (2011a). Mutations in ORC1, encoding the largest subunit of the origin recognition complex, cause microcephalic primordial dwarfism resembling Meier-Gorlin syndrome. Nat. Genet. *43*, 350.

Bicknell, L.S., Bongers, E.M.H.F., Leitch, A., Brown, S., Schoots, J., Harley, M.E., Aftimos, S., Al-Aama, J.Y., Bober, M., Brown, P.A.J., et al. (2011b). Mutations in the pre-replication complex cause Meier-Gorlin syndrome. Nat. Genet. *43*, 356.

Bielinsky, A.-K., Leung, W., Shetty, M., Starr, T., and Moldovan, G.-L. (2019). Responses to replication stress in human cells. FASEB J. *33*, 345.1-345.1.

Bleichert, F., Balasov, M., Chesnokov, I., Nogales, E., Botchan, M.R., and Berger, J.M. (2013). A Meier-Gorlin syndrome mutation in a conserved C-terminal helix of Orc6 impedes origin recognition complex formation. ELife *2*.

Bleichert, F., Botchan, M.R., and Berger, J.M. (2017). Mechanisms for initiating cellular DNA replication. Science *355*, eaah6317.

Blow, J.J., and Dutta, A. (2005). Preventing re-replication of chromosomal DNA. Nat. Rev. Mol. Cell Biol. *6*, 476–486.

Blow, J.J., and Ge, X.Q. (2009). A model for DNA replication showing how dormant origins safeguard against replication fork failure. EMBO Rep. *10*, 406–412.

Blow, J.J., Ge, X.Q., and Jackson, D.A. (2011). How dormant origins promote complete genome replication. Trends Biochem. Sci. *36*, 405–414.

Bodnar, A.G., Ouellette, M., Frolkis, M., Holt, S.E., Chiu, C.-P., Morin, G.B., Harley, C.B., Shay, J.W., Lichtsteiner, S., and Wright, W.E. (1998). Extension of Life-Span by Introduction of Telomerase into Normal Human Cells. Science *279*, 349.

Boles, G., R., Teebi, S., A., Schwartz, F., D., and Harper, F., J. (1994). Further delineation of the ear, patella, short stature syndrome (Meier-Gorlin syndrome). Clin. Dysmorphol. *3*, 207–214.

Bongers, E., Otten, B.J., Toutain, A., and et, al (2003). Meier-Gorlin syndrome: delineation of the adult phenotype. Eur. J. Hum. Genet. 95–96.

Bongers, E., Van Kampen, A., Van Bokhoven, H., and Knoers, N. (2005). Human syndromes with congenital patellar anomalies and the underlying gene defects. Clin. Genet. *68*, 302–319.

Bongers, E.M.H.F., Opitz, J.M., Fryer, A., Sarda, P., Hennekam, R.C.M., Hall, B.D., Superneau, D.W., Harbison, M., Poss, A., Bokhoven, H. van, et al. (2001). Meier-Gorlin syndrome: Report of eight additional cases and review. Am. J. Med. Genet. *102*, 115–124.

Bonizzoni, P., Rizzi, R., and Pesole, G. (2006). Computational methods for alternative splicing prediction. Brief. Funct. Genomics *5*, 46–51.

Bozas, A., Beumer, K.J., Trautman, J.K., and Carroll, D. (2009). Genetic analysis of zinc-finger nuclease-induced gene targeting in Drosophila. Genetics *182*, 641–651.

Brogna, S., and Wen, J. (2009). Nonsense-mediated mRNA decay (NMD) mechanisms. Nat. Struct. Mol. Biol. *16*, 107–113.

Brouwer, A.P.M., Bokhoven, H., and Kremer, H. (2006). Comparison of 12 Reference Genes for Normalization of Gene Expression Levels in Epstein-Barr Virus-Transformed Lymphoblastoid Cell Lines and Fibroblasts. Mol. Diagn. Ther. *10*, 197–204.

Burrage, L.C., Charng, W.-L., Eldomery, M.K., Willer, J.R., Davis, E.E., Lugtenberg, D., Zhu, W., Leduc, M.S., Akdemir, Z.C., Azamian, M., et al. (2015). De Novo GMNN Mutations Cause Autosomal-Dominant Primordial Dwarfism Associated with Meier-Gorlin Syndrome. Am. J. Hum. Genet. *97*, 904–913.

Callebaut, I., Courvalin, J.-C., and Mornon, J.-P. (1999). The BAH (bromo-adjacent homology) domain: a link between DNA methylation, replication and transcriptional regulation. FEBS Lett. *446*, 189–193.

Capp, C., Wu, J., and Hsieh, T. (2010). RecQ4: the second replicative helicase? Crit. Rev. Biochem. Mol. Biol. *45*, 233–242.

Cardoso-Moreira, M., Halbert, J., Valloton, D., Velten, B., Chen, C., Shao, Y., Liechti, A., Ascenção, K., Rummel, C., Ovchinnikova, S., et al. (2019). Gene expression across mammalian organ development. Nature *571*, 505–509.

Casey, J.P., Nobbs, M., McGettigan, P., Lynch, S., and Ennis, S. (2012). Recessive mutations in MCM4 PRKDC cause a novel syndrome involving a primary immunodeficiency and a disorder of DNA repair. J. Med. Genet. *49*, 242.

Cayrou, C., Coulombe, P., Vigneron, A., Stanojcic, S., Ganier, O., Peiffer, I., Rivals, E., Puy, A., Laurent-Chabalier, S., Desprat, R., et al. (2011). Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. Genome Res. *21*, 1438–1449.

Chen, Z., Speck, C., Wendel, P., Tang, C., Stillman, B., and Li, H. (2008). The architecture of the DNA replication origin recognition complex in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U. S. A. *105*, 10326–10331.

Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., and Kühn, R. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat. Biotechnol. *33*, 543–548.

Codd, P.J., Scott, R.M., and Smith, E.R. (2009). Seckel syndrome and moyamoya. J. Neurosurg. Pediatr. PED *3*, 320–324.

Cohen, B., Temple, I.K., Symons, J.C., Hall, C.M., Shaw, D.G., Bhamra, M., Jackson, A.M., and Pembrey, M.E. (1991). Microtia and short stature: a new syndrome. J. Med. Genet. *28*, 786.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science *339*, 819.

Cooper, F.A. (2016). The role of ORC1 in cell proliferation and regulation of organism growth. PhD. University of Edinburgh.

Costa, A., Ilves, I., Tamberg, N., Petojevic, T., Nogales, E., Botchan, M.R., and Berger, J.M. (2011). The structural basis for MCM2–7 helicase activation by GINS and Cdc45. Nat. Struct. Amp Mol. Biol. *18*, 471.

Cottineau, J., Kottemann, M.C., Lach, F.P., Kang, Y.-H., Vély, F., Deenick, E.K., Lazarov, T., Gineau, L., Wang, Y., Farina, A., et al. (2017). Inherited GINS1 deficiency underlies growth retardation along with neutropenia and NK cell deficiency. J. Clin. Invest. *127*, 1991–2006.

Daigaku, Y., Davies, A.A., and Ulrich, H.D. (2010). Ubiquitin-dependent DNA damage bypass is separable from genome replication. Nature *465*, 951–955.

Delgado, S., Gómez, M., Bird, A., and Antequera, F. (1998). Initiation of DNA replication at CpG islands in mammalian chromosomes. EMBO J. *17*, 2426–2435.

van Deursen, F., Sengupta, S., De Piccoli, G., Sanchez-Diaz, A., and Labib, K. (2012). Mcm10 associates with the loaded DNA helicase at replication origins and defines a novel step in its activation. EMBO J. *31*, 2195–2206.

Dhar, S.K., Yoshida, K., Machida, Y., Khaira, P., Chaudhuri, B., Wohlschlegel, J.A., Leffak, M., Yates, J., and Dutta, A. (2001). Replication from oriP of Epstein-Barr Virus Requires Human ORC and Is Inhibited by Geminin. Cell *106*, 287–296.

Diamond, D.J., Clayton, L.K., Sayre, P.H., and Reinherz, E.L. (1988). Exon-intron organization and sequence comparison of human and murine T11 (CD2) genes. Proc. Natl. Acad. Sci. *85*, 1615.

Dickinson, M.E., Flenniken, A.M., Ji, X., Teboul, L., Wong, M.D., White, J.K., Meehan, T.F., Weninger, W.J., Westerberg, H., Adissu, H., et al. (2016). High-throughput discovery of novel developmental phenotypes. Nature *537*, 508–514.

DiNapoli, S.E., Martinez-McFaline, R., Gribbin, C.K., Wrighton, P.J., Balgobin, C.A., Nelson, I., Leonard, A., Maskin, C.R., Shwartz, A., Quenzer, E.D., et al. (2020). Synthetic CRISPR/Cas9 reagents facilitate genome editing and homology directed repair. Nucleic Acids Res.

Dueber, E.C., Costa, A., Corn, J.E., Bell, S.D., and Berger, J.M. (2011). Molecular determinants of origin discrimination by Orc1 initiators in archaea. Nucleic Acids Res. *39*, 3621–3631.

Dueber, E.L.C., Corn, J.E., Bell, S.D., and Berger, J.M. (2007). Replication Origin Recognition and Deformation by a Heterodimeric Archaeal Orc1 Complex. Science *317*, 1210.

Dupré, A., Boyer-Chatenet, L., and Gautier, J. (2006). Two-step activation of ATM by DNA and the Mre11–Rad50–Nbs1 complex. Nat. Struct. Mol. Biol. *13*, 451–457.

Edery, P., Marcaillou, C., Sahbatou, M., Labalme, A., Chastang, J., Touraine, R., Tubacher, E., Senni, F., Bober, M.B., Nampoothiri, S., et al. (2011). Association of TALS Developmental Disorder with Defect in Minor Splicing Component U4atac snRNA. Science *332*, 240.

Eidenschenk, C., Dunne, J., Jouanguy, E., Fourlinnie, C., Gineau, L., Bacq, D., McMahon, C., Smith, O., Casanova, J.-L., Abel, L., et al. (2006). A Novel Primary Immunodeficiency with Specific Natural-Killer Cell Deficiency Maps to the Centromeric Region of Chromosome 8. Am. J. Hum. Genet. *78*, 721–727.

Elsayed, F.A., Kets, C.M., Ruano, D., van den Akker, B., Mensenkamp, A.R., Schrumpf, M., Nielsen, M., Wijnen, J.T., Tops, C.M., Ligtenberg, M.J., et al. (2015). Germline variants in POLE are associated with early onset mismatch repair deficient colorectal cancer. Eur. J. Hum. Genet. *23*, 1080–1084.

Ercilla, A., Llopis Gómez, A., Feu, S., Aranda, S., Ernfors, P., Freire, R., and Agell, N. (2016). New origin firing is inhibited by APC/C-Cdh1 activation in S-phase after severe replication stress. Nucleic Acids Res. *44*.

Evrin, C., Clarke, P., Zech, J., Lurz, R., Sun, J., Uhle, S., Li, H., Stillman, B., and Speck, C. (2009). A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. Proc. Natl. Acad. Sci. *106*, 20240.

Evrony, G.D., Cordero, D.R., Shen, J., Partlow, J.N., Yu, T.W., Rodin, R.E., Hill, R.S., Coulter, M.E., Lam, A.-T.N., Jayaraman, D., et al. (2017). Integrated genome and transcriptome sequencing identifies a noncoding mutation in the genome replication factor as the cause of microcephaly-micromelia syndrome. Genome Res. *27*, 1323.

Fenwick, A.L., Kliszczak, M., Cooper, F., Murray, J., Sanchez-Pulido, L., Twigg, S.R.F., Goriely, A., McGowan, S.J., Miller, K.A., Taylor, I.B., et al. (2016). Mutations in CDC45, Encoding an Essential Component of the Pre-initiation Complex, Cause Meier-Gorlin Syndrome and Craniosynostosis. Am. J. Hum. Genet. *99*, 125–138.

Feringa, F.M., Raaijmakers, J.A., Hadders, M.A., Vaarting, C., Macurek, L., Heitink, L., Krenning, L., and Medema, R.H. (2018). Persistent repair intermediates induce senescence. Nat. Commun. *9*, 3923.

Fernández-Cid, A., Riera, A., Tognetti, S., Herrera, M.C., Samel, S., Evrin, C., Winkler, C., Gardenal, E., Uhle, S., and Speck, C. (2013). An ORC/Cdc6/MCM2-7 Complex Is Formed in a Multistep Reaction to Serve as a Platform for MCM Double-Hexamer Assembly. Mol. Cell *50*, 577–588.

Fragkos, M., and Naim, V. (2017). Rescue from replication stress during mitosis. Cell Cycle *16*, 613–633.

Francis, L.I., Randell, J.C.W., Takara, T.J., Uchima, L., and Bell, S.P. (2009). Incorporation into the prereplicative complex activates the Mcm2-7 helicase for Cdc7-Dbf4 phosphorylation. Genes Dev. *23*, 643–654.

Frit, P., Barboule, N., Yuan, Y., Gomez, D., and Calsou, P. (2014). Alternative end-joining pathway(s): Bricolage at DNA breaks. Recent Dev. Non-Homol. End Join. *17*, 81–97.

Fuchs, F., Pau, G., Kranz, D., Sklyar, O., Budjan, C., Steinbrink, S., Horn, T., Pedal, A., Huber, W., and Boutros, M. (2010). Clustering phenotype populations by genome-wide RNAi and multiparametric imaging. Mol. Syst. Biol. *6*, 370.

Galea, P., and Tolmie, J.L. (1990). Normal growth and development in a child with Baller-Gerold syndrome (craniosynostosis and radial aplasia). J. Med. Genet. *27*, 784.

Gambus, A. (2017). Termination of Eukaryotic Replication Forks. In DNA Replication: From Old Principles to New Discoveries, H. Masai, and M. Foiani, eds. (Singapore: Springer Singapore), pp. 163–187.

Gambus, A., van Deursen, F., Polychronopoulos, D., Foltman, M., Jones, R.C., Edmondson, R.D., Calzada, A., and Labib, K. (2009). A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase alpha within the eukaryotic replisome. EMBO J. *28*, 2992–3004.

Gambus, A., Khoudoli, G.A., Jones, R.C., and Blow, J.J. (2011). MCM2-7 Form Double Hexamers at Licensed Origins in Xenopus Egg Extract. J. Biol. Chem. *286*, 11855–11864.

Garg, P., Stith, C.M., Sabouri, N., Johansson, E., and Burgers, P.M. (2004). Idling by DNA polymerase delta maintains a ligatable nick during lagging-strand DNA replication. Genes Dev. *18*, 2764–2773.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. U. S. A. *109*, E2579–E2586.

Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature *551*, 464–471.

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

Gennery, A.R., Cant, A.J., and Jeggo, P.A. (2000). Immunodeficiency associated with DNA repair defects. Clin. Exp. Immunol. *121*, 1–7.

Gineau, L., Cognet, C., Kara, N., Lach, F.P., Dunne, J., Veturi, U., Picard, C., Trouillet, C., Eidenschenk, C., Aoufouchi, S., et al. (2012). Partial MCM4 deficiency in patients with growth retardation, adrenal insufficiency, and natural killer cell deficiency. J. Clin. Invest. *122*, 821–832.

Giuliano, C.J., Lin, A., Girish, V., and Sheltzer, J.M. (2019). Generating Single Cell–Derived Knockout Clones in Mammalian Cells with CRISPR/Cas9. Curr. Protoc. Mol. Biol. *128*, e100.

Gorlin, R.J., Cervenka, J., Moller, K., Horrobin, M., and Witkop, C.J.J. (1975). Malformation syndromes. A selected miscellany. Birth Defects Orig Artic Ser *11*, 39–50.

Gu, B., Posfai, E., and Rossant, J. (2018). Efficient generation of targeted large insertions by microinjection into two-cell-stage mouse embryos. Nat. Biotechnol. *36*, 632.

Guernsey, D.L., Matsuoka, M., Jiang, H., Evans, S., Macgillivray, C., Nightingale, M., Perry, S., Ferguson, M., LeBlanc, M., Paquette, J., et al. (2011). Mutations in origin recognition complex gene ORC4 cause Meier-Gorlin syndrome. Nat. Genet. *43*, 360.

Gurley, K.E., Vo, K., and Kemp, C.J. (1998). DNA Double-Strand Breaks, p53, and Apoptosis during Lymphomagenesis in scid/scid Mice. Cancer Res. *58*, 3111.

Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., and Taipale, J. (2018). CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. Nat. Med. *24*, 927–930.

Hanel, W., and Moll, U.M. (2012). Links between mutant p53 and genomic instability. J. Cell. Biochem. *113*, 433–439.

Hardy, C.F.J. (1997). Identification of Cdc45p, an essential factor required for DNA replication. Gene *187*, 239–246.

Harley, M.E., Murina, O., Leitch, A., Higgs, M.R., Bicknell, L.S., Yigit, G., Blackford, A.N., Zlatanou, A., Mackenzie, K.J., Reddy, K., et al. (2015). TRAIP promotes DNA damage response during genome replication and is mutated in primordial dwarfism. Nat. Genet. *48*, 36.

He, H., Liyanarachchi, S., Akagi, K., Nagy, R., Li, J., Dietrich, R.C., Li, W., Sebastian, N., Wen, B., Xin, B., et al. (2011). Mutations in U4atac snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I. Science *332*, 238–240.

Hemerly, A.S., Prasanth, S.G., Siddiqui, K., and Stillman, B. (2009). Orc1 Controls Centriole and Centrosome Copy Number in Human Cells. Science *323*, 789.

Herbig, U., Marlar, C.A., and Fanning, E. (1999). The Cdc6 Nucleotide–Binding Site Regulates Its Activity in DNA Replication in Human Cells. Mol. Biol. Cell *10*, 2631–2645.

Horii, T., and Hatada, I. (2016). Genome-Editing Technology in CRISPR/Cas System: How to Increase Knock-In Efficiency in Mouse Zygotes. In Genome Editing, K. Turksen, ed. (Cham: Springer International Publishing), pp. 79–91.

Hossain, M., and Stillman, B. (2016). Meier-Gorlin Syndrome. In The Initiation of DNA Replication in Eukaryotes, D.L. Kaplan, ed. (Cham: Springer International Publishing), pp. 503–524.

Hua, B.L., and Orr-Weaver, T.L. (2017). DNA Replication Control During Drosophila Development: Insights into the Onset of S Phase, Replication Initiation, and Fork Progression. Genetics *207*, 29.

Hughes, C.R., Guasti, L., Meimaridou, E., Chuang, C.-H., Schimenti, J.C., King, P.J., Costigan, C., Clark, A.J.L., and Metherell, L.A. (2012). MCM4 mutation causes adrenal failure, short stature, and natural killer cell deficiency in humans. J. Clin. Invest. *122*, 814–820.

Ibarra, A., Schwob, E., and Méndez, J. (2008). Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication. Proc. Natl. Acad. Sci. U. S. A. *105*, 8956–8961.

Ichijima, Y., Sakasai, R., Okita, N., Asahina, K., Mizutani, S., and Teraoka, H. (2005). Phosphorylation of histone H2AX at M phase in human cells without DNA damage response. Biochem. Biophys. Res. Commun. *336*, 807–812.

Ihry, R.J., Worringer, K.A., Salick, M.R., Frias, E., Ho, D., Theriault, K., Kommineni, S., Chen, J., Sondey, M., Ye, C., et al. (2018). p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. Nat. Med. *24*, 939–946.

Ilves, I., Petojevic, T., Pesavento, J.J., and Botchan, M.R. (2010). Activation of the MCM2-7 Helicase by Association with Cdc45 and GINS Proteins. Mol. Cell *37*, 247–258.

Jackson, D.A., and Pombo, A. (1998). Replicon Clusters Are Stable Units of Chromosome Structure: Evidence That Nuclear Organization Contributes to the Efficient Activation and Propagation of S Phase in Human Cells. J. Cell Biol. *140*, 1285.

Jackson, A.P., Laskey, R.A., and Coleman, N. (2014). Replication proteins and human disease. Cold Spring Harb. Perspect. Biol. *6*, a013060.

Jayathilaka, K., Sheridan, S.D., Bold, T.D., Bochenska, K., Logan, H.L., Weichselbaum, R.R., Bishop, D.K., and Connell, P.P. (2008). A chemical compound that stimulates the human homologous recombination protein RAD51. Proc. Natl. Acad. Sci. U. S. A. *105*, 15848–15853.

Jiang, X.-R., Jimenez, G., Chang, E., Frolkis, M., Kusler, B., Sage, M., Beeche, M., Bodnar, A.G., Wahl, G.M., Tlsty, T.D., et al. (1999). Telomerase expression in human somatic cells does not induce changes associated with a transformed phenotype. Nat. Genet. *21*, 111.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science *337*, 816.

Kääriäinen, H., Ryöppy, S., and Norio, R. (1989). RAPADILINO syndrome with radial and patellar aplasia/hypoplasia as main manifestations. Am. J. Med. Genet. *33*, 346–351.

Kaina, B. (2003). DNA damage-triggered apoptosis: critical role of DNA repair, double-strand breaks, cell proliferation and signaling. Apoptosis - Signal. Pathw. Ther. Tools *66*, 1547–1554.

Kalogeropoulou, A., Lygerou, Z., and Taraviras, S. (2019). Cortical Development and Brain Malformations: Insights From the Differential Regulation of Early Events of DNA Replication. Front. Cell Dev. Biol. *7*, 29.

Kantaputra, P., Tanpaiboon, P., Porntaveetus, T., Ohazama, A., Sharpe, P., Rauch, A., Hussadaloy, A., and Thiel, C.T. (2011). The smallest teeth in the world are caused by mutations in the PCNT gene. Am. J. Med. Genet. A. *155*, 1398–1403.

Kara, N., Hossain, M., Prasanth, S.G., and Stillman, B. (2015). Orc1 Binding to Mitotic Chromosomes Precedes Spatial Patterning during G1 Phase and Assembly of the Origin Recognition Complex in Human Cells. J. Biol. Chem. *290*, 12355–12369.

Karaca, E., Posey, J.E., Bostwick, B., Liu, P., Gezdirici, A., Yesil, G., Coban Akdemir, Z., Bayram, Y., Harms, F.L., Meinecke, P., et al. (2019). Biallelic and De Novo Variants in DONSON Reveal a Clinical Spectrum of Cell Cycle-opathies with Microcephaly, Dwarfism and Skeletal Abnormalities. Am. J. Med. Genet. A. *0*.

Karnani, N., and Dutta, A. (2011). The effect of the intra-S-phase checkpoint on origins of replication in human cells. Genes Dev. *25*, 621–633.

Kashiwagi, H., Shiraishi, K., Sakaguchi, K., Nakahama, T., and Kodama, S. (2018). Repair kinetics of DNA double-strand breaks and incidence of apoptosis in mouse neural stem/progenitor cells and their differentiated neurons exposed to ionizing radiation. J. Radiat. Res. (Tokyo) *59*, 261–271.

Katoh, Y., Michisaka, S., Nozaki, S., Funabashi, T., Hirano, T., Takei, R., and Nakayama, K. (2017). Practical method for targeted disruption of cilia-related genes by using CRISPR/Cas9-mediated, homology-independent knock-in system. Mol. Biol. Cell *28*, 898–906.

Kellermayer, R. (2006). The versatile RECQL4. Genet. Med. *8*, 213–216.

Kemp, M.G., Akan, Z., Yilmaz, S., Grillo, M., Smith-Roe, S.L., Kang, T.-H., Cordeiro-Stone, M., Kaufmann, W.K., Abraham, R.T., Sancar, A., et al. (2010). Tipin-Replication Protein A Interaction Mediates Chk1 Phosphorylation by ATR in Response to Genotoxic Stress. J. Biol. Chem. *285*, 16562–16571.

Kerzendorfer, C., Colnaghi, R., Abramowicz, I., Carpenter, G., and O'Driscoll, M. (2013). Meier–Gorlin syndrome and Wolf–Hirschhorn syndrome: Two developmental disorders highlighting the importance of efficient DNA replication for normal development and neurogenesis. Genome Maint. Nerv. Syst. *12*, 637–644.

Khakhar, R.R., Cobb, J.A., Bjergbaek, L., Hickson, I.D., and Gasser, S.M. (2003). RecQ helicases: multiple roles in genome maintenance. Trends Cell Biol. *13*, 493–501.

Khayrutdinov, B.I., Bae, W.J., Yun, Y.M., Lee, J.H., Tsuyama, T., Kim, J.J., Hwang, E., Ryu, K.-S., Cheong, H.-K., Cheong, C., et al. (2009). Structure of the Cdt1 C-terminal domain: conservation of the winged helix fold in replication licensing factors. Protein Sci. Publ. Protein Soc. *18*, 2252–2264.

Khetarpal, P., Das, S., Panigrahi, I., and Munshi, A. (2016). Primordial dwarfism: overview of clinical and genetic aspects. Mol. Genet. Genomics *291*, 1–15.

Kim, Y., and Kipreos, E.T. (2007). Cdt1 degradation to prevent DNA re-replication: conserved and non-conserved pathways. Cell Div. *2*, 18.

Kim, J., Feng, H., and Kipreos, E.T. (2007). C. elegans CUL-4 Prevents Rereplication by Promoting the Nuclear Export of CDC-6 via a CKI-1-Dependent Pathway. Curr. Biol. *17*, 966–972.

Kim, K., Ryu, S.-M., Kim, S.-T., Baek, G., Kim, D., Lim, K., Chung, E., Kim, S., and Kim, J.-S. (2017a). Highly efficient RNA-guided base editing in mouse embryos. Nat. Biotechnol. *35*, 435–437.

Kim, Y.B., Komor, A.C., Levy, J.M., Packer, M.S., Zhao, K.T., and Liu, D.R. (2017b). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nat. Biotechnol. *35*, 371–376.

Kishore, S., Khanna, A., and Stamm, S. (2008). Rapid generation of splicing reporters with pSpliceExpress. Gene *427*, 104–110.

Kitao, S., Ohsugi, I., Ichikawa, K., Goto, M., Furuichi, Y., and Shimamoto, A. (1998). Cloning of Two New Human Helicase Genes of the RecQ Family: Biological Significance of Multiple Species in Higher Eukaryotes. Genomics *54*, 443–452.

Kitao, S., Shimamoto, A., Goto, M., Miller, R.W., Smithson, W.A., Lindor, N.M., and Furuichi, Y. (1999). Mutations in RECQL4 cause a subset of cases of Rothmund-Thomson syndrome. Nat. Genet. *22*, 82–84.

Klein, K.N., and Gilbert, D.M. (2016). Epigenetic vs. Sequence-Dependent Control of Eukaryotic Replication Timing. In The Initiation of DNA Replication in Eukaryotes, D.L. Kaplan, ed. (Cham: Springer International Publishing), pp. 39–63.

Klingseisen, A., and Jackson, A.P. (2011). Mechanisms and pathways of growth failure in primordial dwarfism. Genes Dev. *25*, 2011–2024.

Knapp, K.M., **Sullivan, R**., Murray, J., Gimenez, G., Arn, P., D'Souza, P., Gezdirici, A., Wilson, W.G., Jackson, A.P., Ferreira, C., et al. (2019). Linked-read genome sequencing identifies biallelic pathogenic variants in DONSON as a novel cause of Meier-Gorlin syndrome. J. Med. Genet. jmedgenet-2019-106396.

Köhler, C., Koalick, D., Fabricius, A., Parplys, A.C., Borgmann, K., Pospiech, H., and Grosse, F. (2016). Cdc45 is limiting for replication initiation in humans. Cell Cycle Georget. Tex *15*, 974–985.

Kolesnikova, T.D. (2013). Regulation of DNA replication timing. Mol. Biol. *47*, 12–33.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature *533*, 420–424.

Kopito, R.R. (2000). Aggresomes, inclusion bodies and protein aggregation. Trends Cell Biol. *10*, 524–530.

Kora, N., Nayek, K., Soren, B., and Das, R. (2016). Meier-Gorlin syndrome with ventriculomegaly and hypoplastic corpus callosum: a rarely reported congenital malformation. J Pediatr Neonat Individ. Med *5*, e050123.

Koressaar, T., and Remm, M. (2007). Enhancements and modifications of primer design program Primer3. Bioinformatics *23*, 1289–1291.

Krasowska, J., Olasek, M., Bzowska, A., Clark, P.L., and Wielgus-Kutrowska, B. (2010). The comparison of aggregation and folding of enhanced green fluorescent protein (EGFP) by spectroscopic studies. Spectroscopy *24*, 186903.

Kunkel, T.A., and Burgers, P.M. (2014). Delivering nonidentical twins. Nat. Struct. Mol. Biol. *21*, 649–651.

Kuo, A.J., Song, J., Cheung, P., Ishibe-Murakami, S., Yamazoe, S., Chen, J.K., Patel, D.J., and Gozani, O. (2012). The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier–Gorlin syndrome. Nature *484*, 115.

Kuznetsova, A.V., Kurinov, A.M., and Aleksandrova, M.A. (2014). Cell Models to Study Regulation of Cell Transformation in Pathologies of Retinal Pigment Epithelium. J. Ophthalmol. *2014*.

Kwart, D., Paquet, D., Teo, S., and Tessier-Lavigne, M. (2017). Precise and efficient scarless genome editing in stem cells using CORRECT. Nat. Protoc. *12*, 329–354.

Labun, K., Montague, T.G., Krause, M., Torres Cleuren, Y.N., Tjeldnes, H., and Valen, E. (2019). CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. Nucleic Acids Res. *47*, W171–W174.

Lahtela, J., Corson, L.B., Hemmes, A., Brauer, M.J., Koopal, S., Lee, J., Hunsaker, T.L., Jackson, P.K., and Verschuren, E.W. (2013). A high-content cellular senescence screen identifies candidate tumor suppressors, including EPHA3. Cell Cycle Georget. Tex *12*, 625–634.

Larizza, L., Roversi, G., and Volpi, L. (2010). Rothmund-Thomson syndrome. Orphanet J. Rare Dis. *5*, 2.

Lee, H.K., Willi, M., Miller, S.M., Kim, S., Liu, C., Liu, D.R., and Hennighausen, L. (2018). Targeting fidelity of adenine and cytosine base editors in mouse embryos. Nat. Commun. *9*, 4804.

Lesly, S., Bandura, J.L., and Calvi, B.R. (2017). Rapid DNA Synthesis During Early Drosophila Embryogenesis Is Sensitive to Maternal Humpty Dumpty Protein Function. Genetics *207*, 935.

Li, B., Reddy, S., and Comai, L. (2011). Depletion of Ku70/80 reduces the levels of extrachromosomal telomeric circles and inhibits proliferation of ALT cells. Aging *3*, 395–406.

Li, G., Zhang, X., Zhong, C., Mo, J., Quan, R., Yang, J., Liu, D., Li, Z., Yang, H., and Wu, Z. (2017). Small molecules enhance CRISPR/Cas9-mediated homology-directed genome editing in primary cells. Sci. Rep. *7*, 8943–8943.

Li, K., Wang, G., Andersen, T., Zhou, P., and Pu, W.T. (2014). Optimization of genome engineering approaches with the CRISPR/Cas9 system. PloS One *9*, e105779–e105779.

Li, S.C., Goto, N.K., Williams, K.A., and Deber, C.M. (1996). Alpha-helical, but not beta-sheet, propensity of proline is determined by peptide environment. Proc. Natl. Acad. Sci. *93*, 6676.

Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu. Rev. Biochem. *79*, 181–211.

Lieber, M.R., Gu, J., Lu, H., Shimazaki, N., and Tsai, A.G. (2010). Nonhomologous DNA end joining (NHEJ) and chromosomal translocations in humans. Subcell. Biochem. *50*, 279–296.

Lin, J.J., Milhollen, M.A., Smith, P.G., Narayanan, U., and Dutta, A. (2010). NEDD8-Targeting Drug MLN4924 Elicits DNA Rereplication by Stabilizing Cdt1 in S Phase, Triggering Checkpoint Activation, Apoptosis, and Senescence in Cancer Cells. Cancer Res. *70*, 10310.

Lindor, N.M., Furuichi, Y., Kitao, S., Shimamoto, A., Arndt, C., and Jalal, S. (2000). Rothmund-Thomson syndrome due to RECQ4 helicase mutations: Report and clinical and molecular comparisons with Bloom syndrome and Werner syndrome. Am. J. Med. Genet. *90*, 223–228.

Ling, Y.H., Wong, C.C., Li, K.W., Chan, K.M., Boukamp, P., and Liu, W.K. (2014). CCHCR1 interacts with EDC4, suggesting its localization in P-bodies. Exp. Cell Res. *327*, 12–23.

Liu, C., Wu, R., Zhou, B., Wang, J., Wei, Z., Tye, B.K., Liang, C., and Zhu, G. (2012). Structural insights into the Cdt1-mediated MCM2-7 chromatin loading. Nucleic Acids Res. *40*, 3208–3217.

Liu, M., Rehman, S., Tang, X., Gu, K., Fan, Q., Chen, D., and Ma, W. (2019). Methodologies for Improving HDR Efficiency. Front. Genet. *9*, 691–691.

Liu, Z., Lu, Z., Yang, G., Huang, S., Li, G., Feng, S., Liu, Y., Li, J., Yu, W., Zhang, Y., et al. (2018). Efficient generation of mouse models of human diseases via ABE- and BE-mediated base editing. Nat. Commun. *9*, 2338.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. Methods *25*, 402–408.

Logan, C.V., Murray, J.E., Parry, D.A., Robertson, A., Bellelli, R., Tarnauskaitė, Ž., Challis, R., Cleal, L., Borel, V., Fluteau, A., et al. (2018). DNA Polymerase Epsilon Deficiency Causes IMAGe Syndrome with Variable Immunodeficiency. Am. J. Hum. Genet. *103*, 1038–1044.

Luo, Y., Na, Z., and Slavoff, S.A. (2018). P-Bodies: Composition, Properties, and Functions. Biochemistry *57*, 2424–2431.

Machida, Y.J., Hamlin, J.L., and Dutta, A. (2005). Right Place, Right Time, and Only Once: Replication Initiation in Metazoans. Cell *123*, 13–24.

Maguire, J.A., Cardenas-Diaz, F.L., Gadue, P., and French, D.L. (2019). Highly Efficient CRISPR-Cas9-Mediated Genome Editing in Human Pluripotent Stem Cells. Curr. Protoc. Stem Cell Biol. *48*, e64.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013a). RNA-Guided Human Genome Engineering via Cas9. Science *339*, 823.

Mali, P., Esvelt, K.M., and Church, G.M. (2013b). Cas9 as a versatile tool for engineering biology. Nat. Methods *10*, 957–963.

Mao, Z., Bozzella, M., Seluanov, A., and Gorbunova, V. (2008). Comparison of nonhomologous end joining and homologous recombination in human cells. DNA Repair *7*, 1765–1771.

Maruyama, T., Dougan, S.K., Truttmann, M.C., Bilate, A.M., Ingram, J.R., and Ploegh, H.L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nat. Biotechnol. *33*, 538–542.

Matsuno, K., Kumano, M., Kubota, Y., Hashimoto, Y., and Takisawa, H. (2006). The N-Terminal Noncatalytic Region of Xenopus RecQ4 Is Required for Chromatin Binding of DNA Polymerase α in the Initiation of DNA Replication. Mol. Cell. Biol. *26*, 4843.

Mazouzi, A., Velimezi, G., and Loizou, J.I. (2014). DNA replication stress: Causes, resolution and disease. DNA DAMAGE REPAIR *329*, 85–93.

McConnell Smith, A., Takeuchi, R., Pellenz, S., Davis, L., Maizels, N., Monnat, R.J., and Stoddard, B.L. (2009). Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. Proc. Natl. Acad. Sci. *106*, 5099.

McIntosh, D., and Blow, J.J. (2012). Dormant origins, the licensing checkpoint, and the response to replicative stresses. Cold Spring Harb. Perspect. Biol. *4*, a012955.

Méchali, M. (2010). Eukaryotic DNA replication origins: many choices for appropriate answers. Nat. Rev. Mol. Cell Biol. *11*, 728.

Meehan, T.F., Conte, N., West, D.B., Jacobsen, J.O., Mason, J., Warren, J., Chen, C.-K., Tudose, I., Relac, M., Matthews, P., et al. (2017). Disease model discovery from 3,328 gene knockouts by The International Mouse Phenotyping Consortium. Nat. Genet. *49*, 1231.

Meier, Z., Poschiavo, and Rothschild, M. (1959). Case of arthrogryposis multiplex congenita with mandibulofacial dysostosis (Franceschetti syndrome). Helv Paediatr Acta *14*, 213–216.

Mendoza-Maldonado, R., Paolinelli, R., Galbiati, L., Giadrossi, S., and Giacca, M. (2010). Interaction of the Retinoblastoma Protein with Orc1 and Its Recruitment to Human Origins of DNA Replication. PLOS ONE *5*, e13720.

Menon, V., and Povirk, L. (2014). Involvement of p53 in the repair of DNA double strand breaks: multifaceted Roles of p53 in homologous recombination repair (HRR) and non-homologous end joining (NHEJ). Subcell. Biochem. *85*, 321–336.

Miyazawa-Onami, M., Araki, H., and Tanaka, S. (2017). Pre-initiation complex assembly functions as a molecular switch that splits the Mcm2-7 double hexamer. EMBO Rep. *18*, 1752–1761.

Mo, D., Zhao, Y., and Balajee, A.S. (2018). Human RecQL4 helicase plays multifaceted roles in the genomic stability of normal and cancer cells. Cancer Lett. *413*, 1–10.

Montanari, M., Macaluso, M., Cittadini, A., and Giordano, A. (2006). Role of geminin: from normal control of DNA replication to cancer formation and progression? Cell Death Differ. *13*, 1052–1056.

Moon, S.B., Kim, D.Y., Ko, J.-H., and Kim, Y.-S. (2019). Recent advances in the CRISPR genome editing tool set. Exp. Mol. Med. *51*, 1–11.

Moyer, S.E., Lewis, P.W., and Botchan, M.R. (2006). Isolation of the Cdc45/Mcm2–7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. Proc. Natl. Acad. Sci. *103*, 10236.

Müller, E., Dunstheimer, D., Klammt, J., Friebe, D., Kiess, W., Kratzsch, J., Kruis, T., Laue, S., Pfäffle, R., Wallborn, T., et al. (2012). Clinical and functional characterization of a patient carrying a compound heterozygous pericentrin mutation and a heterozygous IGF1 receptor mutation. PloS One *7*, e38220–e38220.

de Munnik, S.A., Bicknell, L.S., Aftimos, S., Al-Aama, J.Y., van Bever, Y., Bober, M.B., Clayton-Smith, J., Edrees, A.Y., Feingold, M., Fryer, A., et al. (2012a). Meier–Gorlin syndrome genotype–phenotype studies: 35 individuals with pre-replication complex gene mutations and 10 without molecular diagnosis. Eur. J. Hum. Genet. *20*, 598.

de Munnik, S.A., Otten, B.J., Schoots, J., Bicknell, L.S., Aftimos, S., Al-Aama, J.Y., van Bever, Y., Bober, M.B., Borm, G.F., Clayton-Smith, J., et al. (2012b). Meier–Gorlin syndrome: Growth and secondary sexual development of a microcephalic primordial dwarfism disorder. Am. J. Med. Genet. A. *158A*, 2733–2742.

de Munnik, S.A., Hoefsloot, E.H., Roukema, J., Schoots, J., Knoers, N.V., Brunner, H.G., Jackson, A.P., and Ernie MHF Bongers (2015). Meier-Gorlin syndrome. Orphanet J. Rare Dis. *10*.

Muramatsu, S., Hirai, K., Tak, Y.-S., Kamimura, Y., and Araki, H. (2010). CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol (epsilon}, and GINS in budding yeast. Genes Dev. *24*, 602–612.

Musiałek, M.W., and Rybaczek, D. (2015). Behavior of replication origins in Eukaryota – spatio-temporal dynamics of licensing and firing. Cell Cycle *14*, 2251–2264.

Nambiar, T.S., Billon, P., Diedenhofen, G., Hayward, S.B., Taglialatela, A., Cai, K., Huang, J.-W., Leuzzi, G., Cuella-Martin, R., Palacios, A., et al. (2019). Stimulation of CRISPR-mediated homology-directed repair by an engineered RAD18 variant. Nat. Commun. *10*, 3395.

Neuwald, A.F., Aravind, L., Spouge, J.L., and Koonin, E.V. (1999). AAA+: A Class of Chaperone-Like ATPases Associated with the Assembly, Operation, and Disassembly of Protein Complexes. Genome Res. *9*, 27–43.

Nielsen, S., Bassler, N., Grzanka, L., Swakon, J., Olko, P., Andreassen, C.N., Alsner, J., and Sørensen, B.S. (2018). Optimal reference genes for normalization of qPCR gene expression data from proton and photon irradiated dermal fibroblasts. Sci. Rep. *8*, 12688.

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K.Y., et al. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science *353*, aaf8729.

Noguchi, K., Vassilev, A., Ghosh, S., Yates, J.L., and DePamphilis, M.L. (2006). The BAH domain facilitates the ability of human Orc1 protein to activate replication origins in vivo. EMBO J. *25*, 5372.

van Nuland, R., and Gozani, O. (2016). Histone H4 Lysine 20 (H4K20) Methylation, Expanding the Signaling Potential of the Proteome One Methyl Moiety at a Time. Mol. Cell. Proteomics MCP *15*, 755–764.

O'Brien, A.R., Wilson, L.O.W., Burgio, G., and Bauer, D.C. (2019). Unlocking HDR-mediated nucleotide editing by identifying high-efficiency target sites using machine learning. Sci. Rep. *9*, 2788–2788.

O'Donnell, M., Langston, L., and Stillman, B. (2013). Principles and Concepts of DNA Replication in Bacteria and Eukarya. Cold Spring Harb. Perspect. Biol. *5*, 10.1101/cshperspect.a010108 a010108.

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

Ogawa, T., and Okazaki, T. (1980). Discontinuous DNA Replication. Annu. Rev. Biochem. *49*, 421–457.

Pacek, M., and Walter, J.C. (2004). A requirement for MCM7 and Cdc45 in chromosome unwinding during eukaryotic DNA replication. EMBO J. *23*, 3667.

Pachlopnik Schmid, J., Lemoine, R., Nehme, N., Cormier-Daire, V., Revy, P., Debeurme, F., Debré, M., Nitschke, P., Bole-Feysot, C., Legeai-Mallet, L., et al. (2012). Polymerase ε1 mutation in a human syndrome with facial dysmorphism, immunodeficiency, livedo, and short stature ("FILS syndrome"). J. Exp. Med. *209*, 2323–2330.

Palles, C., Cazier, J.-B., Howarth, K.M., Domingo, E., Jones, A.M., Broderick, P., Kemp, Z., Spain, S.L., Guarino, E., Salguero, I., et al. (2013). Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nat. Genet. *45*, 136–144.

Pan, Y., Shen, N., Jung-Klawitter, S., Betzen, C., Hoffmann, G.F., Hoheisel, J.D., and Blau, N. (2016). CRISPR RNA-guided FokI nucleases repair a PAH variant in a phenylketonuria model. Sci. Rep. *6*, 35794–35794.

Panico, K., and Forti, F.L. (2013). Proteomic, Cellular, and Network Analyses Reveal New DUSP3 Interactions with Nucleolar Proteins in HeLa Cells. J. Proteome Res. *12*, 5851–5866.

Panina, Y., Germond, A., Masui, S., and Watanabe, T.M. (2018). Validation of Common Housekeeping Genes as Reference for qPCR Gene Expression Analysis During iPS Reprogramming Process. Sci. Rep. *8*, 8716.

Paquet, D., Kwart, D., Chen, A., Sproul, A., Jacob, S., Teo, S., Olsen, K.M., Gregg, A., Noggle, S., and Tessier-Lavigne, M. (2016). Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. Nature *533*, 125–129.

Park, J.S., Helble, J.D., Lazarus, J.E., Yang, G., Blondel, C.J., Doench, J.G., Starnbach, M.N., and Waldor, M.K. (2019). A FACS-Based Genome-wide CRISPR Screen Reveals a Requirement for COPI in Chlamydia trachomatis Invasion. IScience *11*, 71–84.

Parker, M.W., Botchan, M.R., and Berger, J.M. (2017). Mechanisms and regulation of DNA replication initiation in eukaryotes. Crit. Rev. Biochem. Mol. Biol. *52*, 107–144.

de la Paz Sanchez, M., and Gutierrez, C. (2009). Arabidopsis ORC1 is a PHD-containing H3K4me3 effector that regulates transcription. Proc. Natl. Acad. Sci. *106*, 2065.

Pellegrini, L. (2012). The Pol α-Primase Complex. In The Eukaryotic Replisome: A Guide to Protein Structure and Function, S. MacNeill, ed. (Dordrecht: Springer Netherlands), pp. 157–169.

Peng, R., Lin, G., and Li, J. (2016). Potential pitfalls of CRISPR/Cas9-mediated genome editing. FEBS J. *283*, 1218–1231.

Pervouchine, D.D. (2018). Towards Long-Range RNA Structure Prediction in Eukaryotic Genes. Genes *9*.

Petry, S. (2016). Mechanisms of Mitotic Spindle Assembly. Annu. Rev. Biochem. *85*, 659–683.

Piatti, S., Lengauer, C., and Nasmyth, K. (1995). Cdc6 is an unstable protein whose de novo synthesis in G1 is important for the onset of S phase and for preventing a “reductional” anaphase in the budding yeast Saccharomyces cerevisiae. EMBO J. *14*, 3788–3799.

Pinder, J., Salsman, J., and Dellaire, G. (2015). Nuclear domain ‘knock-in’ screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing. Nucleic Acids Res. *43*, 9379–9392.

Pozo, P.N., and Cook, J.G. (2016). Regulation and Function of Cdt1; A Key Factor in Cell Proliferation and Genome Stability. Genes *8*, 2.

Qiu, H., Honey, D.M., Kingsbury, J.S., Park, A., Boudanova, E., Wei, R.R., Pan, C.Q., and Edmunds, T. (2015). Impact of cysteine variants on the structure, activity, and stability of recombinant human α-galactosidase A. Protein Sci. Publ. Protein Soc. *24*, 1401–1411.

Quan, Y., Xia, Y., Liu, L., Cui, J., Li, Z., Cao, Q., Chen, X.S., Campbell, J.L., and Lou, H. (2015). Cell-Cycle-Regulated Interaction between Mcm10 and Double Hexameric Mcm2-7 Is Required for Helicase Splitting and Activation during S Phase. Cell Rep. *13*, 2576–2586.

Quinet, A., Carvajal Maldonado, D., Lemacon, D., and Vindigni, A. (2017). DNA Fiber Analysis: Mind the Gap! In Methods in Enzymology, p.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013a). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.

Ran, F.A., Hsu, P.D., Lin, C.-Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., et al. (2013b). Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Cell *155*, 479–480.

Rauch, A., Thiel, C.T., Schindler, D., Wick, U., Crow, Y.J., Ekici, A.B., van Essen, A.J., Goecke, T.O., Al-Gazali, L., Chrzanowska, K.H., et al. (2008). Mutations in the Pericentrin (PCNT) Gene Cause Primordial Dwarfism. Science *319*, 816.

Remus, D., Beuron, F., Tolun, G., Griffith, J.D., Morris, E.P., and Diffley, J.F.X. (2009). Concerted Loading of Mcm2–7 Double Hexamers around DNA during DNA Replication Origin Licensing. Cell *139*, 719–730.

Reynolds, J.J., Bicknell, L.S., Carroll, P., Higgs, M.R., Shaheen, R., Murray, J.E., Papadopoulos, D.K., Leitch, A., Murina, O., Tarnauskaitė, Ž., et al. (2017). Mutations in DONSON disrupt replication fork stability and cause microcephalic dwarfism. Nat. Genet. *49*, 537.

Richardson, H.E. (2006). Vinegar Flies Turn to Porto for Cell Division Cycle Meeting. Dev. Cell *11*, 141–146.

Richardson, C.D., Ray, G.J., DeWitt, M.A., Curie, G.L., and Corn, J.E. (2016). Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nat. Biotechnol. *34*, 339–344.

Ricke, R.M., and Bielinsky, A.-K. (2004). Mcm10 Regulates the Stability and Chromatin Association of DNA Polymerase-α. Mol. Cell *16*, 173–185.

Riera, A., and Speck, C. (2015). Opening the gate to DNA replication. Cell Cycle Georget. Tex *14*, 6–8.

Robert, F., Barbeau, M., Éthier, S., Dostie, J., and Pelletier, J. (2015). Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. Genome Med. *7*, 93–93.

Royba, E., Miyamoto, T., Natsuko Akutsu, S., Hosoba, K., Tauchi, H., Kudo, Y., Tashiro, S., Yamamoto, T., and Matsuura, S. (2017). Evaluation of ATM heterozygous mutations underlying individual differences in radiosensitivity using genome editing in human cultured cells. Sci. Rep. *7*, 5996.

Rozov, S.M., Permyakova, N.V., and Deineko, E.V. (2019). The Problem of the Low Rates of CRISPR/Cas9-Mediated Knock-ins in Plants: Approaches and Solutions. Int. J. Mol. Sci. *20*, 3371.

Ryu, S.-M., Koo, T., Kim, K., Lim, K., Baek, G., Kim, S.-T., Kim, H.S., Kim, D., Lee, H., Chung, E., et al. (2018). Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. Nat. Biotechnol. *36*, 536–539.

Ryu, S.-M., Hur, J.W., and Kim, K. (2019). Evolution of CRISPR towards accurate and efficient mammal genome engineering. BMB Rep. *52*, 475–481.

Saha, P., Chen, J., Thome, K.C., Lawlis, S.J., Hou, Z., Hendricks, M., Parvin, J.D., and Dutta, A. (1998). Human CDC6/Cdc18 Associates with Orc1 and Cyclin-cdk and Is Selectively Eliminated from the Nucleus at the Onset of S Phase. Mol. Cell. Biol. *18*, 2758.

Sancar, A., Lindsey-Boltz, L.A., Ünsal-Kaçmaz, K., and Linn, S. (2004). Molecular Mechanisms of Mammalian DNA Repair and the DNA Damage Checkpoints. Annu. Rev. Biochem. *73*, 39–85.

Sangrithi, M.N., Bernal, J.A., Madine, M., Philpott, A., Lee, J., Dunphy, W.G., and Venkitaraman, A.R. (2005). Initiation of DNA Replication Requires the RECQL4 Protein Mutated in Rothmund-Thomson Syndrome. Cell *121*, 887–898.

Sansam, C.L., Shepard, J.L., Lai, K., Ianari, A., Danielian, P.S., Amsterdam, A., Hopkins, N., and Lees, J.A. (2006). DTL/CDT2 is essential for both CDT1 regulation and the early G2/M checkpoint. Genes Dev. *20*, 3117–3129.

Sargent, R.G., Brenneman, M.A., and Wilson, J.H. (1997). Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination. Mol. Cell. Biol. *17*, 267–277.

Satomura, A., Nishioka, R., Mori, H., Sato, K., Kuroda, K., and Ueda, M. (2017). Precise genome-wide base editing by the CRISPR Nickase system in yeast. Sci. Rep. *7*, 2095.

Savic, N., Ringnalda, F.C., Lindsay, H., Berk, C., Bargsten, K., Li, Y., Neri, D., Robinson, M.D., Ciaudo, C., Hall, J., et al. (2018). Covalent linkage of the DNA repair template to the CRISPR-Cas9 nuclease enhances homology-directed repair. ELife *7*, e33761.

Schiroli, G., Conti, A., Ferrari, S., della Volpe, L., Jacob, A., Albano, L., Beretta, S., Calabria, A., Vavassori, V., Gasparini, P., et al. (2019). Precise Gene Editing Preserves Hematopoietic Stem Cell Function following Transient p53-Mediated DNA Damage Response. Cell Stem Cell *24*, 551-565.e8.

Schulz, S., Mensah, M.A., de Vries, H., Fröber, R., Romeike, B., Schneider, U., Borte, S., Schindler, D., and Kentouche, K. (2018). Microcephaly, short stature, and limb abnormality disorder due to novel autosomal biallelic DONSON mutations in two German siblings. Eur. J. Hum. Genet. *26*, 1282–1287.

Shaikh, T.H., Gottlieb, S., Sellinger, B., Chen, F., Roe, B.A., Oakey, R.J., Emanuel, B.S., and Budarf, M.L. (1999). Characterization of CDC45L: a gene in the 22q11.2 deletion region expressed during murine and human development. Mamm. Genome *10*, 322–326.

Shalev, S.A., and Hall, J.G. (2003). Another adult with Meier-Gorlin syndrome - insights into the natural history. Clin. Dysmorphol. *12*.

Shawky, R.M., Elabd, H.S.A.-E., Gamal, R., Mohammad, S.A., and Gad, S. (2014). Meier-Gorlin syndrome: Report of an additional patient with congenital heart disease. Egypt. J. Med. Hum. Genet. *15*, 393–398.

Sheu, Y.-J., and Stillman, B. (2006). Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression. Mol. Cell *24*, 101–113.

Shimada, M., Tsuchiya, H., and Matsumoto, Y. (2017). Effect of different dose rates of ionizing radiation on ciliogenesis in hTERT-RPE1 cells. Spec. Issue Fifth Int. Symp. Innov. Nucl. Energy Syst. *131*, 444–447.

Siitonen, H.A., Kopra, O., Kääriäinen, H., Haravuori, H., Winter, R.M., Säämänen, A.-M., Peltonen, L., and Kestilä, M. (2003). Molecular defect of RAPADILINO syndrome expands the phenotype spectrum of RECQL diseases. Hum. Mol. Genet. *12*, 2837–2844.

Siitonen, H.A., Sotkasiira, J., Biervliet, M., Benmansour, A., Capri, Y., Cormier-Daire, V., Crandall, B., Hannula-Jouppi, K., Hennekam, R., Herzog, D., et al. (2009). The mutation spectrum in RECQL4 diseases. Eur. J. Hum. Genet. *17*, 151–158.

Singh, P., Schimenti, J.C., and Bolcun-Filas, E. (2015). A Mouse Geneticist's Practical Guide to CRISPR Applications. Genetics *199*, 1.

Sokolova, M., Turunen, M., Mortusewicz, O., Kivioja, T., Herr, P., Vähärautio, A., Björklund, M., Taipale, M., Helleday, T., and Taipale, J. (2017). Genome-wide screen of cell-cycle regulators in normal and tumor cells identifies a differential response to nucleosome depletion. Cell Cycle Georget. Tex *16*, 189–199.

Sone, K., Piao, L., Nakakido, M., Ueda, K., Jenuwein, T., Nakamura, Y., and Hamamoto, R. (2014). Critical role of lysine 134 methylation on histone H2AX for γ-H2AX production and DNA repair. Nat. Commun. *5*, 5691–5691.

Song, J., Yang, D., Xu, J., Zhu, T., Chen, Y.E., and Zhang, J. (2016). RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. Nat. Commun. *7*, 10548.

Spurdle, A.B., Couch, F.J., Hogervorst, F.B.L., Radice, P., Sinilnikova, O.M., and IARC Unclassified Genetic Variants Working Group (2008). Prediction and assessment of splicing alterations: implications for clinical testing. Hum. Mutat. *29*, 1304–1313.

Srivastava, M., Nambiar, M., Sharma, S., Karki, S.S., Goldsmith, G., Hegde, M., Kumar, S., Pandey, M., Singh, R.K., Ray, P., et al. (2012). An Inhibitor of Nonhomologous End-Joining Abrogates Double-Strand Break Repair and Impedes Cancer Progression. Cell *151*, 1474–1487.

Stiess, M., Maghelli, N., Kapitein, L.C., Gomis-Rüth, S., Wilsch-Bräuninger, M., Hoogenraad, C.C., Tolić-Nørrelykke, I.M., and Bradke, F. (2010). Axon Extension Occurs Independently of Centrosomal Microtubule Nucleation. Science *327*, 704.

Stiff, T., Alagoz, M., Alcantara, D., Outwin, E., Brunner, H.G., Bongers, M.H., and Jeggo, P.A. (2013). Deficiency in Origin Licensing Proteins Impairs Cilia Formation: Implications for the Aetiology of Meier-Gorlin Syndrome. PLoS Genet. *9*.

Stillman, B. (2001). Genomic Views of Genome Duplication. Science *294*, 2301.

Sun, X., Bizhanova, A., Matheson, T.D., Yu, J., Zhu, L.J., and Kaufman, P.D. (2017). Ki-67 Contributes to Normal Cell Cycle Progression and Inactive X Heterochromatin in p21 Checkpoint-Proficient Human Cells. Mol. Cell. Biol. *37*, e00569-16.

Symington, L.S., and Gautier, J. (2011). Double-Strand Break End Resection and Repair Pathway Choice. Annu. Rev. Genet. *45*, 247–271.

Takara, T.J., and Bell, S.P. (2011). Multiple Cdt1 molecules act at each origin to load replication-competent Mcm2-7 helicases. EMBO J. *30*, 4885–4896.

Tanaka, H., Katou, Y., Yagura, M., Saitoh, K., Itoh, T., Araki, H., Bando, M., and Shirahige, K. (2009). Ctf4 coordinates the progression of helicase and DNA polymerase α. Genes Cells *14*, 807–820.

Tanaka, S., Umemori, T., Hirai, K., Muramatsu, S., Kamimura, Y., and Araki, H. (2007). CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. Nature *445*, 328–332.

Técher, H., Koundrioukoff, S., Azar, D., Wilhelm, T., Carignon, S., Brison, O., Debatisse, M., and Le Tallec, B. (2013). Replication Dynamics: Biases and Robustness of DNA Fiber Analysis. Replicon Theory Regul. DNA Replication Genomic Instab. *425*, 4845–4855.

Thiffault, I., Saunders, C., Jenkins, J., Raje, N., Canty, K., Sharma, M., Grote, L., Welsh, H.I., Farrow, E., Twist, G., et al. (2015). A patient with polymerase E1 deficiency (POLE1): clinical features and overlap with DNA breakage/instability syndromes. BMC Med. Genet. *16*, 31.

Ticau, S., Friedman, L.J., Champasa, K., Corrêa Jr, I.R., Gelles, J., and Bell, S.P. (2017). Mechanism and timing of Mcm2–7 ring closure during DNA replication origin licensing. Nat. Struct. Amp Mol. Biol. *24*, 309.

Ting, C.Y., Bhatia, N.S., Lim, J.Y., Goh, C.-Y.J., Vasanwala, R.F., Ong, C.C.-P., Seow, W.T., Yeow, V.K.-L., Ting, T.W., Ng, I.S.-L., et al. (2019). Further delineation of CDC45-related Meier-Gorlin syndrome with craniosynostosis and review of literature. Eur. J. Med. Genet. 103652.

Truong, L.N., and Wu, X. (2011). Prevention of DNA re-replication in eukaryotic cells. J. Mol. Cell Biol. *3*, 13–22.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3—new capabilities and interfaces. Nucleic Acids Res. *40*, e115–e115.

Uzcanga, G., Lara, E., Gutiérrez, F., Beaty, D., Beske, T., Teran, R., Navarro, J.-C., Pasero, P., Benítez, W., and Poveda, A. (2017). Nuclear DNA replication and repair in parasites of the genus Leishmania: Exploiting differences to develop innovative therapeutic approaches. Crit. Rev. Microbiol. *43*, 156–177.

Van Maldergem, L., Siitonen, H.A., Jalkh, N., Chouery, E., De Roy, M., Delague, V., Muenke, M., Jabs, E.W., Cai, J., Wang, L.L., et al. (2006). Revisiting the craniosynostosis-radial ray hypoplasia association: Baller-Gerold syndrome caused by mutations in the RECQL4 gene. J. Med. Genet. *43*, 148.

Vargas, F.R., de Almeida, J.C.C., Llerena Jr., J.C., and Reis, D.F. (1992). Rapadilino syndrome. Am. J. Med. Genet. *44*, 716–719.

Varma, D., Chandrasekaran, S., Sundin, L.J.R., Reidy, K.T., Wan, X., Chasse, D.A.D., Nevis, K.R., DeLuca, J.G., Salmon, E.D., and Cook, J.G. (2012). Recruitment of the human Cdt1 replication licensing protein by the loop domain of Hec1 is required for stable kinetochore–microtubule attachment. Nat. Cell Biol. *14*, 593–603.

Vesela, E., Chroma, K., Turi, Z., and Mistrik, M. (2017). Common Chemical Inductors of Replication Stress: Focus on Cell-Based Studies. Biomolecules *7*, 19.

Vetro, A., Savasta, S., Russo Raucci, A., Cerqua, C., Sartori, G., Limongelli, I., Forlino, A., Maruelli, S., Perucca, P., Vergani, D., et al. (2017). MCM5: a new actor in the link between DNA replication and Meier-Gorlin syndrome. Eur. J. Hum. Genet. *25*, 646.

Wang, B., Feng, L., Hu, Y., Huang, S.H., Reynolds, C.P., Wu, L., and Jong, A.Y. (1999). The Essential Role of Saccharomyces cerevisiae CDC6 Nucleotide-binding Site in Cell Growth, DNA Synthesis, and Orc1 Association. J. Biol. Chem. *274*, 8291–8298.

Wang, L.L., Gannavarapu, A., Kozinetz, C.A., Levy, M.L., Lewis, R.A., Chintagumpala, M.M., Ruiz-Maldanado, R., Contreras-Ruiz, J., Cunniff, C., Erickson, R.P., et al. (2003). Association

Between Osteosarcoma and Deleterious Mutations in the RECQL4 Gene in Rothmund–Thomson Syndrome. JNCI J. Natl. Cancer Inst. *95*, 669–674.

Watase, G., Takisawa, H., and Kanemaki, M.T. (2012). Mcm10 Plays a Role in Functioning of the Eukaryotic Replicative DNA Helicase, Cdc45-Mcm-GINS. Curr. Biol. *22*, 343–349.

Wolter, F.P., Fritz, C.C., Willmitzer, L., Schell, J., and Schreier, P.H. (1988). rbcS genes in Solanum tuberosum: conservation of transit peptide and exon shuffling during evolution. Proc. Natl. Acad. Sci. *85*, 846.

Wong, P.G., Winter, S.L., Zaika, E., Cao, T.V., Oguz, U., Koomen, J.M., Hamlin, J.L., and Alexandrow, M.G. (2011). Cdc45 Limits Replicon Usage from a Low Density of preRCs in Mammalian Cells. PLOS ONE *6*, e17533.

Woodward, A.M., Göhler, T., Luciani, M.G., Oehlmann, M., Ge, X., Gartner, A., Jackson, D.A., and Blow, J.J. (2006). Excess Mcm2-7 license dormant origins of replication that can be used under conditions of replicative stress. J. Cell Biol. *173*, 673–683.

Wynn, S.L., Fisher, R.A., Pagel, C., Price, M., Liu, Q.Y., Khan, I.M., Zammit, P., Dadrah, K., Mazrani, W., Kessling, A., et al. (2000). Organization and Conservation of the GART/SON/DONSON Locus in Mouse and Human Genomes. Genomics *68*, 57–62.

Xu, X., Gao, D., Wang, P., Chen, J., Ruan, J., Xu, J., and Xia, X. (2018). Efficient homology-directed gene editing by CRISPR/Cas9 in human stem and primary cells using tube electroporation. Sci. Rep. *8*, 11649.

Yang, L., Guell, M., Byrne, S., Yang, J.L., De Los Angeles, A., Mali, P., Aach, J., Kim-Kiselak, C., Briggs, A.W., Rios, X., et al. (2013). Optimization of scarless human stem cell genome editing. Nucleic Acids Res. *41*, 9049–9061.

Yao, X., Wang, X., Hu, X., Liu, Z., Liu, J., Zhou, H., Shen, X., Wei, Y., Huang, Z., Ying, W., et al. (2017). Homology-mediated end joining-based targeted integration using CRISPR/Cas9. Cell Res. *27*, 801.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics *13*, 134–134.

Yeeles, J.T.P., Deegan, T.D., Janska, A., Early, A., and Diffley, J.F.X. (2015). Regulated eukaryotic DNA replication origin firing with purified proteins. Nature *519*, 431–435.

Yiu, G., Tieu, E., Nguyen, A.T., Wong, B., and Smit-McBride, Z. (2016). Genomic Disruption of VEGF-A Expression in Human Retinal Pigment Epithelial Cells Using CRISPR-Cas9 Endonuclease. Invest. Ophthalmol. Vis. Sci. *57*, 5490–5497.

Yoshida, K., Oyaizu, N., Dutta, A., and Inoue, I. (2004). The destruction box of human Geminin is critical for proliferation and tumor growth in human colon cancer cells. Oncogene *23*, 58.

You, Z., Komamura, Y., and Ishimi, Y. (1999). Biochemical analysis of the intrinsic Mcm4-Mcm6-mcm7 DNA helicase activity. Mol. Cell. Biol. *19*, 8003–8015.

Yu, N., Yang, J., Mishina, Y., and Giannobile, W.V. (2018). Genome Editing: A New Horizon for Oral and Craniofacial Research. J. Dent. Res. *98*, 36–45.

Zegerman, P., and Diffley, J.F.X. (2007). Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. Nature *445*, 281–285.

Zeman, M.K., and Cimprich, K.A. (2014). Causes and consequences of replication stress. Nat. Cell Biol. *16*, 2–9.

Zhang, Y.-Q., and Rao, R. (2007). Global Disruption of Cell Cycle Progression and Nutrient Response by the Antifungal Agent Amiodarone. J. Biol. Chem. *282*, 37844–37853.

Zhang, J.-P., Li, X.-L., Li, G.-H., Chen, W., Arakaki, C., Botimer, G.D., Baylink, D., Zhang, L., Wen, W., Fu, Y.-W., et al. (2017). Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. Genome Biol. *18*, 35–35.

Zhang, W., Sankaran, S., Gozani, O., and Song, J. (2015). A Meier-Gorlin Syndrome Mutation Impairs the ORC1-Nucleosome Association. ACS Chem. Biol. *10*, 1176–1180.

Zhu, W., Ukomadu, C., Jha, S., Senga, T., Dhar, S.K., Wohlschlegel, J.A., Nutt, L.K., Kornbluth, S., and Dutta, A. (2007). Mcm10 and And-1/CTF4 recruit DNA polymerase alpha to chromatin for initiation of DNA replication. Genes Dev. *21*, 2288–2299.

Zou, L., and Stillman, B. (2000). Assembly of a Complex Containing Cdc45p, Replication Protein A, and Mcm2p at Replication Origins Controlled by S-Phase Cyclin-Dependent Kinases and Cdc7p-Dbf4p Kinase. Mol. Cell. Biol. *20*, 3086.

Zou, L., Mitchell, J., and Stillman, B. (1997). CDC45, a novel yeast gene that functions with the origin recognition complex and Mcm proteins in initiation of DNA replication. Mol. Cell. Biol. *17*, 553.

Appendix A

General reagents and primers

A.1 General reagents

- 0.25% Trypsin-EDTA (Thermo Scientific - 25200-072)
- 0.5 M EDTA pH8 (Thermo Scientific - 180-500G)
- 10% APS (Sigma-Aldrich - A3678-25G)
- 10% SDS (Sigma-Aldrich - 436143-100G)
- 100 bp DNA ladder (Genesearch - N3231L)
- 10x Buffer G (Thermo Scientific - BG6)
- 10x FastAP Buffer (Thermo Scientific - EF0651)
- 1 kb DNA ladder (Genesearch N3232S)
- 1 M CaCl2 (Thermo Scientific - AJA127-500G)
- 1 M NaCl (LabServ - BSPSL944.500)
- 1 M NaOH (Sigma-Aldrich - S5881-500G)
- 1x cOmplete™ Mini, EDTA-free Protease Inhibitor Cocktail mix (Sigma-Aldrich – 4693159001)
- 2x ReddyMix PCR Master Mix (Thermo Scientific - ABGAB-0575/DC/LD/A)
- 2.1 buffer (NEB - B7202S)
- 2x Phusion Flash PCR Master Mix (Thermo Scientific - F548S)
- 3 M Sodium acetate (Thermo Scientific - R1181)
- 50% Glycerol (Thermo Scientific - BSPGL885-500ML)
- 5x big dye buffer (GAS Otago)
- 70% ethanol (Pathology Store)
- Acetic acid (BDH – Lot K286717101)
- Agarose (MediRay - FBAG100)
- Amphotericin B aka Fungizone (Thermo Scientific – 15290018)
- Ampicillin (Sigma-Aldrich - A5354-10ML)
- Anti-mouse 488: Alexa Fluor (Thermo Scientific - A-11001)
- Anti-phospho-Histone H2A.X (Ser139) (merk millipore 05-636)
- Anti-Rat 594: Alexa Fluor (Thermo Scientific - A-11007)
- Bbsl Restriction enzyme (Thermo Scientific - ER1011)
- Big Dye Master mix (GAS Otago)
- Blocking solution - (7% FBS, 0.1% Triton X-100 in PBS)
- BSA (Sigma-Aldrich - A3733-50G)
- BsrGI restriction enzyme (Genesearch - R0575S)
- CDC45 antibody (Bio-Strategy - sc-55569)
- Chemically competent *E.coli* Top10
- CldU: (Sigma-Aldrich - C6891-100 mg)
- Cutsmart buffer (NEB - B7204S)
- DAPI (Sigma-Aldrich - D9542-1MG)
- DMEM (Thermo Scientific - 11995.065)

- DMEM/F12 (Thermo Scientific – 11330032)
- DMSO (Sigma-Aldrich - D8418-50ML)
- DNeasy Blood & Tissue Kit (Bio-Strategy - 69504)
- DpnI restriction enzyme (Genesearch - R0129S)
- DreamTaq Green PCR Master Mix (2X) (Thermo Scientific - K1082)
- Ethanol (Pathology Store)
- Exonuclease (Thermo Scientific - EN0582)
- FastAP alkaline phosphatase (1 u/µL) (Thermo Scientific - EF0651)
- Fetal bovine serum (Morgate Biotech - Batch #48827103)
- GeneJET Gel Extraction Kit (Thermo Scientific - K0691)
- Gibson Assembly® Master Mix (Genesearch - E2611S)
- Glycerol (Thermo Scientific - BSPGL885-500ML)
- HCl: 30% (Lab Supply - MER103181000)
- Hygromycin B (Sigma-Aldrich – 10843555001)
- IdU: SIGMA (I7125-5 g) (Abcam - ab142581)
- IRDye 800CW goat anti-mouse secondary antibody diluted 1:25,000 (Millenium Science NZ)
- Isopropanol (BDH - 102246L)
- Kanamycin (Sigma-Aldrich K0254-20ML)
- LB Agar (Oxoid - CM0131)
- Loading Dye Purple (6X) no SDS (NEB - B7025S)
- Lipofectomine 2000 (Thermo Scientific - 11668-027)
- Lipofectomine 3000 (Thermo Scientific - L3000001)
- Methanol (Fisher Scientific - M/4000/17)
- GeneJET Midi Plasmid prep kit (Thermo Scientific - K0481)
- GeneJET Mini Plasmid prep kit (Thermo Scientific - K0502)
- Mounting medium: SIGMA (F6182-20 ml)
- Mouse anti-BrdU: Becton Dickinson (347580)
- Mouse anti-CDC45 (Bio-Strategy cat. no. sc-55569)
- Opti-MEM reduced serum media (Thermo Scientific - 31985062)
- Paraformaldehyde (Sigma-Aldrich - P6148-500G)
- PCS2+Cas9-mSA (Addgene plasmid #103882)
- Penicillin/Streptomycin (Thermo Scientific - 15140-122)
- Pierce™ Rapid Gold BCA Protein Assay Kit (Thermo Scientific - A53226)
- PowerUp™ SYBR™ Green Master Mix (Thermo Scientific - A25741)
- Precision Plus Protein Dual Colour Standard (Bio-Rad – 1610374)
- ProLong™ Diamond Antifade Mountant (Thermo Scientific - P36965)
- pSpCas9(BB)-2A-GFP (PX458) (Addgene plasmid #48138)
- Qubit™ dsDNA BR Assay Kit (Thermo Scientific - Q32850)
- QuickExtract™ DNA Extraction Solution (Mediray - QE0905T)
- Rat anti-BrdU: (AbD SeroTec (OBT0030G) 0.25 mg)
- RIPA buffer (Thermo Scientific - 89900)
- RNeasy Mini Kit (Bio-Strategy - 74104)
- SacI restriction enzyme (Genesearch - R3156S)
- Sodium acetate solution (Thermo Scientific - R1181)

- Sodium chloride (NaCl) (LabServ - BSPSL944.500)
- SuperScript™ IV VILO™ Master Mix with ezDNase enzyme (Thermo Scientific - 11766050)
- SYBRSafe (Thermo Scientific - S33102)
- 10X T4 ligation buffer (NEB - B0202S)
- T4 PNK (10 u/µL) (NEB - M0201S)
- Taq (α) 1 restriction enzyme (Genesearch - R0149S)
- TEMED (Sigma-Aldrich - T9281-25ML)
- Tris UltraPure (Invitrogen 15504020)
- Triton X-100 (Sigma-Aldrich - T8787-50ML)
- Tryptone (Oxoid - LP0042)
- Tween 20 (Sigma-Aldrich - P1379)
- Yeast extract (Oxoid - LP0021)
- Zombie NIR™ Fixable Viability Kit (MediRay BIO423105)

A.2 General buffers

1x Phosphate Buffered Saline (PBS)

- Add one Oxoid Phosphate Buffered Saline (Dulbecco A) tablet per 100 mL of distilled water.
- Mix until dissolved and autoclave at 121 °C for 20 minutes + to sterilise.

LB media

- Dissolve 10 g Tryptone, 10 g NaCl, & 5 g Yeast extract in 1 litre of distilled water.
- Add no more than 400 mL LB per 500 mL Schott bottle and autoclave at 121 °C for at least 20 minutes to sterilise.

MACS buffer

- Autoclave PBS and EDTA solutions
- In a clean hood make a 1X PBS pH 7.2, 2 mM EDTA solution
- Add FBS to a final concentration of 1% and filter sterilise (0.2 µm) MACS buffer into sterile 50 mL Falcon tubes
- Store at 4 °C.

Tris-HCl

- 0.5 M, 1 M pH 7.4, 1.5 M pH7.5 or 2 M pH5 Tris-HCl solutions were made by dissolving the appropriate amount of UltraPure Tris in MilliQ water before adjusting to the desired pH using HCl.

Tris/Borate/EDTA (TBE) solution (2.5x stock)

- Dissolve 26 g Tris and 12.75 g Boric acid in ~900 mL distilled water.

- Once dissolved add 10 mL 0.5 M EDTA pH8 and make up solution to 1 litre.
- Dilute 2.5x TBE stock to 0.5x TBE before use.

10x Western Buffer

- Dissolve 144 g Glycine and 30 g Tris in a total volume of 1 litre.

A.3 Primer list

CRISPR	
ORC1 sgRNA_F	caccggCCACAGAAGGTCGAGctgcc
ORC1 sgRNA_R	aaacggcagCTCGACCTTCTGTGGcc
ORC1 pL622P ssODN	gaaacatgcagctaaagcagtggttataggtcttcattttgaagcaccagctcagtctctgtgatt gactctgcacctctctcctttgcTctggcagCCCGACCTTCTGTGGACTCACAAACA AGACATAATGTACAATCTCTTTGACTGGCCCACTCATAAGGAGGCCCG GCTTGTGGTCCTGGCAA
ORC1 CRISPR seq_F	aaagtcctggctcccttctc
ORC1 CRISPR seq_R	aaagtgcggggtttacagg
DONSON sgRNA_F	caccggGGCTCTGCTGGAAGGTACAA
DONSON sgRNA_R	aaacTTGTACCTTCCAGCAGAGCCcc
DONSON pR211C ssODN	gtctttaataaaattctctatttctcaatgggcaaaaactgcccactgtcacctttggattttagGA TCCCAAACTCTCCTCTGAGCTATGTTGTACCTTCCAGCAGAGCCTTATC TATTGGCTCCACCCTGCTTTGTCTTGGCTACCACTGTTCCCTCGTATTG GAGCTGATAGAAAA
DONSON CRISPR Seq_F	cggcttaccattacatctgg
DONSON CRISPR Seq_R	tgccattaacgtatccagtcac
AS-PCR	
AS-PCR_*DONSON*_WT_F	CCAAACTCTCCTCTGAGCTCC
AS-PCR_*DONSON*_M_F	CACCTTTGGATTTTAGGATCCCAAACTCTCCTCTGAGCTAT
AS-PCR_*DONSON*_R	ACATGCTGCAGGGTTGCATC
*ORC1*_pL622P_ssODN_biotin	/5Biosg/GAAACATGCAGCTAAAGCAGTGGTTATAGGTCTTCATTTTG AAGCACCAGCTCAGTCTCTGTGATTGACTCTGCACCTCTCTCCTTTGCT CTGGCAGCCCGACCTTCTGTGGACTCACA AACAAGACATAATGTACAATCTCTTTGACTGGCCCACTCATAAGGAGG CCCGGCTTGTGGTCCTGGCAA
*DONSON*_pR211C_ssODN_biotin	/5Biosg/GTCTTTAATAAAATTCTCTATTTCTCAATGGGCAAAAACTG CCCACTGTCACCTTTGGATTTTAGGATCCCAAACTCTCCTCTGAGCTAT GTTGTACCTTCCAGCAGAGCCTTATCTATTGGCTCCACCCTGCTTTGTC TTGGCTACCACTGTTCCCTCGTATTGGAGCTGATAGAAAA
GIBSON	
Gibson_F	ACTCTCGGCATGGACGAGCTGTACAAGGAATTCATGGCGGAAG
Gibson_R	AGCTCTAGTTAGAATTCCTTTCAGGATCCAGACGCCGC
mSA_F	GTACAAGGAATTCATGGCGGAAGCGGGTATCAC
mSA_R	TCAGGATCCAGACGCCGCAGAC
PX458_mSA_seq_F	GGTGAACTTCAAGATCCGCC
PX458_mSA_seq_R	TGGCCAACTCCATCACTAGG
mSA_cPCR_F	CGGGTATCACCGGCACGTGG
mSA_cPCR_R	CTGGCAACTAGAAGGCACAGTC
mSAfrag_cPCR_F	CTGGAATGGCGTGTTGAATGGA
mSAfrag_cPCR_R	TCCATTCAACACGCCATTCCAG

DONSON Mutagenesis	
*DONSON*_pR211C_S	CAAACTCTCCTCTGAGCTCTGTTGTACCTTCCAGCAG
*DONSON*_pR211C_AS	CTGCTGGAAGGTACAACAGAGCTCAGAGGAGAGTTTG
*DONSON*_pY270C_S	GTGAGCTTTACTTCTCTATGTAATTTGCTGAAGACAAAAC
*DONSON*_pY270C_AS	GTTTTGTCTTCAGCAAATTACATAGAGAAGTAAAGCTCAC
*DONSON*_pP545L_S	GAGCAACTTAGTCAAATACTGTTACTTGGGAAATCATC
*DONSON*_pP545L_AS	GATGATTTCCCAAGTAACAGTATTTGACTAAGTTGCTC
*DONSON*_pP224S_S	CTTATCTATTGGCTCCACTCTGCTTTGTCTTGGCTAC
*DONSON*_pP224S_AS	GTAGCCAAGACAAAGCAGAGTGGAGCCAATAGATAAG
*DONSON*_pF165S_S	GTCTATCAAAACGCGACTCCTTTCCACCTCTTCTCAACCC
*DONSON*_pF165S_AS	AAGGGTTGAGAAGAGGTGCTAAGGAGTCGCGTTTTGATAG
***DONSON* Sequencing primers**	
*DONSON*_seq_F1	GTACTGAGTTACCCGTAGATTGGTCTATC
*DONSON*_seq_F2	GACATACTTTCTATCAAGCTGCG
*DONSON*_seq_R1	ACAGATTCAGGTCTGTGATCC
*DONSON*_seq_R2	TGCCCAGGTAAAGGGTTGAG
*DONSON*_seq_F3	TCATGTTAGCTGATAAAATCTTTAGG
*DONSON*_seq_R3	TGCAAAGGAGAGCATTTCAC
*DONSON*_seq_F4	GGCTTACCATTACATCTGGCTC
*DONSON*_seq_R4	CCTTCTGCCATTAACGTATCC
*DONSON*_seq_F5	GCGAGTCCTCATAACCAATTTC
*DONSON*_seq_R5	AAAAGTTTGCGGACTGCTG
*DONSON*_seq_F6	TGTCTCATGCATCTCAGTTGG
*DONSON*_seq_R6	ATTTCCCATTTGCAAGGAAG
EGFP_seq_F1	GGAGTACAACTACAACAGCC
EGFP_seq_R1	TCCGTCGGAGCCGCACTAC
pEGFP_seq_F1	CAGAGCTGGTTTAGTGAACCGTCAG
pEGFP_seq_R1	AGGCACAGTCGAGGCTGATC
M13F	GTAAAACGACGGCCAG
M13R	CAGGAAACAGCTATGAC
RT-PCR	
RT-PCR_F	GCCTCTCCAGGAGTAACCTG
RT-PCR_R	GGACAGGAGGGAAATAAGTG
qPCR	
Fenwick_qPCR_Ex1-3_F	TCCGATTTCCGCAAAGAGTTCTACG
Fenwick_qPCR_Ex1-3_R	CAGCGTATATTGCACGTGGTCACA
Fenwick_qPCR_Ex5-6_F	GTTCCCGCCTATGAAGACAT
Fenwick_qPCR_Ex5-6_R	CTCCTCCGCATGGTTTGCTC
*CDC45*_qPCR_Ex15_F	GCACTCCAGATGTCATGCTG
*CDC45*_qPCR_Ex16_R	GTCACTGTGCCATGCTCCAT
*CDC45*_qPCR_Ex17_F	ATGCTGCACAACCATTTTGA
*CDC45*_qPCR_Ex18_R	CCAGAAACTTGCTCCGATCC
*RPS18*_F	CATTAAGGGTGTGGGCCGAA
*RPS18*_R	CTTCAGTCGCTCCAGGTCTT
CDC45-mutagenesis	
Mut_*CDC45*_del1584-1589_s	gtcctttgtgtgttcgaaccggcgctgcaaac
Mut_*CDC45*_del1584-1589_as	gtttgcagcgccggttcgaacacacaaaggac
Gateway cloning	

attB1_*ORC1*_L622P_F	GGGGACAAGTTTCTACAAAAAAGCAGGCTATTTTAAGGGGGCTTCTGGC
attB2_*ORC1*_L622P_R	GGGGACCACTTTGTACAAGAAAGCTGGGTAACTGCTTGCTCCCTGAATGA
cPCR_R	CTG TAA GGG TCT TAC TCC CTT G
cPCR_R_2	TGC CGC ATG GTT GGC TGT TG
RT-PCR oligonucleotides	
Rat_F	Rat_INS2_Ex2_F CCTGCTCATCCTCTGGGAGC
Rat_R	Rat_INS2_Ex3_R AGGTCTGAAGGTCACGGGCC
*ORC1*_Ex12 _R	TGC CGC ATG GTT GGC TGT TG
*ORC1*_Ex13_R	CTCTGGCAGGTCCATTGTGT
*ORC1*_Ex14_R	CTACCAGCTGGATGGCATCAT

Appendix B

General protocols

B.1 Chemically competent *E.coli* top10 protocol

50 mL of the following solutions were prepared and autoclaved:

1 M $CaCl_2$
1 M Tris-HCL pH 7.4
50% Glycerol

Chemically competent *E.coli* Top10 preparation:

- 10 mL LB was inoculated with approx. 5 µL of an aliquot of Karen's competent *E.coli* Top10 cells.
- The culture was incubated overnight at 37 °C, 200 rpm

Chemically competent *E.coli* Top10 prep:

- 2 mL of the overnight culture was used to inoculate 200 mL LB in a sterile 1 L conical flask.
- The culture was incubated at 37 °C, 250 rpm for 2-4 hours until an OD600 nm of 0.2-0.5 (~5 x 10^7 cells/mL) was reached (measurements were take every 30-60 minutes).
- OD600 nm = 0.047 (NanoDrop measurement)
- The flask was chilled on ice for 10 minutes.
- The culture was transferred to sterile 50 mL falcon tubes and centrifuged (3220 g, 15 min, 4 °C).
- The supernatant was discarded, and the cells were resuspended in 100 mL ice-cold 50 mM $CaCl_2$, 10 mM Tris-HCL pH 7.4.
- Cells were incubated on ice for 15 minutes and then re-centrifuged (3220 g) 15 min, 4 °C).
- The supernatant was discarded, and the cells were resuspended in 5 mL ice-cold 50 mM $CaCl_2$, 10 mM Tris-HCL pH 7.4, 10% Glycerol.
- Cells were aliquoted (100 µL) into small Eppendorf tubes, snap frozen with liquid nitrogen and placed in the -80 °C freezer.

A transformation efficiency of 10^5-10^7 transformants/ 1 µg plasmid is expected.

B.2 SDS PAGE gel

2x 10% 1 mm separating gels:

Separating buffer	– 3 ml
40% bis acrylamide	– 3 mL
10% SDS	– 120 µL
10% APS	– 120 µL
TEMED	– 10 µL
MilliQ water	– 5.75 mL

2x 1 mm stacking gels:

Stacking buffer	– 2.5 ml
acrylamide	– 1 mL
10% SDS	– 100 µL
10% APS	– 100 µL
TEMED	– 10 µL
MilliQ water	– 6.29 mL

B.3 CRISPR FACS protocol

Bulk FACS sort

Cells were visualised under the cell culture room microscope and appeared healthy. Cells were visualised under the Olympus IX71 inverted light microscope with brightfield and GFP (488) filter. Both bright field and GFP images were taken for reference.

- Approx. 10 mL of prewarmed fresh Complete Growth Medium (10% FBS, 0.01 mg/mL Hygromycin B, 1x penicillin/streptomycin in Hyclone DMEM: F12 medium) was added to each well.
- The samples were then passed through sterile 70 µm cell strainers into a 50 mL falcon tubes.
- Cells were harvested by centrifugation (300 g, 5 minutes at room temperature).
- Cell pellets were resuspended in 1 mL sterile MACS buffer.

*Note: MACS buffer (1X PBS pH 7.2, 2 mM EDTA, 1% FBS) was generated in the clean hood from autoclaved PBS and EDTA stocks. The buffer was filter sterilise (0.2 µm) into sterile 50 mL falcon tubes after FBS was added and stored at 4 °C.

- Cells were transferred to 15 mL falcon tubes.
- A 250 µL aliquot of the not transfected hTERT-RPE-1 control sample was taken for heat-kill treatment.
- The heat-kill aliquot was incubated at 80 °C for 2 minutes.
- 250 µL non heat-killed not transfected hTERT-RPE-1 control sample was added to the heat-killed aliquot resulting in a 1:1 ratio of live to dead cells.
- Cells were again harvested by centrifugation (300 g, 5 minutes at room temperature).

- Cell pellets were resuspended in 100 µL sterile MACS buffer.
- 1 µL Zombie dye was added to each of the cell suspensions to generate the following samples:

C1. hTERT-RPE-1 cells – not transfected (unstained).
C2. hTERT-RPE-1 cells – not transfected (with Zombie dye).
C.3 hTERT-RPE-1 cells – GFP positive cells.
1. hTERT-RPE-1 cells – CRISPR Experiment.

- The samples were kept in the dark after the addition of the Zombie dye.
- Samples were incubated at room temperature for 30 minutes.
- The cells were harvested by centrifugation (300 g, 5 minutes at room temperature).
- Cell pellets were resuspended in 0.5 mL sterile MACS buffer and placed on ice.

15 mL labelled falcon tubes containing 5 mL CGM, the samples and spare sterile MACS buffer (on ice), were taken over to the Microbiology building, 4th floor Room 414.

- Michelle Wilson calibrates and tests the drop delay of the FACS Aria instrument.
- Samples were bulk sorted.
- Once transported back to Pathology cells were plated ~2000-4000 cells per 24 well.
- Cell cultures were incubated 37 °C, 5% CO_2.

Single cell FACS sort (1 week later)

- The cells were visualised under the cell culture room microscope and appeared healthy and ~80% confluent.
- Following the removal of media from the 24-well plate (using aseptic technique in the cell culture hood), cells were trypsinized by adding 0.1 mL Trypsin solution (per well) and incubating at 37 °C, 5% CO2 for 10 minutes.
- Two mL of prewarmed fresh Complete Growth Medium (10% FBS, 0.01 mg/mL Hygromycin B, 1x penicillin/streptomycin in DMEM:F12 medium) was added to each well.
- Samples were passed through sterile 70 µm cell strainers into 50 mL falcon tubes.
- Cells were harvested by centrifugation (300 g, 5 minutes at room temperature).
- Cell pellets were resuspended in 1 mL sterile MACS buffer.
- Control cells were split into two 0.5 mL aliquots.
- A 250 µL aliquot of the untransfected control sample was taken for heat-kill treatment.
- The heat-kill aliquot was incubated at 80 °C for 2 minutes.
- This was added to the remaining 250 µL control cells (not heat killed) resulting in a 1:1 ratio of live to dead cells.
- Cells were again harvested by centrifugation (300 g, 5 minutes at room temperature).

- Cell pellets were resuspended in 100 µL sterile MACS buffer.
- Cells were transferred to 15 mL falcon tubes.
- One µL Zombie dye was added to each of the cell suspensions to generate the following samples:

1. hTERT-RPE-1 cells – not transfected (unstained).
2. hTERT-RPE-1 cells – heat killed (with Zombie dye).
3. hTERT-RPE-1 CRISPR sample (with Zombie dye).

- Samples were kept in the dark after the addition of the Zombie dye.
- Samples were incubated at room temperature for 30 minutes.
- All samples were harvested by centrifugation (300 g, 5 minutes at room temperature).
- Cell pellets were resuspended in 0.5 mL sterile MACS buffer and placed on ice.
- 96 well U bottom plates containing 100 µL CGM per well were prepared.
- The samples, 96 well U bottom plates (containing 100 µL CGM) and spare MACS buffer were taken over to the Microbiology building, 4th floor Room 414.
- A single cell was sorted into the inside wells in the 96 well plate (60 cells/plate).
- Plates were transported back to the laboratory on ice.
- Incubate cell cultures at 37 °C, 5% CO_2.

Media change Every 3-4 days

- Replace media in 96-well trays with prewarmed fresh Complete Growth Medium.
- Monitor growth by microscopy, expand clones over 60% confluent after 1.5-2 weeks.

B.4 Immunocytochemistry

Solutions:

PBTx-BSA

PBS	100 mL
Bovine serum albumin (BSA)	0.25 g
Triton X-100	300 µL

Blocking solution

PBTx-BSA	980 µL
Normal serum (species that secondary was made in)	20 µL

- Wash cells 4 x 2 minutes in PBS at room temperature. After removing the final PBS wash, add 500 µL blocking solution & incubate at room temperature for 1 h.
- Dilute the primary antibody in 200 µL blocking solution add to each well.
- Incubate overnight at 4 °C.
- Wash 4 x 2 minutes in PBS.

- Make up fluorescent secondary antibody solution in PBTx-BSA and add ~250 µL room temperature for 1.5 h.
- Wash 4 x 2 minutes in PBS. Leave coverslips in the final rinse. Mount coverslips onto glass slides by inverting onto drops of fluorescent mounting medium. Leave in the dark to cure for a couple of hours.

B.5 DNA fibre combing

Reagents

Spreading buffer: 200 mM Tris pH 7.5, 50 mM EDTA, 0.5% SDS.
HCl: 37% Fisher Scientific.
Blocking solution (200 ml): 2 g BSA, 200 µl Tween 20, made with PBS (non-sterile).
Rat anti-BrdU (1:750): AbD SeroTec (OBT0030G) 0.25 mg.
Mouse anti-BrdU (1:750): Becton Dickinson (347580).
Anti-Rat 594 (1:500): Alexafluor.
Anti-mouse 488 (1:500): Alexafluor.
Mounting medium: SIGMA (F6182-20 ml).
Fixative: Methanol / acetic acid (3:1).
CldU: SIGMA (C6891-100 mg).
IdU: SIGMA (I7125-5 g).

CldU: 6.6 mg in 10 mL medium. Vortex to dissolve. Store at -20 °C.
IdU: 8.9 mg in 10 mL medium. Heat to 60 °C and vortex to dissolve. Store at 4 °C. Prior to use, pour into tissue culture dish (6 cm dish) and equilibrate in CO_2 for ~1 hour.

Protocol

Prepare everything for the experiment so cells are maintained in correct conditions at all times, including ice box for PBS and tubes.

- Grow sub-confluent cells within 6-well plate (minimal size being a 24-well plate) with volume of 2 mL medium.
- Add **25 µM CIdU** to cell samples and incubate cells at 37 °C for 20 minutes.
- Remove CldU and rinse 3x with warm PBS.
- Add **250 µM IdU** to the cells and incubate cells at 37 °C for 20 minutes.
- Remove medium/ IdU solution and wash cells 2 x with PBS (non-sterile).
- Trypsinize cells from the well.
- Re-suspend cells in 1 mL ice-cold PBS and transfer to ice-cold 1.5 mL eppendorf.

Counting cells

Make sure cells are suspended well in the tube before taking a small sample.
Count cells using the MOXI counter.
Optimal concentration: 50 x 10^4 cells/ml.
Calculate concentration and adjust accordingly to optimal.

At this stage, cells can be stored on ice for a maximum of 1-2 hours.

Spreading fibres on slides

- Upon a piece of paper towel, lay out 5 slides per sample (accounting for things going wrong) and number with pencil.
- Make small aliquots of the spreading buffer (200 mM Tris pH 7.5, 50 mM EDTA, 0.5% SDS) for each sample i.e. ~40 µl/tube for each sample.
- Place 2 µl of the sample onto each slide, near to the frosting, and allow to slightly dry, i.e. the edges of the drop become crinkly: ~6 minutes.
- Directly on top of each, place 7 µl of the spreading buffer (using the same tip, taking care not to touch the slide) and then stir the two together on the slide with the tip (~ 5 x).
- Incubate for 2 minutes at RT.
- Tilt the slides briefly to set the drop moving and then leave at a smaller angle (slide sitting on the end of a PCR test tube rack), for drops to slide down the slide gradually, however don't allow the drop to dry.
- Allow excess liquid to run off the end, and leave to air-dry for at least 2 minutes.
- Place slides directly into MeOH/ AcOH (3:1) in the coplin jar for at least 10 minutes (can be left for longer). This fixes the DNA to the slide and de-proteinises the sample.
- Air-dry the slides for 5-10 minutes and store in the fridge for up to 1 year.

Immunostaining

Place the slides in a black tray and keep covered as much as possible.

- Wash (~1 minute) slides 2 x with 1 mL H_2O (pour off liquid into plastic beaker).
- Rinse 1 x with 1 mL **freshly made** 2.5 M HCl.
- Denature with 2.5 M HCl for 1 hour 15 minutes (NO LONGER; optimised for all human cells). This makes the ssDNA accessible to the antibody.
- Rinse 2 x with PBS.
- Wash (~1 minute) 2 x with blocking solution (1 g BSA, 100 µl Tween 20, in 100 mL PBS).
- Incubate in blocking solution for 30 minutes - 1 hour.
- Tip off the Blocking solution and add **115 µl primary antibodies** i.e. **1:750 Rat anti-BrdU** (anti-CIdU) and **1:750 Mouse anti-BrdU** (anti-IdU).
- Cover each slide with a long cover slip and incubate in the dark overnight 4 °C (can leave for 1 hour at room temperature but found overnight gives cleaner images).
- Tip off the antibodies and cover slip into the beaker and rinse 3 x with PBS.
- Fix with 0.5 mL 4% paraformaldehyde for 10 minutes. This fixes the antibody to the antigen.
- Rinse 3 x with PBS.
- Wash (~1 minute) 3 x with blocking solution.
- Add **115 µl secondary antibodies** i.e. **1:500 anti-rat Alexafluor 594** (red) and **1:500 anti-mouse Alexafluor 488** (green).

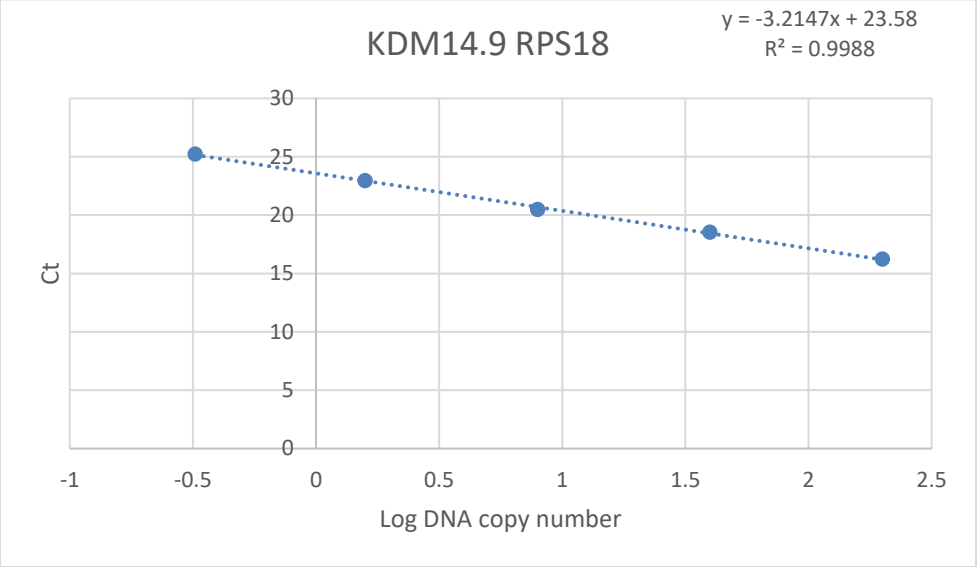
- Cover each slide with a long cover slip and incubate in the dark for 1.5 hours.
- Rinse 2 x with PBS.
- Wash (~1 minute) 3 x with blocking solution.
- Rinse 2 x with PBS.
- Spread a drop of mounting medium over the surface and place a long cover slip over.
- Allow to dry for 5-10 minutes before dabbing dry with a tissue and sealing edges with nail varnish.
- Store in freezer.

Image under Olympus BX53 microscope.

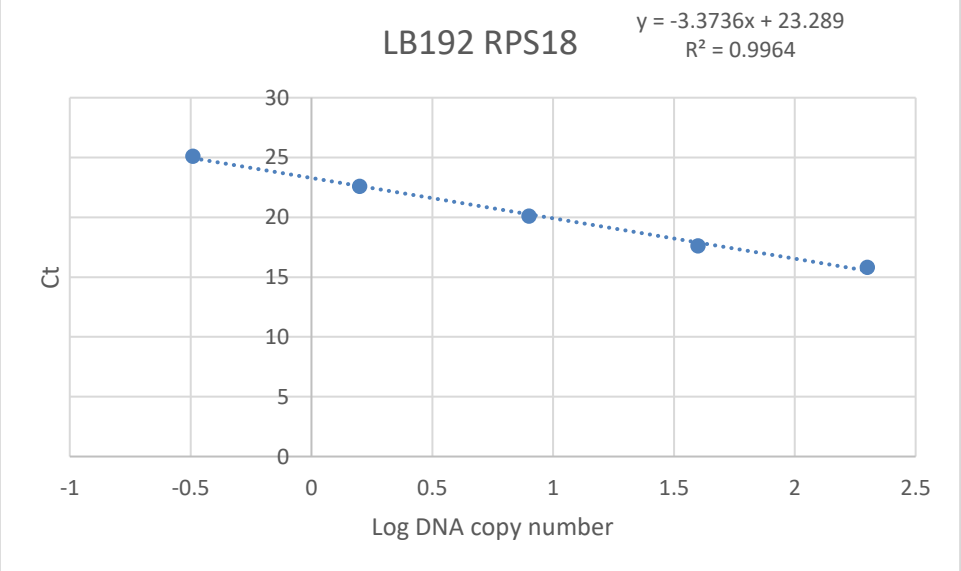
Appendix C CDC45

C.1 qPCR standard curves for calculating primer efficiency

RPS18 – Housekeeping gene



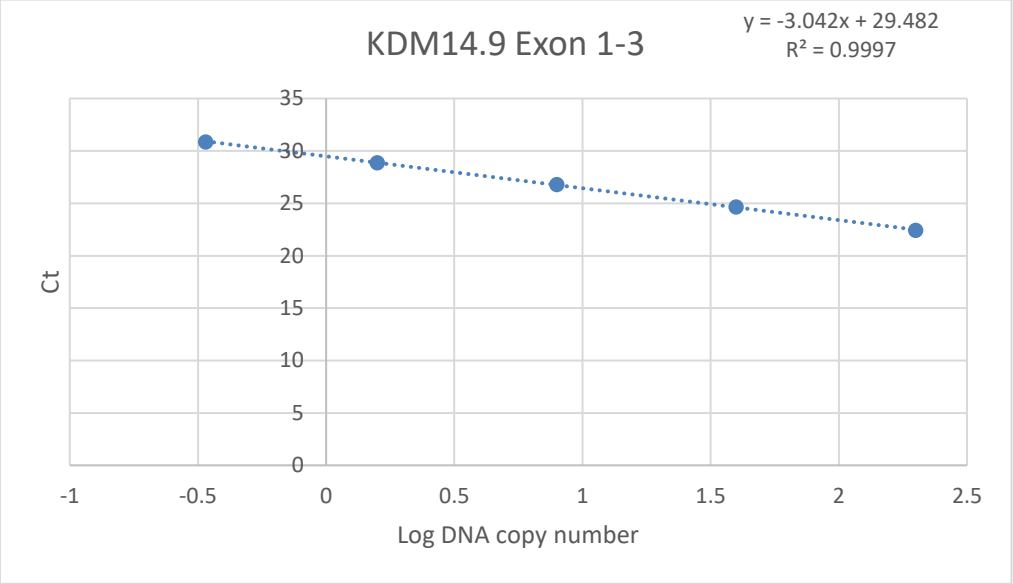
Primer efficiency: 104.6779307



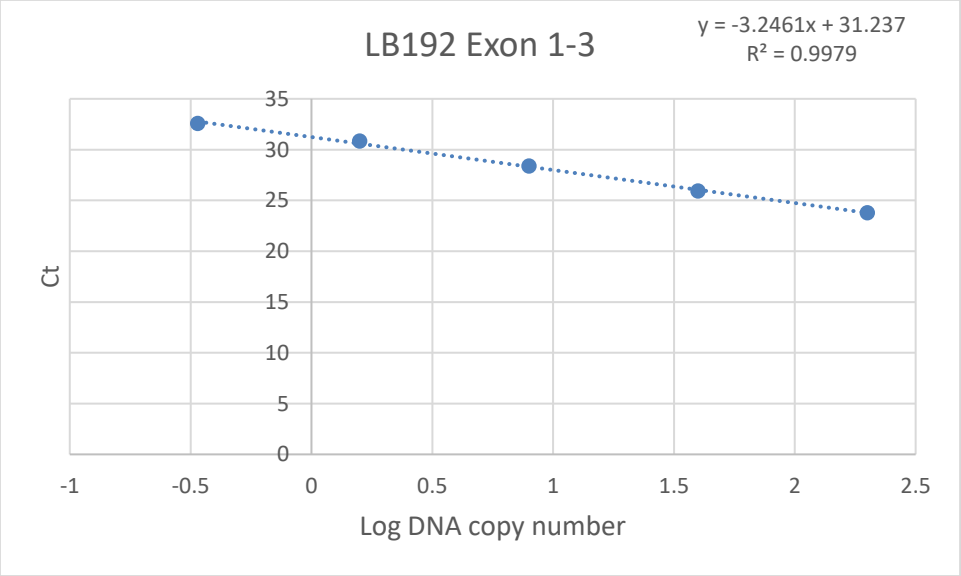
Primer efficiency: 97.88790712

Figure C1.1 RT-qPCR Standard curves for calculating RPS18 primer efficenies with control and patient DNA. A standard curve was set up for control (KDM14.9) or patient (LB192) cDNA using 100ng, 20ng, 4ng, 0.8ng or 0.16ng cDNA and RPS18 primers. The Ct value was plotted against the log value of the dilution and the primer efficey was calculated using the following equation: Primer efficieny = (10^(-1/slope value)-1)*100.

CDC45 Exons 1-3



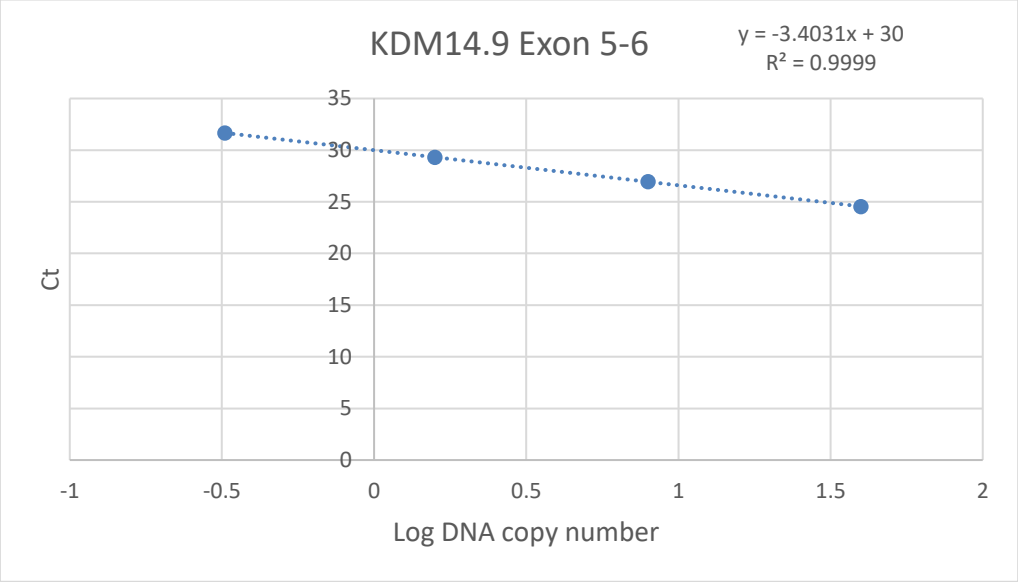
Primer efficiency: 113.1724605



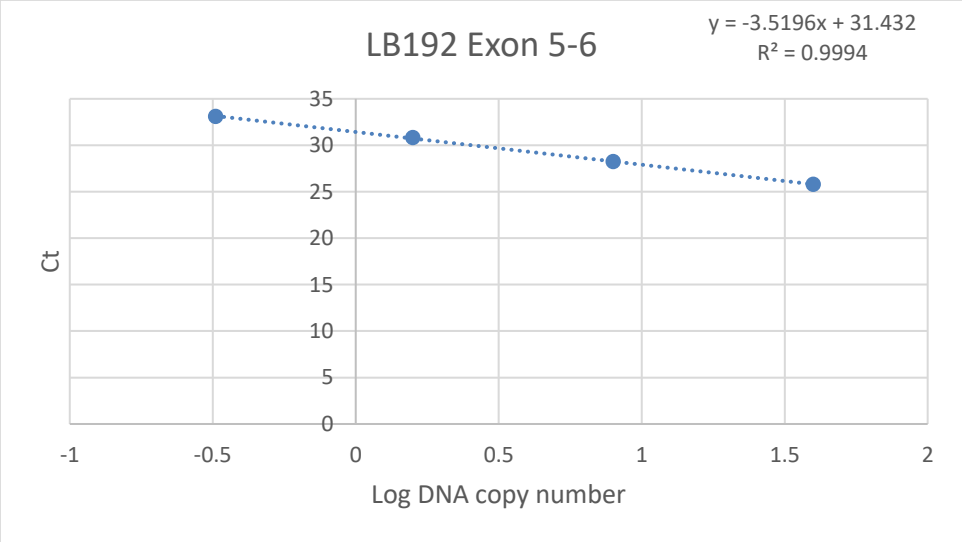
Primer efficiency: 103.2647088

Figure C1.2 RT-qPCR Standard curves for calculating CDC45 Exon1-3 primer efficenies with control and patient DNA. A standard curve was set up for control (KDM14.9) or patient (LB192) cDNA using 100ng, 20ng, 4ng, 0.8ng or 0.16ng cDNA and CDC45 Exon 1-3 primers. The Ct value was plotted against the log value of the dilution and the primer efficey was calculated using the following equation: Primer efficieny = (10^(-1/slope value)-1)*100.

CDC45 Exons 5-6



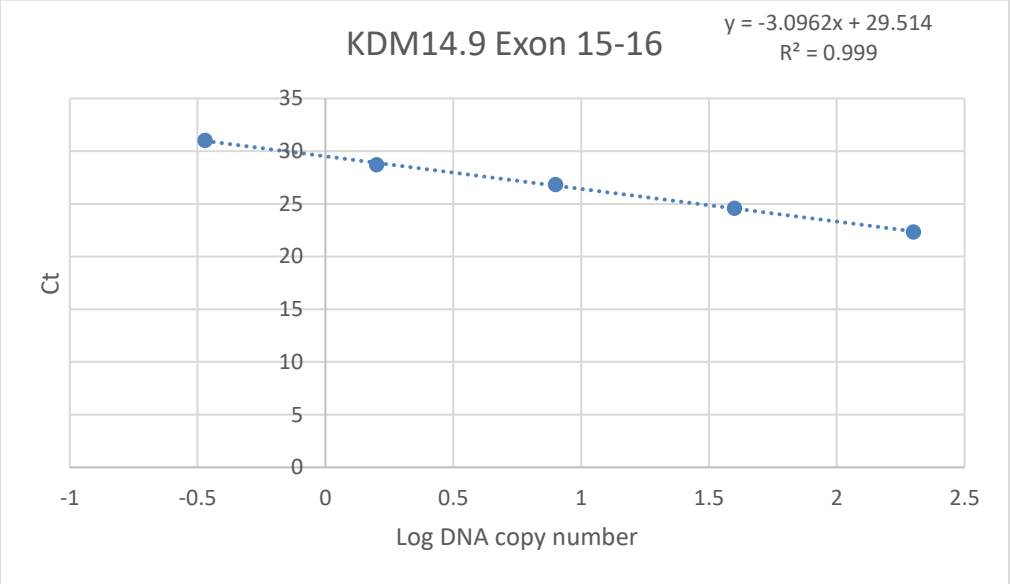
Primer efficiency: 96.72054787



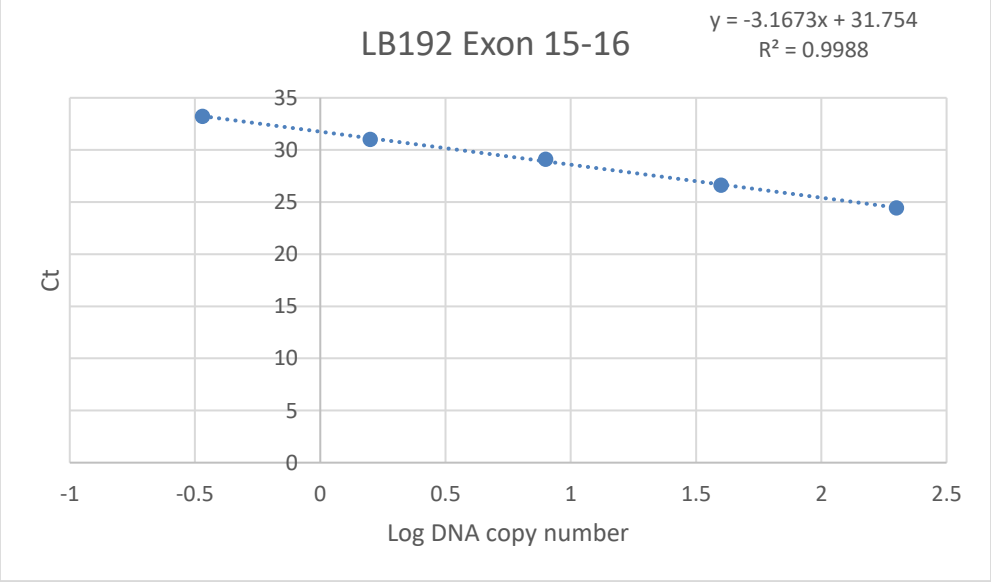
Primer efficiency: 92.36373275

Figure C1.3 RT-qPCR Standard curves for calculating CDC45 Exon5-6 primer efficenies with control and patient DNA. A standard curve was set up for control (KDM14.9) or patient (LB192) cDNA using 100ng, 20ng, 4ng, 0.8ng or 0.16ng cDNA and CDC45 Exon 5-6 primers. The Ct value was plotted against the log value of the dilution and the primer efficey was calculated using the following equation: Primer efficieny = (10^(-1/slope value)-1)*100.

CDC45 Exons 15-16



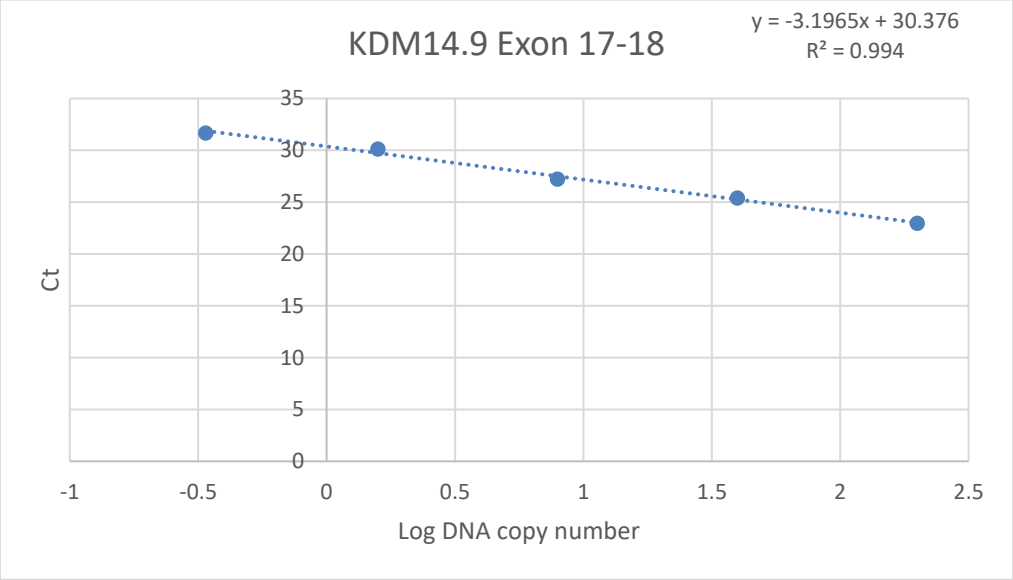
Primer efficiency: 110.3664857



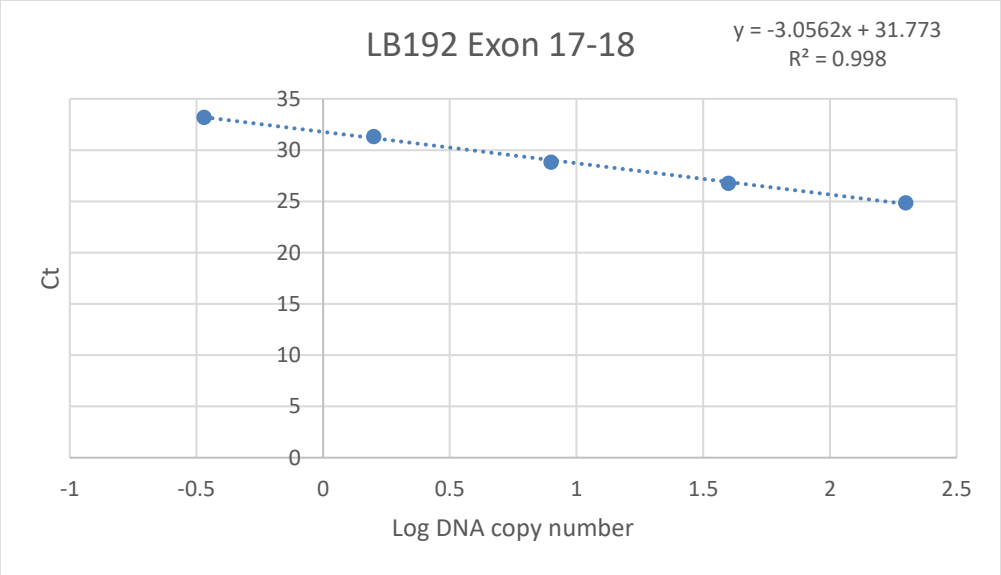
Primer efficiency: 106.8837256

Figure C1.4 RT-qPCR Standard curves for calculating CDC45 Exon 15-16 primer efficenies with control and patient DNA. A standard curve was set up for control (KDM14.9) or patient (LB192) cDNA using 100ng, 20ng, 4ng, 0.8ng or 0.16ng cDNA and CDC45 Exon 15-16 primers. The Ct value was plotted against the log value of the dilution and the primer efficey was calculated using the following equation: Primer efficieny = (10^(-1/slope value)-1)*100.

CDC45 Exons 17-18



Primer efficiency: 105.5143591



Primer efficiency: 112.4240658

Figure C1.5 RT-qPCR Standard curves for calculating CDC45 Exon 17-18 primer efficenies with control and patient DNA. A standard curve was set up for control (KDM14.9) or patient (LB192) cDNA using 100ng, 20ng, 4ng, 0.8ng or 0.16ng cDNA and CDC45 Exon 17-18 primers. The Ct value was plotted against the log value of the dilution and the primer efficey was calculated using the following equation: Primer efficieny = (10^(-1/slope value)-1)*100.

C.2 RT-qPCR biological replicates

RT-qPCR biological replicate 1

Table C.1 RT-qPCR biological replicate 1 expression fold change normalised to RPS18.

RNA 1 replicate		
	KDM14.9	LB192
Ex 1-3	1.07	0.28
Ex 5-6	1.02	0.25
Ex 15-16	1.02	0.17
Ex 17-18	1.01	0.27

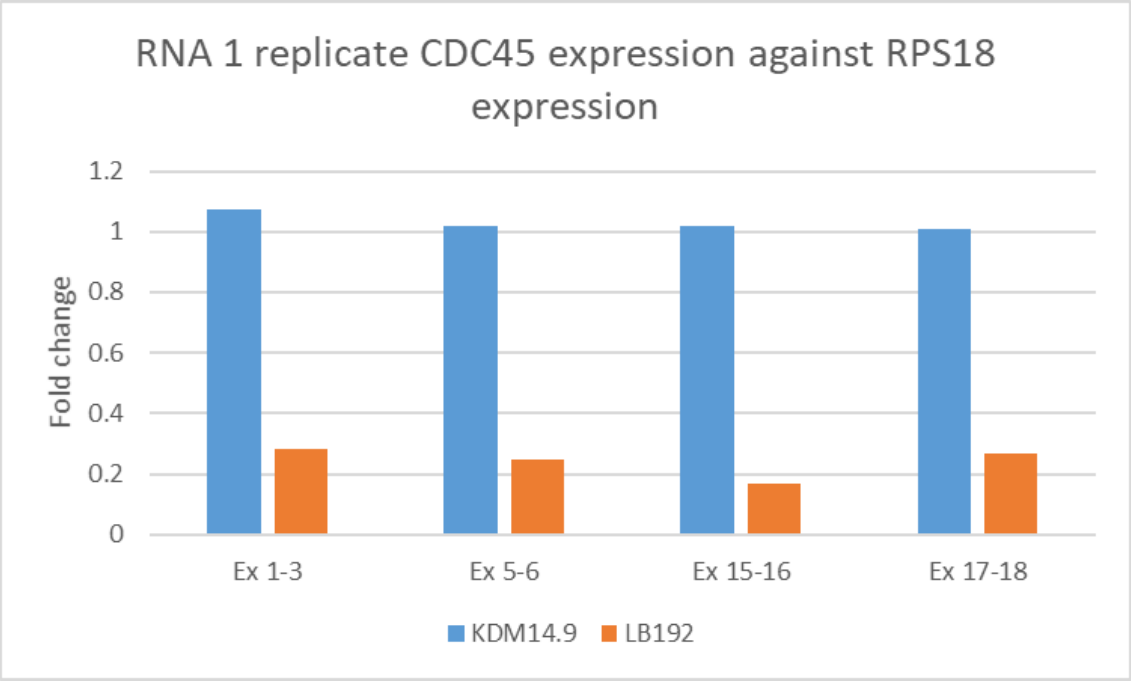


Figure C2.1 RT-qPCR biological replicate one shows a knockdown of CDC45 expression in LB192 Patient cells (orange) across the gene when compared to control DNA (KDM14.9 cells – Blue). Expression fold change was normalised to the hosekeeping gene RPS18.

RT-qPCR biological replicate 2

Table C.1 RT-qPCR biological replicate 2 expression fold change normalised to RPS18.

RNA 2 replicate		
	KDM14.9	LB192
Ex 1-3	1.00	0.21
Ex 5-6	1.00	0.23
Ex 15-16	1.01	0.16
Ex 17-18	1.00	0.17

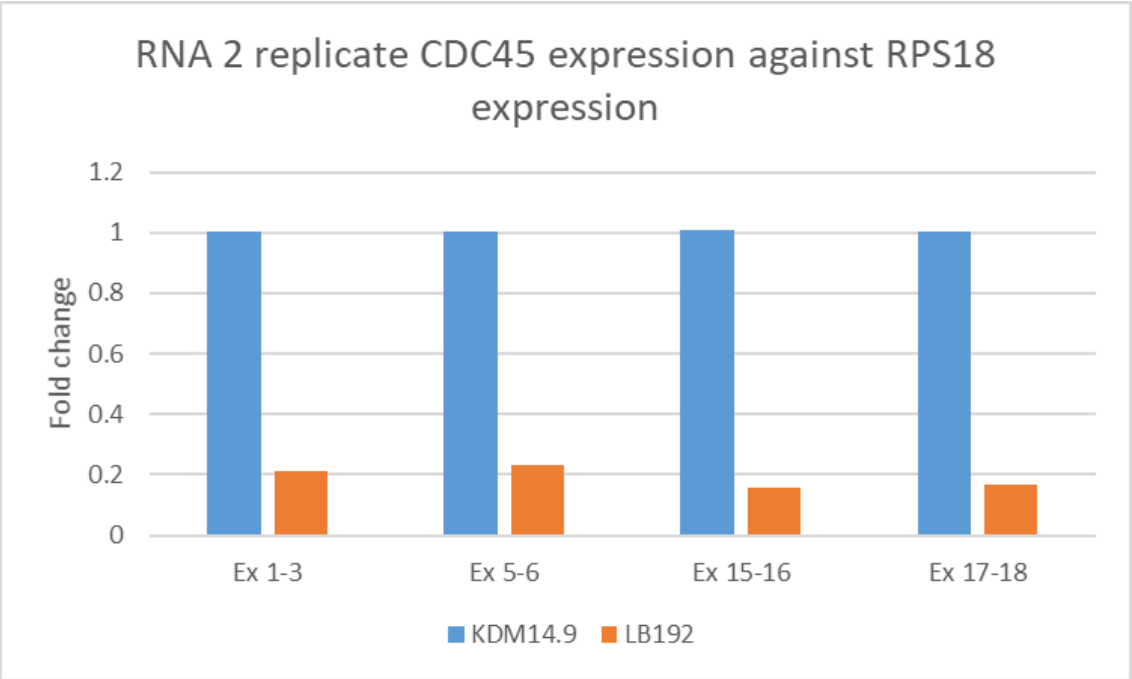


Figure C2.2 RT-qPCR biological replicate two shows a knockdown of CDC45 expression in LB192 Patient cells (orange) across the gene when compared to control (KDM14.9 – Blue). Expression fold change was normalised to the hosekeeping gene RPS18.

RT-qPCR biological replicate 3

Table C.1 RT-qPCR biological replicate 3 expression fold change normalised to RPS18.

RNA 3 replicate 1		
	KDM14.9	LB192
Ex 1-3	1.00	0.39
Ex 5-6	1.00	0.49
Ex 15-16	1.00	0.25
Ex 17-18	1.01	0.44

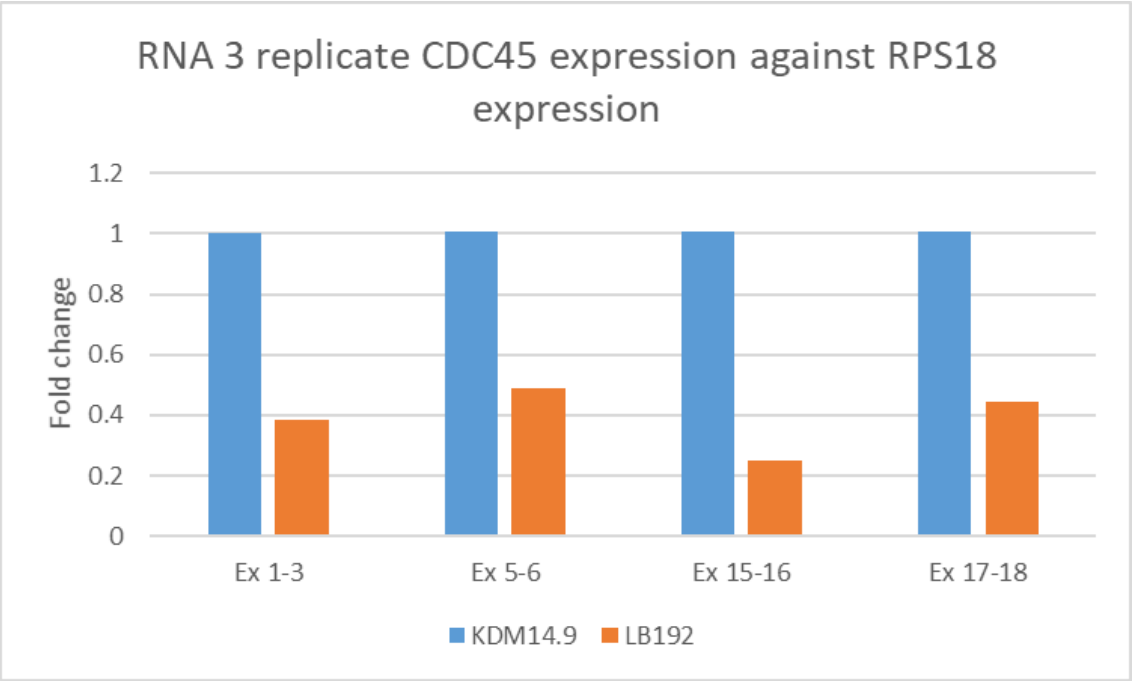


Figure C2.3 RT-qPCR biological replicate three shows a knockdown of CDC45 expression in LB192 Patient cells (orange) across the gene when compared to control (KDM14.9 – Blue). Expression fold change was normalised to the hosekeeping gene RPS18.