2021

Microsatellites and Their Association with Break Induced Replication

French J. Damewood IV
Wright State University

MICROSATELLITES AND THEIR ASSOCIATION WITH BREAK INDUCED REPLICATION

A thesis submitted in partial fulfillment of the
requirements for Master of Science

by

FRENCH J. DAMEWOOD IV

B.S., Wright State University, 2018

B.S.C.L.S., Wright State University, 2018

2021

Wright State University

WRIGHT STATE UNIVERSITY

GRADUATE SCHOOL

12/10/2021

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY French J. Damewood IV ENTITLED Microsatellites and Their Association with Break Induced Replication BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science.

_____

Michael I. Leffak PhD.
Thesis Director

_____

John V. Paietta, PhD
Chair, Biochemistry and
Molecular Biology Dept.

Committee on Final Examination:

_____

Michael P. Markey, PhD

_____

Kwang-Jin Cho, Ph.D.

_____

Barry Milligan, Ph.D.
Vice Provost for Academic Affairs
Dean of the Graduate School

ABSTRACT

Damewood IV, French J. M.S. Department of Biochemistry and Molecular Biology, Wright State University, 2021. Microsatellites And Their Association With Break Induced Replication

To study microsatellites instability and their repair pathways a dual fluorescent (DF2) and selectable (ganciclovir sensitive/ thymidine kinase (TK) expressing) cell system was assayed using replication fork stalling agents hydroxyurea and telomestatin. These cell lines carried ectopically integrated microsatellites derived from the Dystrophia Myotonica Protein Kinase (DMPK) gene ($(CTG)_{102}$ microsatellite), or an 88 bp polypurine/ polypyrimidine (Pu/Py) repeat from the PKD-1 locus, inserted into a FLP recombinase target site. These microsatellites form non-B DNA structures *in -vivo* and *in-vitro* causing replication fork stalling and double strand breaks. DF2 myc $(CTG)_{102}$ - TK cells treated with hydroxyurea were assayed for mutagenesis of the thymidine kinase gene (ganciclovir resistance) in the presence or absence of Polymerase Deltas 3rd subunit (POLD3). Knockdown in POLD3 lead to a decrease in TK mutagenesis numerous enough in cells possessing wild type POLD3 activity that mutation lead to increased survivability in the presence of ganciclovir but cell senescence without as demonstrated via Resazurin assay. Because break induced replication (BIR) and its mutagenic potential rely on the POLD 3rd subunit Polymerase Delta these results indicate that hydroxyurea damage of $(CTG)_{102}$ microsatellite is repaired by BIR. Cell lines containing an ectopically integrated Pu/Py repeat were also constructed and analyzed for BIR mutagenesis.

TABLE OF CONTENTS

LIST OF FIGURES

PCR Primer Tables

Acknowledgements

I would like to thank all of the people in my lab and in my life who inspired me, endured me, and helped me to accomplish this work. You all know who you are, and I couldn't have done it without you.

I: Introduction

Homology mediated repair mechanisms are a family of repair pathways, the intricacies of which are yet to be fully understood. One phenomenon known to be closely involved with these pathways is the inherent instability of microsatellites. These variable stretches of repetitive nitrogenous base pairs possess the ability to break under replication stress. These microsatellites also may form inter- and intra-molecular non-B DNAs, and also possess the ability to invade other DNA strands and generate structural variations such as translocations. This makes microsatellites an excellent suspect to implicate in a recently recognized form of homology mediated repair known as Break Induced Replication (BIR). To that end plasmids possessing a gene cassette carrying a $(CTG)_{102}$ microsatellite, and Alu elements were stable integrated into 406 Hela acceptor cell lines. These plasmids possess three identical ~300bp Alu repeats who's homology make them preferred regions for recombination in particular locations in between, upstream, and/or downstream of the microsatellite and selectable markers. These plasmids carried a thymidine kinase gene capable of assaying for DNA mutagenesis via induced ganciclovir resistance and a pair of fluorescent marker genes to assay for DNA double strand breaks by flow cytometry.

These cell lines were used to investigate the $(CTG)_{102}$ microsatellite's involvement in break induced replication (BIR). This was assessed with a cell viability assay utilizing resazurin's reduction to resorufin in the presence of aerobic metabolic conditions was used to measure cell senescence post ganciclovir exposure. In addition cell lines were constructed possessing the same dual fluorescent and selectable properties but with an 88 bp homopurine/ homopyrimidine (Pu/ Py) mirror repeat found in the PKD1 gene at intron 19 as well as 2 Alu elements. This repeat is prone to form non-Watson-Crick (non-B) DNA structures including G-quadruplexes and H-DNA triple helices, as demonstrated in vivo (Larson, et al., 2020).

Finally, data were gathered on cell lines containing a Pu/ Py 88 bp repeat previously shown to break under replication stress using 3rd generation long-read sequencing from Pac Bio (Eid, et al., 2009). These data were generated from inverse PCR experiments using genomic DNA harvested from cells previously treated with a G-quadruplex stabilizing agent called telomestatin. A data analysis pipeline utilizing both NGLMR and BWA-MEM was implemented to assess mutations accumulated in sequences downstream of the putative break site than upstream, consistent with the BIR repair mechanism and detect translocations at nonallelic loci consistent with genome instability resulting from BIR.

CoNvex Gap-cost alignMents for Long Reads also known as (NGLMR) is a mapping software that when paired with another program called SNIFFLES accurately maps PacBio or Oxford Nanopore (standard and ultra-long) sequence reads to a reference genome with a focus on reads that span structural variations. Further processing by a

Burrows-Wheeler aligner called BWA-MEM and import into a software like Integrated genome viewer then allows for visualization and assessment of data. (Li & Durbin, 2010)

To add to this data a large section upstream and downstream of the putative break site were also sequenced. All of this was performed to find a better understanding of the forms of replication fork restart and DNA damage repair pathways at play in our cell lines.

DNA Replication:

From meiosis to mitosis, no process is more crucial during a cell's life cycle than accurate DNA replication. With its semi-conservative nature (Stahl, 1958), being diploid or sometimes polyploid (Nelson & Cox, 2017), and utilizing enzymes for synthesis such as the exonucleolytic proofreader DNA polymerase I in bacteria and DNA polymerase epsilon in mammals (Nelson & Cox, 2017) (Mason & Cox, 2012), one might conclude that DNA replication is an error proof process. If the replication process was perfect, we would undoubtedly see far fewer ailments attributable to genomic disorders. Sadly, this is not the case. As termed by Lupski (Lupski, 2016), genomic disorders are a group of diseases caused by mutations and rearrangements of the human genome due to inherent genomic instability, resulting in susceptibility to structural variation mutagenesis (Carvalho & Lupski, 2016). These mutations often arise during the replicative process due to damage, mutagen exposures or failings in DNA replication and damage repair.

DNA Damage:

Genetic mutations may be introduced through both endogenous and exogenous means (Beadle & Tatum, 1941). Malefactors such as free-radicals, alkylating agents, as well as ionizing radiation and cytotoxic compounds have been implicated as causative agents for replication stress, DNA damage and mutagenesis (Roots, Kraft, & Gosschalk, 1985), (Shrivastav, Li, & Essigmann, 2010). As shown in Figure 1, these may be as simple as a neutron, gamma or beta radiation. They may also be as complex as a macrocyclic structure such as telomestatin or TmpyP4. These offenders may cause oxidation or thymine dimer formation as seen with exposures to UV and high ionizing radiation (Setlow, Swenson, & Carrier, 1963), (Beadle & Tatum, 1941), or they may stall replication as observed with hydroxyurea's inhibition of ribonucleotide reductase (Singh & Xu, 2016). Synthetically manufactured molecules can also stabilize non-B DNA structures as seen with telomestatin (TMS) and TMPYP4 (Seenisamy, et al., 2004), (Kim, Vankayalapati, Shin-ya, Wierzba, & Hurley, 2002) as well as cause the formation of inter-strand crosslinking that we see with cisplatin (Rocha, Silva, Quinet, Cabaral-Neto, & Menck, 2018). This damage can accumulate in the form of mutations, causing or even fighting cancers, or introducing developmental abnormalities like those seen in newborns exposed to a guanosine/ cytosine intercalating agent called thalidomide (Cooper-Roth, 2010), (Drucker, et al., 2003).

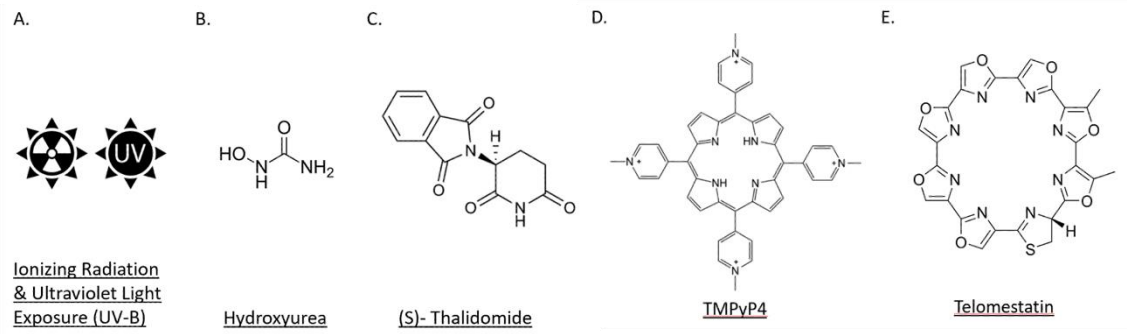


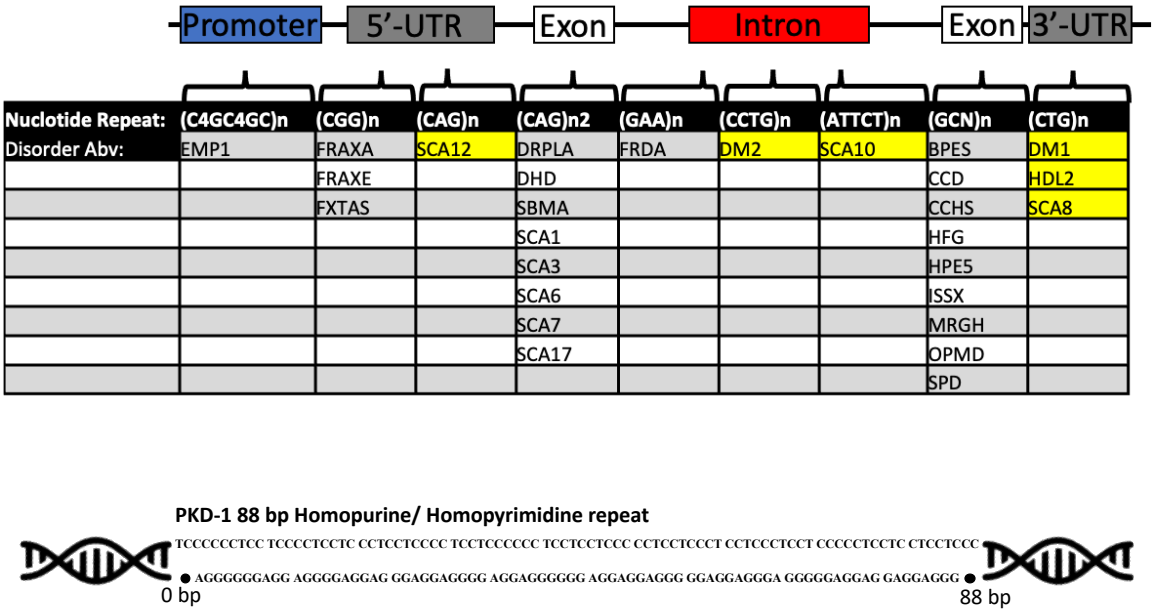
Figure 1: DNA and DNA replication assailing agents.

Examples of DNA assailing agents implicated in various cancers or known to cause DNA damage and replication stalling include, A: Ionizing radiation, B: Hydroxyurea, C: Thalidomide, D: TMPyP4, and E: Telomestatin.

Microsatellites:

In most cases as stated above, the stressor may be exogenous requiring an environmental exposure and can be as simple as exposure to the sun (Hader & Sinha, 2002). In some most cases the stressor is endogenous and may be attributed to nucleotide to the type or repetition or of the sequence's inter-molecular characteristics (Liu, et al., 2012), (Kramara, Malkova, & Osia, 2018). In these latter cases, a stretch of sequence known as a microsatellite, might be causing mutations globally with respect to the genome and not just locally at the break such as those seen in nonallelic homologous recombination (NAHR) (Hurles & Lupski, 2006) and might be full translocations as well (Leffak, 2017). Microsatellites have been characterized as tandemly repeated motifs possessing 1 to 12 nucleotides per repeat. Despite differences in DNA sequence, a common feature of these repeats is the ability to expand, contract, and/or form non-B DNA structures during replication or repair. These resulting non-B DNA structures can include hairpins, triplexes, and G-quadruplexes with the majority of expansion linked diseases being associated with di- and tri-nucleotide repeats (Kim & Mirkin, 2013). Under stressful conditions these non-B DNAs cause fork stalling, and will break in a single or double stranded manner (Kim & Mirkin, 2013). These short tandem repeats have been implicated in a number of diseases including Friedreich's ataxia FRDA (GAA)n, Fragile X syndrome FXTAS (CGG)n, Huntington's disease HD (CAG)n, and Myotonic Dystrophy type one and two, DM1(CTG)n and DM2(CCTG) (Mirkin, 2007). For a more exhaustive list, see Figure 2. Microsatellites can also be found in a variety of positions within the chromosomal landscape whether that be introns, exons, promoters, or untranslated regions (Mirkin, 2007). In addition, it has also been demonstrated that due to

the homology present at microsatellites structural variants in the form of translocations might arise during replication fork rescue and strand repair (Barthelemy, Hanenberg, & Leffak, 2016).



Nuclotide Repeat:	(C4GC4GC)n	(CGG)n	(CAG)n	(CAG)n2	(GAA)n	(CCTG)n	(ATTCT)n	(GCN)n	(CTG)n
Disorder Abv:	EMP1	FRAXA	SCA12	DRPLA	FRDA	DM2	SCA10	BPES	DM1
		FRAXE		DHD				CCD	HDL2
		FXTAS		SBMA				CCHS	SCA8
				SCA1				HFG	
				SCA3				HPE5	
				SCA6				ISSX	
				SCA7				MRGH	
				SCA17				OPMD	
								SPD	

Figure 2: Microsatellites and Chromosomal positions.

Examples of microsatellites their associated chromosomal positions. Microsatellites of interest associated with disease. SCA8: Spinocerebellar ataxia type 8, SCA10: Spinocerebellar ataxia type 10, SCA12: Spinocerebellar ataxia type 12, DM1&2: Myotonic dystrophy type 1&2, and HDL2: Huntington disease-like 2 are of importance to this work and highlighted in the table above. Below them, a diagram of the PKD-1 88 bp Homopurine/ Homopyrimidine repeat that is referenced later.

Replication Stress, DNA Damage, Double and Single Strand Breaks:

On their own, or in the presence of replication stressors, microsatellites can form non-B DNAs. Hairpins and Hoogsteen-DNA (non-canonical major groove base pairing resulting in G-quadruplexes or triplexes) are relevant examples and diagramed in Figures 3 and 4. The simple hairpin which is associated with the CTG microsatellite and the unimolecular lateral loop G-quadruplex associated with the homopurine/ homopyrimidine Pu/ Py 88 bp mirror repeat found within the PKD1 gene are the two of most importance to this paper and are found in Figure 3.

As an example, the CTG/CAG microsatellite has been implicated in hairpin formation, while the Pu/Py repeat has been shown to form a type of Hoogsteen DNA called G-quadruplexes. It has been demonstrated *in vitro* that these quadruplexes can be further stabilized using intercalating agents like telomestatin, and that *in vivo* treatment of cells containing this Pu/Py 88 bp repeat associated with the PKD1 gene may form replication barriers, causing DNA double strand breaks (Liu, Chen, Bissler, Siden, & Leffak, 2010), (Liu, et al., 2012) Figure 4. It has also been demonstrated that constructs integrated into 406 HeLa acceptor cells carrying this mirror repeat show preferentiality such that replication is stalled and breaks occur more frequently when the poly-purine strand of the mirror repeat is in the lagging strand (Liu, et al., 2012).

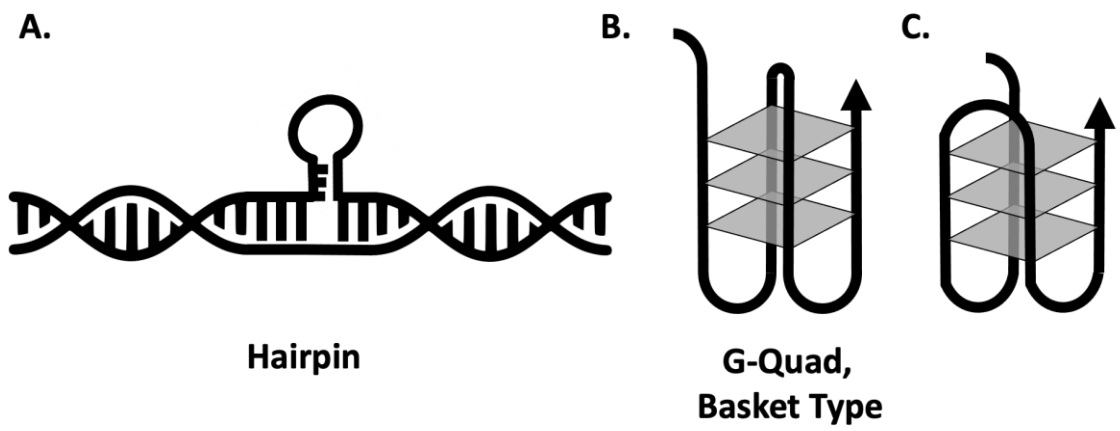


Figure 3: Non-B DNAs

Panel A: Simple hairpin seen with CTG/CAG repeats, Panel B & C: Possible forms of G-quadruplex associated with the Py/Pu 88 bp repeat.

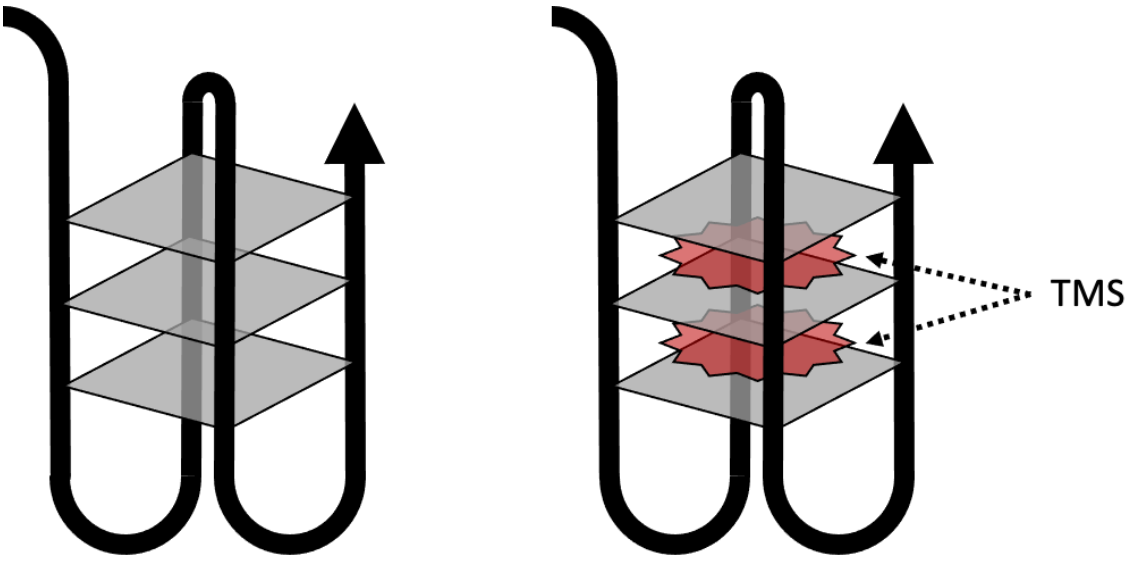


Figure 4: Telomestatin G-Quadruplex Intercalation

Simplified diagram of TMS intercalation between the intramolecular base stacking seen in G-quadruplex formation.

As stated above, these non-B DNAs have been shown to form barriers to replication and cause replication fork stalling and double stranded DNA breaks after treatment with replication stressors such as hydroxyurea as well as in the absence of exogenous replication stress. (Liu, Chen, Bissler, Siden, & Leffak, 2010), (Liu, et al., 2012). These breaks come in two forms, double and single stranded. Among these two, double-strand breaks (DSBs) are particularly detrimental resulting in insertions, deletions, or chromosomal translocations that are implicated in many human cancers (Pannunzio, Watanabe, & Lieber, 2018).

A double strand break is a break across both strands of DNA on a single chromatid whereas a single strand break only involves one strand (Sung, 2018). A single strand break leaves the opposite strand intact which is far more favorable for repair. With double strand breaks being the most deleterious form of break, the cell has developed several methods to rescue the replication fork and repair the damaged strand(s). These strategies are often chosen based on the post-break environment. This post break environment is incompletely characterized but includes cell cycle phase, chromatin structure, the proximity of replication and repair enzymes, and the accessibility of homologous sequences close to the break as well. While some of these repair methods are fairly accurate, as the post break environment becomes more complicated to work with the cell must choose from less-than-ideal repair methods (Her & Bunting, 2018). Homologous recombination (Figure 5) is the most accurate method of repair, while nonhomologous end joining, single strand annealing and break induced replication are more mutagenic.

It has been demonstrated due to the homology present at the microsatellites that structural variants such as translocations might arise during fork rescue and strand repair (Lewis, et al., 2019). Within the context of human disease, structural variants are typically characterized as deletions, duplications, triplications, amplifications (for example, quadruplications) and other larger copy number variants (CNVs) as well as copy number-neutral inversions, insertions and translocations (Yilong Li, 2020). Structural variants are often implicated in human disease, the most apparent of which are translocations. One example, the “Philadelphia Chromosome” BCR-ABL fusion gene, produces severe consequences such as a transcription factor gain of function due to a reciprocal translocation - resulting in an increase in Chronic Myelogenous Leukemia’s disease progression (Nowell & Hungerford, 1960), (Annelies de Klein, 1982).

DNA Repair:

In order of decreasing repair fidelity, the 3 most common methods of double strand break repair are: homologous recombination, non-homologous DNA end joining, and alternative end joining (Sung, 2018).

Homologous Recombination:

Homologous recombination diagramed in Figure 5 is a repair method relying on ~200 bp of homologous sequence (Rubnitz & Subramani, 2021) most often on the sister chromatid in proximity to a DNA double strand break. Once broken, exonuclease resection of the break occurs in a 5' to 3' manner generating 3'-OH single-stranded DNA tails by either the 5' to 3' exonuclease Exo1 or the Dna2 nuclease. Free 3' ends locate and invade a homologous donor sequence to prime and repair DNA synthesis via formation of a RAD51 recombinase nucleoprotein filament. D-loop formation and DNA polymerase extension from the 3' end of the invading strand and capture of the second DSB end by annealing to the extended D-loop forms two crossed strands or Holliday junctions. These holiday junctions are resolved by resolvases (nucleases) to give either crossover or non-crossover products (Wright, Shah, & Heyer, 2018) (Filippo, Sung, & Klein, 2008).

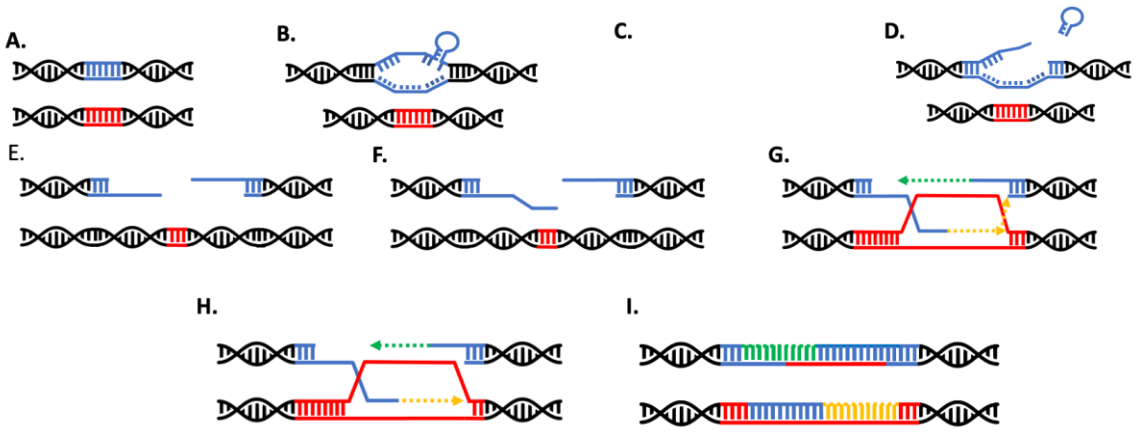


Figure 5: Homologous Recombination

Panel A: Two sister chromatids. Panel B: Replication stressor or non-B DNA. Panel C: DNA damage. Panel D: Single strand break. Panel E: 5' resection generating 3' single strand that searches for homology and upon finding it invades a sister chromatid annealing to this sequence. Panel F: D loop is formed and replication of missing sequence using sister chromatid as template begins. Panel G: resolution by strand displacement and strand annealing proceeded by DNA ligation and synthesis can cause the formation of non- crossover products. Double holiday junction formation via second end annealing followed by DNA synthesis and Ligation leading to crossover or non-crossover products via decatenation and or nucleolytic resolution

Alternative End Joining:

Alternative end joining (a-EJ) is a blanket term used to characterize pathways used to repair double strand breaks that are not homologous recombination or non-homologous end joining and depend on little to no nucleotide homology to reseal DSBs resulting in the loss or gain of a small number of nucleotides at the repair joint. Like homologous recombination, repair of DNA double-strand breaks (DSBs) is initiated by exonucleolytic end resection generating 3' single strands. This is typically carried out by MRE 11, CTIP and later EXO1 or DNA2 following RPA coating of the region exposing 3' end homology allowing the two broken ends to synapse and stabilize. Stabilization is followed by a ligation by either DNA ligase 1, or 3 (Lu, et al., 2016). As shown in Figure 6, there are no D-Loops or replication bubbles to resolve making this one of the simpler repair methods. This method or repair's simplicity can lead to errors though. If homology used for this synapsis is found further down from the break the ligation may leave 3' tails. These 3' tails are then cleaved by ERCC1/ XPF leading to a loss in sequence (Sallmyr & Tomkinson, 2018). Although the a-EJ pathways make only a minor and poorly understood contribution to DSB repair in nonmalignant cells, there is growing interest in these pathways as they generate large deletions, translocations, and end-to-end chromosome fusions, genomic rearrangements that are frequently observed in cancer cells (Boboila, 2010), (Mateos-Gomez, 2015), (Sallmyr & Tomkinson, 2018) Figure 6.

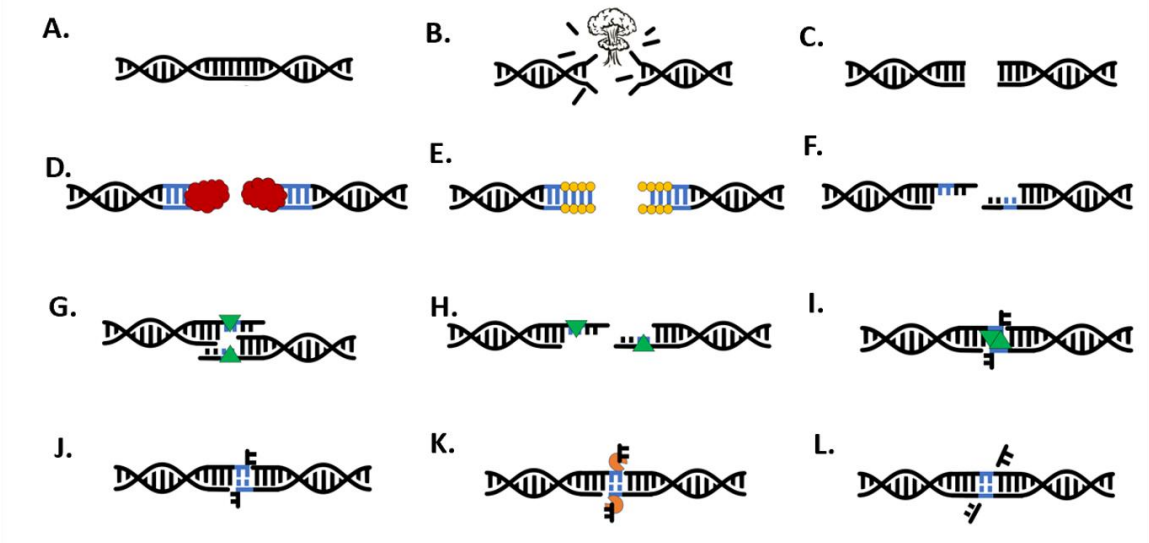


Figure 6: Alternative End-Joining

Panel A: Single chromatid. Panel B: DNA damage. Panel C: Double strand break. Panel D: PARP-1 localization to the break. PARP-1 facilitated MRN and CtIP recruitment to the break. Panel E: RPA recruitment. Panel F: EXO1 and DNA2 recruited via RPA and further end processing to reveal compatibility needed for annealing. Panel G: Rad52 localization. Panel H: Rad52 nucleofilament stabilization. Panel I: Strand annealing. Panel J: 3' non homologous tails exposed. Panel K: ERCC1 and XPF localization. Panel L: 3' tail cleavage and possible sequence loss

Non-Homologous End Joining:

Non-homologous DNA end joining (NHEJ) in human cells, appears to repair nearly all DSBs outside of S and G2 cell cycle phases and even about 80% of DSBs within S and G2 that are not proximal to a replication fork, NHEJ usually requires 1 to 4 base pairs of microhomology. This process requires Ku 70/80 heterodimer recruitment to stabilize the two broken ends. This is followed by enzymatic processing by Artemis, Aprataxin and PNK-like factor via forkhead-associated nuclease activity to resect and process both ends of the break (Kanno, et al., 2007) (Sicong Li, 2011). After end polishing the LIG4/XRCC4/XLF/PAXX complex brings together and ligates the two ends. Enzymatic end processing can result in the loss of nucleotides and increases repair outcome diversity (Pannunzio, Watanabe, & Lieber, 2018) (Mladenova, Mladenov, & Iliakis, 2019).

The repair pathway chosen by the cell may have more consequences than just a loss of a chromosome from damage, garbled code from mutations, or a quality repair. It was hypothesized the types of breaks and post break environment seen in mammalian cell lines possessing a $(CTG)_{100}$ microsatellite and its associated non-B DNA structure (hairpin), might be repaired and replication restarted, using a highly mutagenic repair pathway known as break induced replication (BIR) that, until recently was only demonstrated in yeast (Saini, et al., 2013).

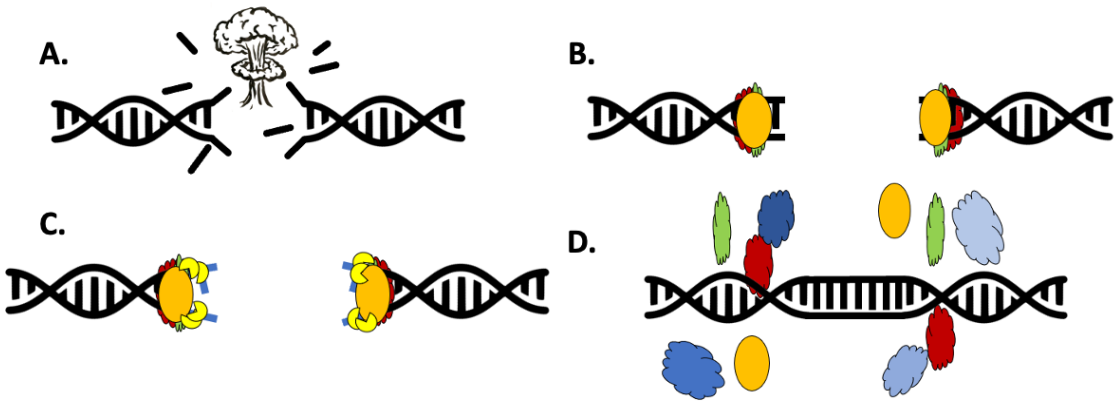


Figure 7: Non-homologues End joining

Panel A: Double strand break. Panel B: The two broken ends with little to no homology are first recognized by a heterodimer consisting of Ku70 and Ku80 followed by binding heterodimer and strand stabilization. DNA-PKcs also possessing high affinity for DNA ends is recruited to the break as-well as p53BP1. stabilizing the two broken ends and recruit other factors. Panel C: Other nucleolytic end processors such as Artemis, Aprataxin, and PNK-like factor localize to the break site and bind facilitating end processing. DNA polymerase Pol λ/Pol μ and LIG4/XRCC4/XLF/PAXX localization occurs and Pol λ/Pol μ facilitated synthesis as well as blunt end ligation via LIG4/XRCC4/XLF/PAXX seals and repairs the damaged strands. Panel D: this is followed by Enzyme degradation/ disassociation.

Break Induced Replication/BIR:

Break induced replication is a rare double strand break repair mechanism once thought to occur only in yeast. Now thought to be a mutagenic repair method, BIR is making its way into mammalian and even human models for DNA replication repair studies. While there are thought to be many forms of BIR such as Rad51 dependent, Rad51 independent, or microhomology mediated (MMBIR) they all share some similarities. The BIR repair pathway is similar to Homologous Recombination (HR) where a single end of the DSB invades a homologous sequence. BIR's occurrence is most frequent, but not limited to when the DSB manifests as a 'one-ended double strand break'. These may occur during replication through a DNA lesion that results in fork stalling and collapse, or telomere erosion exposing a single 3' end (Kramara, Malkova, & Osia, 2018). Double strand breaks as previously stated, occur readily during mammalian DNA replication in the presence of microsatellites or upon treatment with DNA assailing compounds. However, one of the hallmarks for BIR, the single ended double strand break, is rare.

BIR differs in many ways from HR in its accuracy, its machinery, and the method of its "resolution". Two different pathways of BIR have been demonstrated in yeast, Rad51 dependent (canonical) and Rad51 independent (noncanonical) (Kramara, Malkova, & Osia, 2018) (Malkova, Naylor, Yamaguchi, Ira, & Haber, 2020). While these may be distinctly different pathways, they both rely on common repair machinery such as RPAs strand stabilization and repair machinery recruitment (Ruff, Donnianni, Glancy, Oh, & Symington, 2016) and Rad52 nucleofilament stabilization (Sotiriou, et al., 2016). One characteristic unique from other forms of homology mediated repair is the involvement/ dependance of polymerase delta's 3rd subunit. This has been shown to be indispensable

for BIR in both yeast (Pol32) and mammals (POLD3) and is thought to be involved in BIR's mutagenic effects (Kononenko, Ebersole, Vasquez, & Mirkin, 2018).

Of the two, Rad 51 dependent BIR is far better characterized, as shown below this begins with a single 3' ended double strand break. Following resection, this 3' end synapses with homologous sequence to form a displacement loop (D-loop) and synthesis is then initiated at the 3' end. Leading-strand synthesis is carried out by Pol Delta utilizing a helicase that has yet to be determined. This generates a long ssDNA and accumulates unrepaired DNA lesions in this new leading strand by an unknown method (Figure 8 E). Lagging-strand synthesis is thought to copy this strand and any mutations with it. While the exact mechanism of the D-loops resolution is hotly debated it is believed BIR's outcome can result in conservative inheritance of any mutations accumulated (Malkova, Naylor, Yamaguchi, Ira, & Haber, 2020), (Kononenko, Ebersole, Vasquez, & Mirkin, 2018).

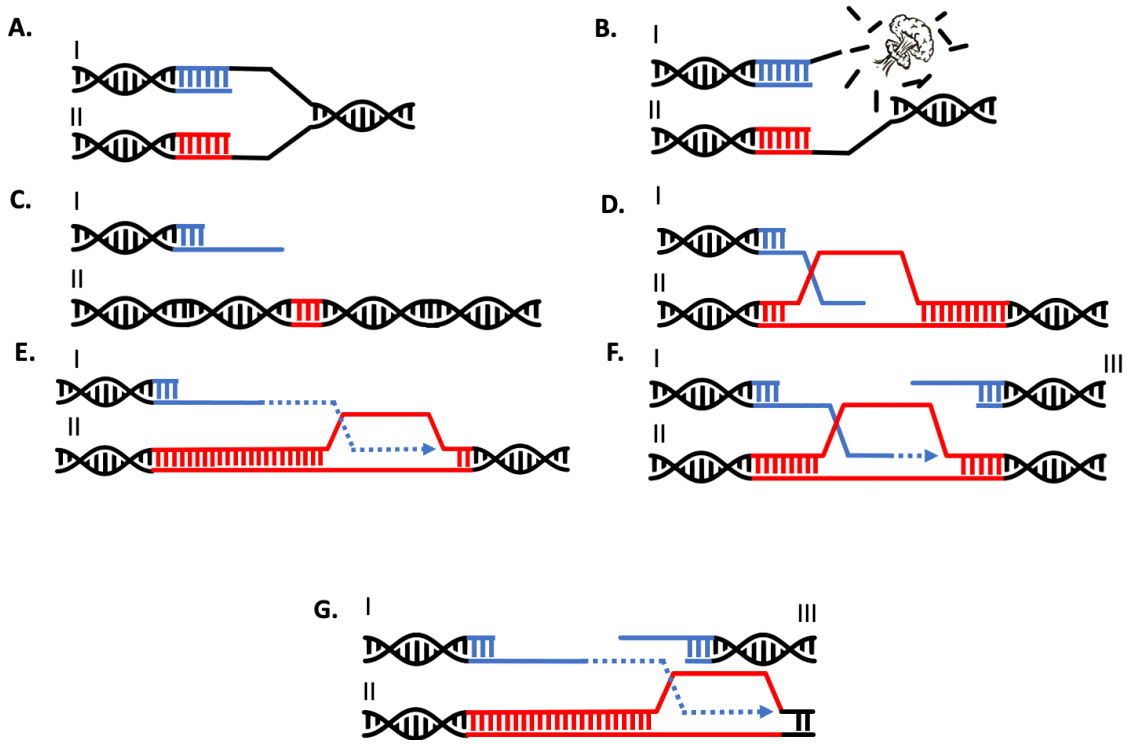


Figure 8: Diagram of proposed BIR repair pathways.

A: Two replicating sister chromatids. B: Replication fork encountering an obstacle or damage occurring. C: Single ended double strand break revealing 3' tail exposed. D: Nascent DNA (II) ligates with broken template strand leaving a single ended DSB (I). 5' resection generates an invading 3' end for BIR. Strand invasion and annealing leads to the formation of a D-Loop 3' tail extension using the sister chromatid or foreign homologous sequence as template. E: D loop progression and DNA synthesis using PolDelta as the primary replicative polymerase as shown in yeast (Pol32) and in mammalian cell lines as well leading to an accumulation of mutations or other abnormalities such as insertions, deletions half-crossover products F & G: Chromosome fragment downstream of DSB (III) is replicated from a downstream origin (leftward moving fork, not shown) generating double sided DSB for BIR.

II. Specific Aims

With BIR's activity having been demonstrated by others to be dependent on POLD3 expression, our lab's demonstration of CTG/ CAG and Pu/Py microsatellite instability (Lewis, et al., 2019), and the demonstration of our cell lines ability to assay for mutations and translocations in the face of Hu or TMS exposure (Gadgil, et al., 2020), the following research was address these three questions and set up upcoming experiments.

1. What is the effect of POLD3 knockdown on BIR from ectopically integrated (CTG/CAG) $_{102}$ microsatellites?

2. What is the effect of replication polarity on the stability of a homopurine/ homopyrimidine microsatellite?

3. What is the pattern of mutagenesis of the homopurine/homopyrimidine (Pu/Py) G4 forming mirror repeat under replication stress?

III. Methods

Plasmid culture and isolation:

All plasmids requiring initial isolation or culture from dH5a E. coli were either streaked to selective LB agar plates using a sterile loop or glass bead technique. These plates were left at 37°C in the absence of light for 24 to 48 hours or until colonies could be isolated. These colonies were picked via sterile loop, and cultured over night at 37°C in a 75 ml flask containing LB broth containing either ampicillin, kanamycin, or chloramphenicol.

Plasmid preparations:

Plasmid preparations were purified via commercial column (Omega-BioTek) and eluted using heated elution buffer. Plasmid DNAs were photometrically quantified using a Nano Drop One (Thermo Scientific) and confirmed using gel electrophoresis. To confirm correct sequence, plasmids were Sanger sequenced by Retrogen.

Plasmid transformations:

Plasmids and ligation products were transformed into bacteria by adding 10 µl of reaction or 20 ng of plasmid DNA to "Stable Cells" (New England Biolabs) on ice after thorough mixing for 30 minutes. Preparations were heat shocked for 30 seconds at 42 °C and placed on ice for 2 minutes. Transformations were incubated in 500 µl stable outgrowth media for 1 hour, and spun down at 300 rpm for 5 minutes. Transformed bacterial cell pellets were resuspended in 20-30 µl of LB/SOC media with no antibiotic and plated for isolation on antibiotic/ LB agar plates.

Restriction digestions:

All restriction digestions were carried out according to NEB™ protocols for the designated enzymes with the following modifications for Swa1 and Pac1 double digestion. Restriction digestion for these enzymes was initially carried out on both the vector backbone and the PCR fragments with PacI (NEB™) using Cutsmart® Buffer (NEB™) overnight at 37 °C. For these double digests, after Pac1 reactions were heat inactivated by placing the reaction in a 65 °C sand bath for 20 min., 1.2 M NaCl was added to this preparation at 1/11 ratio before adding Swa1 enzyme. This second digestion by SwaI (NEB™) was carried out at room temperature for 4 hours and quenched in a 65 °C sand bath for 20 minutes.

Dharmacon SMARTvector™ Inducible mCMV Turbo GFP (POLD3- 1,2, and 3; PuroR & BlastR) PCR:

To make a blasticidin resistant version of the commercially purchased puromycin resistant Dharmacon SMARTvector™ Inducible mCMV Turbo GFP POLD3 vector 25 µl PCR reactions were performed in duplicate using Takara Bio GXL PCR protocols with CMV/BSD Invitrogen® and Dharmacon® SMARTvector™ Inducible mCMV Turbo GFP POLD3 PuroR plasmids as templates. These plasmids came as a set of 3 with their respective shRNAs targeting different portions of the POLD3 subunit. These preparations were run for 30 cycles using the following PCR scheme and utilizing Takara GXL polymerase, 98 °C denaturation step for 10 seconds, variable annealing step temperature according to primer Tm for 15 seconds, and a 68 °C extension step of 15-30 sec per kb using 60 ng template in each preparation (Figure 9).

Primer	Sequence
OUTER CMV FORWARD 1	**5'-GGTATAAGAGGCGCGACCAGCGTCGGTA-3'**
OL BSD-P2A REV2	**5'-TCCAGGCCCTCCCACACATAACC-3'**
OUTER REVERSE 3	**5'-TTGTAAGTCATTGGTCTTAAAGGTA-3'**
OL BSD-P2A FOR 3	**5'-GGTTATGTGTGGGAGGGCCTGGAGCAGCTGTTGAATTTT-3'**
OL CMV-BSD REV1	**5'-TTGGCCATGGTGGCGCTGTTAACT-3'**
OL CMV-BSD FOR2	**5'-GAGTTAACAGCGCCACCATGGCCAAGCCTTTGTCTC-3'**

Primer Table 1: Primers used for PCR of fragments used in Gibson assembly of blasticidin resistance carrying shRNA lentiviral vector carrying shRNA against POLD3.

TTF (Pu) and TTR (Py) Cloning PCR:

PCR preparations were run in duplicate using a previously made plasmid carrying the PKD1 88 bp Pu/PY repeat Py (TTF) orientation as template. PCR was carried out for 30 cycles using Takara GXL polymerase utilizing the following PCR scheme. 98 °C for 10 seconds, 58 °C for 15 seconds, and 68 °C for 1 minute. All reactions were purified via commercial column (Omega BioTek).

Primer	Sequence
TTFPACF	**5'-TAGCTGTTAATTAAGCCAGTGAATTCGCAACGGCTACAA-3'**
TTFSWAR	**5'-ACGTGAATTTAAATCCTACCGAGCTCCATTAGTGAAGAT-3'**
TTRSWAF	**5'-TGACTGATTTAAATGCCAGTGAATTCGCAACGGCTACA-3'**
TTRPACR	**5'-ACGTGATTAATTAACCTACCGACGTCCACCAGTGAAGAT-3'**

Primer Table 2: Primers used to clone TTF/88 base pair Poly pyrimidine repeat and TTR/ 88 base pair Poly purine repeat out of previously made DF2 construct generating the necessary tails for digestion with Pac1 and Swa1 for ligation.

Inverse PCR (iPCR) & dTomato/ eGFP Genomic PCR:

i-PCR & dTomato/ eGFP genomic PCR was carried out following Q5 Polymerase (NEB®) protocols and included a "touchdown" annealing step wherein the first 10 cycles of annealing were carried out at 66 °C for 15 seconds with the temperature reducing by .6 degrees each cycle then the subsequent 25 cycles at 60 °C for 15 seconds. Each cycle also had an extension time of 3 minutes per. PCR products were run on a gel and extracted via commercial column (Omega BioTek). The i-PCR samples were pooled and sent to GeneWiz® for Pac-Bio® long read sequencing. See (Figure 20) for i-PCR mechanism.

Primer	Sequence
Fwd eGFP	5'-CTAGGGATAACAGGGTAATATAGCTCG-3'
Rev eGFP	5'-GGCCGCTTGCTAGCTATAGTTCT-3'
Fwd dTomato	5'-CTTCTTCAACATCTGGTCCAAAC-3'
Rev dTomato	5'-TTATCCCTAGCGTAACTAGATGCT-3'

Primer	Sequence
NheFor	5'-AAGCTTGCCTTGAGTGCTTC-3'
eGFP Rev	5'-GTCTTGTAGTTGCCGTCGTC-3'

Primer Table 3: Primers used for dTomato and eGFP PCR and i-PCR to generate products for Pac Bio long read sequencing.

Diagnostic PCR/ Traditional PCR for sequencing:

Diagnostic PCR was preformed using Q5 Hot Start polymerase and manufacturer described protocols to test for proper orientation and integration using primers flanking the multiple cloning site containing the PKD1 88 bp repeat. Traditional PCR for correct integration was also performed using Q5 Hot Start protocals and MCSF/MCSR primer sets below. Purified plasmids demonstrating the correct PCR products were sent to Retrogen™ for sequence confirmation. These plasmids were also transfected into FRT integration site containing 406 HeLa acceptor cells using the methods described below and stably integrated cell lines were analyzed using flow cytometry to confirm their dual fluorescent properties. See Figure 17 for diagnostic PCR mechanism for orientation determination.

Primer	Sequence
MCSF	5'-GGCCGCGTTTAAACTTAATTAAATTTAAATC-3'
MCSR	5'-CTAGGATTTAAATTTAATTAAGTTTAAACGC-3'
Red	5'-GCGGCCGCGTTTAAACCCAGTGA-3'
Blue	5'-TTAACCGAGCTCCACCAGTGAAGATG-3'

Primer Table 4: Diagnostic PCR primers to test microsatellite orientation.

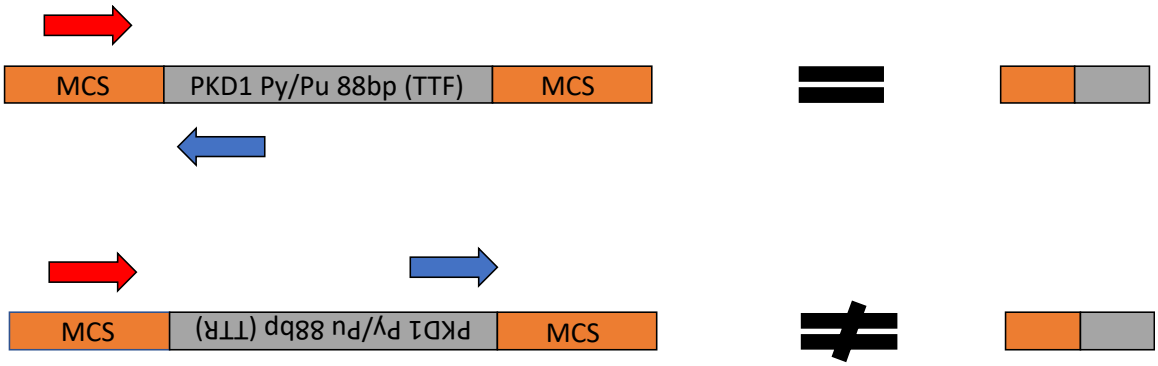


Figure 9: Diagnostic Primer Mechanism

Diagnostic PCR testing orientation was performed using primers that give (or do not give) products dependent on primer direction. Forward primers are displayed in red and reverse are in blue. When flipping the gene and integrating it we can see that no product will form due to both primers facing the same direction.

Ligations and Gibson assembly:

Ligations were performed using T4 DNA Ligase (NEB™) and a molar ratio of 1:3 plasmid backbone to insert at 25°C 1 hour, then overnight at 5-10 °C and later heat inactivated in a 65 °C sand bath for 20 minutes. Ligation products were transformed into bacteria by adding 10 µl to "Stable Cells"(New England Bio Labs) on ice after thorough mixing for 30 minutes, heat shocked for 30 seconds at 42 °C, then placed on ice for 2 minutes. The bacteria were then incubated in 500 µl SOC media for 1 hour, and spun down at 300 rpm for 5 minutes before plating to selective LB agar plates..

Gibson assembly was performed on digested Dharmacon SMARTvector™ Inducible mCMV Turbo GFP (POLD3- 1,2, and 3) and fragments pcr'd using this vector or CMV/BSD Invitrogen® containing a blasticidin resistance gene as template. Using the 4-6 fragment Gibson assembly protocol from New England Biolabs (Gibson, et al. 2006) and these digest/ PCR products containing the required homology the assembly this gene was swapped for a BsdR gene. PCR products and digested vector were ligated by combining on ice 20 ρmol of a 1:1 molar ratio of all fragments and vector along with 10 µl of Gibson assembly master mix (2x) and dH2O to 20 µl ending reaction volume. Gibson preparations were incubated in a thermocycler for one hour at 50 °C. Gibson products were then removed and placed on ice until heat shock transformation was able to be performed on DH5a E.coli in a similarly described manner as above.

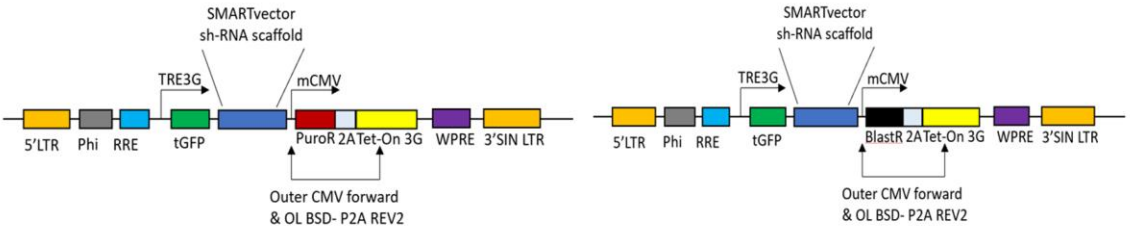


Figure 10 :Dharmacon SMARTvector™ Inducible mCMV Turbo GFP (POLD3-1,2, and 3) Plasmids.

Dharmacon SMARTvector plasmids Pre(a) and Post(b) ligation with PCR product containing BlastR gene from Invivogen.

Gel electrophoresis / Gel Extraction:

All gel electrophoresis were run on lab made gels containing 0.8% agarose and ethidium bromide for one hour at 90-100 volts for 6”x6” gels and 15-20 minutes for 2”x3” gels. For all restriction digests, requiring gel extraction, bands displaying the correct electrophoretic mobility were remove using the back of sterile 1ml pipet tips as punches. These fragments were dissolved and purified via commercial column (Omega-Biotek).

Plasmid Transfection:

Using 406 HeLa cells plated to six well cell culture dishes containing DMEM High Glucose media, 10% NCS and (100 IU/ ml) penicillin/(100ug/ ml)streptomycin antibiotics, cultures were grown to 45% confluence and then transfected first by washing with PBS and then with Opti-MEM, then administration of the transfection complexes as follows: for each well, 2 tubes were labeled A and B. Tube A received 250 µl of Opti-MEM™ (Gibco) reduced serum media plus 1.6 ug of Dharmacon SMARTvector™ Inducible mCMV Turbo GFP (POLD3- 1,2, and 3) with a Blasticidin gene integrated. Tube B received 250 µl Opti-MEM™ (Gibco) plus 5 µl of Lipofectamine 2000™ (Thermo-Fisher® 11668030). Tube A was added to tube B after a 10-minute incubation followed by another 30-minute incubation at room temperature to combine, forming micelles. Aliquots (500 ul) of the DNA-lipid complexes were added to each well and incubated at 37 °C for ~1 hour, 1.5 ml of Opti-MEM was added on top of these complexes gently and left for an additional 8 hours. Complexes were removed after 8 hours and DMEM High Glucose media containing 20% doxycycline/ tetracycline free FBS with no antibiotic was added. Cells were re-plated to 10-centimeter dishes after a 12-hour rest period and 12 hours later selection was performed using 3 µg blasticidin per

ml of media and only 10% Dox/ Tet free FBS. Cells were replated at low confluency in 6 well dishes containing DMEM High Glucose media + 10% Dox/ Tet free FBS and 3 µg/ ml blasticidin to ensure plate purity and left-over night to adhere to the bottom of the plate.

Plasmid Co-Transfection:

All DF2 Myc 406 cell lines (Figure 12) were made by the following protocol utilizing the chromosomally integrated single copy FRT/FLP integration site (Figure 11). This requires co-transfection with the FLP-Recombinase expressing plasmid pOG44 (Stratagene). HeLa 406 acceptor cells were plated to a six well dish using in DMEM High Glucose media and 10% NCS + penicillin/streptomycin. Cultures were grown to 25% confluence. Transfections were carried out as above using 1.8 µg of pOG44 plasmid plus 0.6 µg of dual fluorescent, negatively selectable, microsatellite carrying plasmid maintaining a molar ratio of 1:9 vector of interest to integrate and FLP recombinase expressing pOG44 and 12.5 µl of Lipofectamine 2000™. Twenty-four hours later and at 80 percent confluency, cells were re-plated to a 15 cm dish and given media containing (600 µg/ml) G418 (Fisher®). Antibiotic selection was performed for 10 days or until the control cells (un-transfected) were dead and colonies were isolated to separate 10 cm dishes. Genomic DNA was harvested from cells using the E.Z.N.A™ Tissue DNA Kit (Omega Bio-Tek®). PCR products using this genomic DNA as template and the MCSF/ MCSR primer set were sequenced to ensure correct integration (Primer Table 4). Cell lines currently employed in our lab and manufactured using this methodology are found in Figure 12.

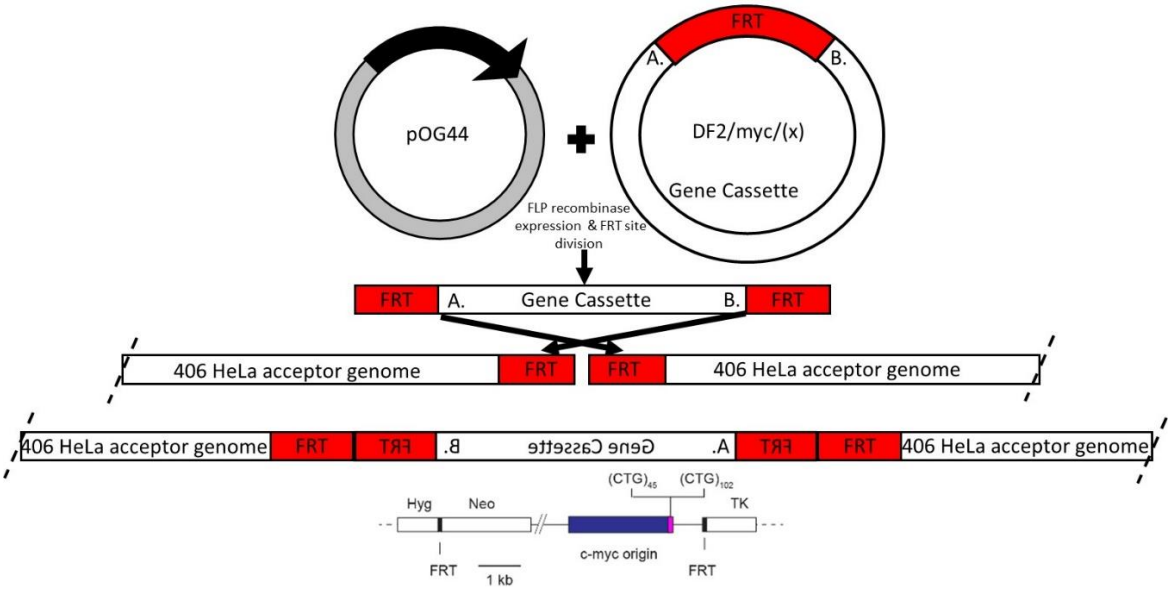


Figure 11: FRT / FLP-Recombinase Mechanism.

Co-transfection and FLP-recombinase expression using single FRT integration site, FLP recombinase localizing to the FRT site, and FLP recombinase dividing and integrating the gene cassette in-between the two FRT sites.

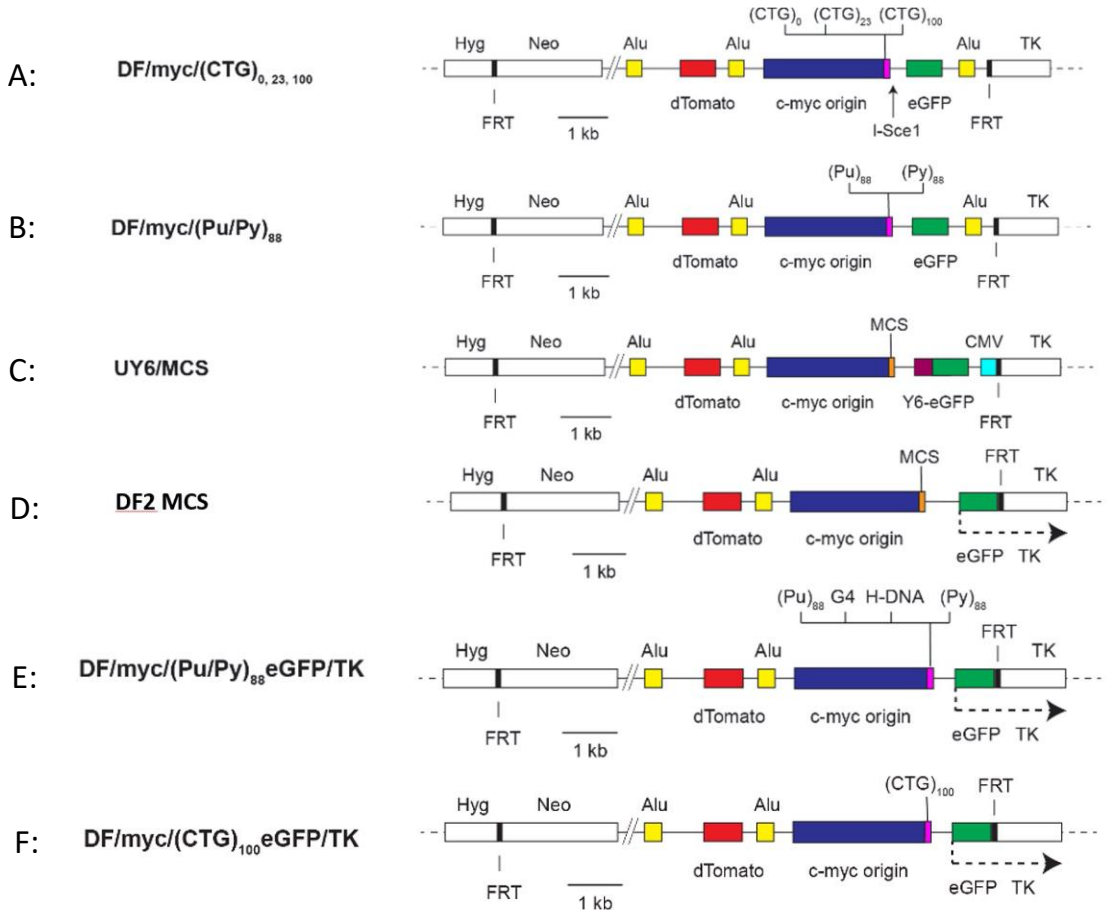


Figure 12: Dual- fluorescent / selectable cell lines.

A: Dual fluorescent cell line containing 3 Alu elements ad used previously to demonstrate (CTG) 102 mer microsatellites instability.

B: Dual fluorescent cell line containing 3 Alu elements ad used previously to demonstrate Pu/Py 88 bp PKD-1 microsatellite instability.

C: Similar to A and B lacking 3rd Alu and possessing MCS and eGFP CMV-TK fusion for constitutive gene expression.

D: Similar to A and B lacking 3rd Alu and possessing MCS and eGFP-TK fusion for constitutive gene expression.

E: Similar to D lacking 3rd Alu and possessing Pu/Py 88 bp PKD-1 microsatellite oriented with Pu or Py repeat in lagging strand orientation and eGFP-TK fusion for constitutive gene expression.

F: Similar to D lacking 3rd Alu and possessing (CTG)102mer and eGFP CMV-TK fusion for constitutive gene expression

Cell Culture:

Unless otherwise specified, standard cell culture was carried out using DMEM High Glucose media (Sigma®) and 10% NCS (BioTechne®) + penicillin/streptomycin (Corning®). Cultures were kept in the dark at 5% CO^2 and 37°C.

Knockdown induction:

Cell lines possessing our negatively selectable dual fluorescent reporter construct and the CTG102mer microsatellite (F, Figure 19) were transfected with SMARTvector™ Inducible mCMV Turbo GFP POLD3-2 BlastR using previously described methods. Cells transfected with or without POLD3-2 Blast R were plated into 6 well dishes containing DMEM High Glucose media (Sigma®) + 10% Dox/ Tet free FBS (R&D systems®) and (5 µg/ml) blasticidin to ensure plate purity of previously transfected cells. Cells were allowed to grow until 15% confluency. 5 mM hydroxyurea (Sigma) and/ or 5 µg/ml doxycycline (Clonetech®) treatments were made in DMEM High Glucose media (Sigma®) + 10% Dox/ Tet free FBS (R&D systems®) every 24 hours for 5 days with doxycycline treatments starting on the first day (120 hr. treatment) and hydroxyurea starting on the third day (72-hour treatment). At 120 hours, cells were allowed to rest in in DMEM High Glucose media and 10% Dox/ Tet free FBS 12 hours before being plated to 15 cm dishes. These cultures were exposed to 5 mg/ ml ganciclovir in DMEM High Glucose media and 10% Dox/ Tet free FBS to select for inactivity of the thymidine kinase gene due to mutations induced from the hydroxyurea treatments. This media was changed every 24 hours for 5 days until cell mortality was observed in control plates.

Western blot:

Western blot analysis was preformed using the following methods. Cells were washed with PBS, harvested using trypsin (Gibco), placed into 15 ml conical tubes and pelleted in DMEM High Glucose media and 10% NCS (BioTechne), to inactive the trypsin. Media was removed by vacuum, cells were resuspended in PBS and transferred to 1.5ml Eppendorf tubes. Suspensions were re-pelleted and washed twice with PBS. Washed cell pellets were subjected Mammalian Protein Extraction Reagent (Thermo-Fisher) and Protease Inhibitor Cocktail (Thermo-Fisher) and sonicated for 30 seconds on ice at 40 kHz. Proteins were quantified using a BCA (Pierce) kit with bovine serum albumin as a standard to generate a curve for quantitation. Samples were measured by absorbance at 562 nm in duplicate on a Tecan Safire 2, sample absorbances were averaged, and did not deviate by more than 5%. Protein electrophoreses was carried out by loading 50 micrograms protein for each sample in a SDS 10% polyacrylamide gel and run over night for 8 hours at 43 volts. This gel was transferred to a PVD membrane (BioRad) at 0.9 amps for 53 minutes. The membrane was divided and blocked by placing the membranes in 50 ml of 5% milk TBST (Tris buffered saline and Tween-20) and allowed to rock for 1 hour. This was replaced with 50 ml 5% milk TBST buffer along with anti-POLD3 (Invitrogen #PAS36951) or anti-GAPDH (Abcam #181602) antibody at a dilution of 1:10,000. These were covered and left to rock on a shaker overnight at 4-6 °C. Blots were given two 15-minute washes in TBST buffer and subjected to their associated HRP conjugated secondary anti-rabbit or anti-goat antibodies from Peirce™ prepared in TBST buffer at a dilution of 1:20,000 and allowed to incubate at 25 °C for one hour.

Secondary antibody was removed followed by two 3 second rinses in TBST buffer. Blots were exposed to ECL2 Western Blotting Substrate (Peirce™ 80196) in the absence of light and left to develop for 15-20 seconds. Blots were examined at 428 nm on an Azure Biosystems C300 to visualize protein transfer using Page Ruler from Thermofisher™ (26617) to monitor migration and size.

Flow cytometry:

Flow cytometry was performed using a BD Accuri C6 flow-cytometer and its associated data analysis program. Dual fluorescent cells carrying the ectopically integrated Pu or Py (TTR/ TTF) 88 bp repeat orientation were harvested as stated above, washed, pelleted, and then resuspended in room temperature PBS. These suspensions were also further prepared by using a cell screen. Dual fluorescent profiles were analyzed with 6.3% compensation for green filter due to spectrum overlap. Gates were set using size forward and side scatter.

Resazurin assay:

Cells either treated with or without doxycycline, hydroxyurea or both trypsinized, resuspended and counted using a trypan blue solution (.4%) at a ratio of 1:1 on a hemocytometer. Densities were calculated and cells replated to DMEM, 10%FBS, and 0.2 mmolar ganciclovir media at 5000 cells per well. Selection was allowed to proceed for 14 days, with DMEM, 10% FBS, and 0.2 mmolar ganciclovir on the seventh day. A resazurin assay was performed on all plates by replacing the media with a 1:10 mixture of resazurin to DMEM and 10% FBS with no antibiotics. These plates were incubated for 2-4 hours at 37°C and 5% CO2. Using a Tecan Safire plate reader absorbance was

measured at 570 nm using 3 read averages. Fluorescence was measured at Ex 550/ Em 600 with adjustments for optimal gain, 40 millisecond integration, using 3 scans at 50 millisecond intervals. This measurement was related to cell density using a standard curve. Controls were used to adjust or normalize the data to background.

DF2 Pu/ Py 88 base pair homopurine homopyrimidine repeat TMS assay:

Cells were grown in a 6 well dish, and treated with 0.5 uM TMS 24 hours after plating, media and TMS was changed every 24 hours for 96 hours. Cells were harvested and genomic DNA was quantified. The i-PCR was carried out following Q5 Polymerase (NEB®) protocols and included a “touchdown” annealing step. A portion of this reaction was run on a 0.8% agarose EtBr gel to ensure expected iPCR products were present. The i-PCR samples were pooled and sent to GeneWiz® for Pac-Bio® long read sequencing

IV. Results

1. What is the effect of POLD3 knockdown on BIR from ectopically integrated (CTG/CAG) 102 microsatellites?

To investigate the possibility of BIR occurring in a DF2 $(CTG)_{102}$ cell line (previously characterized by our laboratory) (Lewis, et al., 2019), an inducible shRNA producing lentiviral vector was employed to knock down the third subunit of Polymerase delta, POLD3. This protein, thought to be essential for break induced replication (Kramara, Malkova, & Osia, 2018), (Kononenko, Ebersole, Vasquez, & Mirkin, 2018) made a perfect target to demonstrate that the types of breaks seen in our cell lines and associated with the $(CTG)_{102}$ microsatellite might be repaired via break induced replication. An assay that was to take place after this one relied on a similar selectable marker present in the lentiviral vector. To ensure compatibility modifications were required pre-transfection that removed a puromycin resistance marker and replaced it with one for blasticidin. These modifications removed a selectable marker for mammalian puromycin resistance from this vector and replaced it with one for blasticidin resistance (Figure 9). These plasmids were used in a DNA damage/ mutagenesis assay to demonstrate our system's ability to detect break induced replication. The replication assay that we designed was to culminate with a photometric resazurin assay for cell survival and provide quantitation of mutagenesis post selection with ganciclovir. To that end three SMARTvector™ Inducible mCMV Turbo GFP plasmids from Dharmacon® capable of producing shRNAs against different regions of the Polymerase deltas 3rd subunit were purchased as E.coli stocks from Dharmacon®. This commercially available lentiviral system is capable of integrating into a host genome and expresses an shRNA upon treatment with

doxycycline. As stated above this vector also contained a puromycin gene for mammalian cell selection post transfection that was incompatible with upcoming assays. This gene was replaced with a blasticidin resistance gene using restriction digestion, standard PCR, and Gibson assembly methods (Figure 10).

Purified SMARTvector™ Inducible mCMV Turbo GFP interchangeably called POLD3 – (1, 2, or 3) each producing a different shRNA against the POLD3 subunit were used as template in concert with a suite of primers (Primer table 1) designed to produce the necessary homologies required for the Gibson assembly methods 5' exonuclease activity and subsequent extension and ligation steps. Vector from Invitrogen™ containing a blasticidin resistance marker was used as template generating the replacement resistance fragment. Bacteria containing original Dharmacon SMARTvector™ Inducible mCMV Turbo GFP (POLD3- 1,2, and 3, PuroR) plasmids were cultured, purified and plasmids sequenced by Retrogen. This sequence was used to design primers needed to build a gene cassette containing a blasticidin resistance gene instead of Puromycin resistance gene to be swapped in using homology from a pair of Kpn1 restriction sites present in the plasmid. In a stepwise manner, multiple fragments were generated via Takara Bio GXL PCR and primer sets containing a homology to the blasticidin resistance marker in place of the puromycin resistance marker as well as the backbone vector. The resulting ends of the assembled cassette also retained two Kpn1 restriction sites. This allowed for exonucleolytic activity and subsequent assembly by the Gibson assembly method (Gibson, et al., 2009), as well as ligation of this cassette into further assemblies (Figures 13,14). Following final vector preparation, restriction digestion confirmed vector size and cassette integration using a Pst1 digest. This restriction digest of the new plasmid

was predicted to yield one less fragment than the old one due to the puromycin resistance gene containing a Pst1 site and blasticidin resistance gene lacking one. This results in 4 bands for puromycin resistance containing vector and 3 bands for blasticidin resistance containing vector (Figure 15).

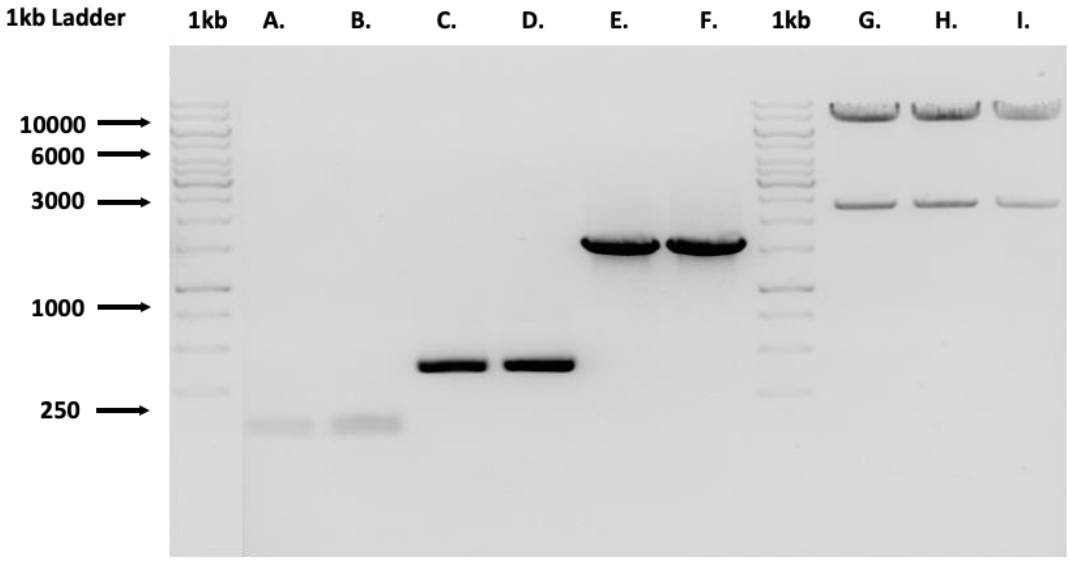


Figure 13: Cloning of BSD gene and POLD3-2 vector preparation

PCR products from POLD3-2 and BSD templates PCR'd in duplicate. Lanes: A. , B. , C. , D. , E. &, F. , PCR products. G. ,H. ,and I. 2.5ug digestion of POLD3-2 PuroR using Kpn1.

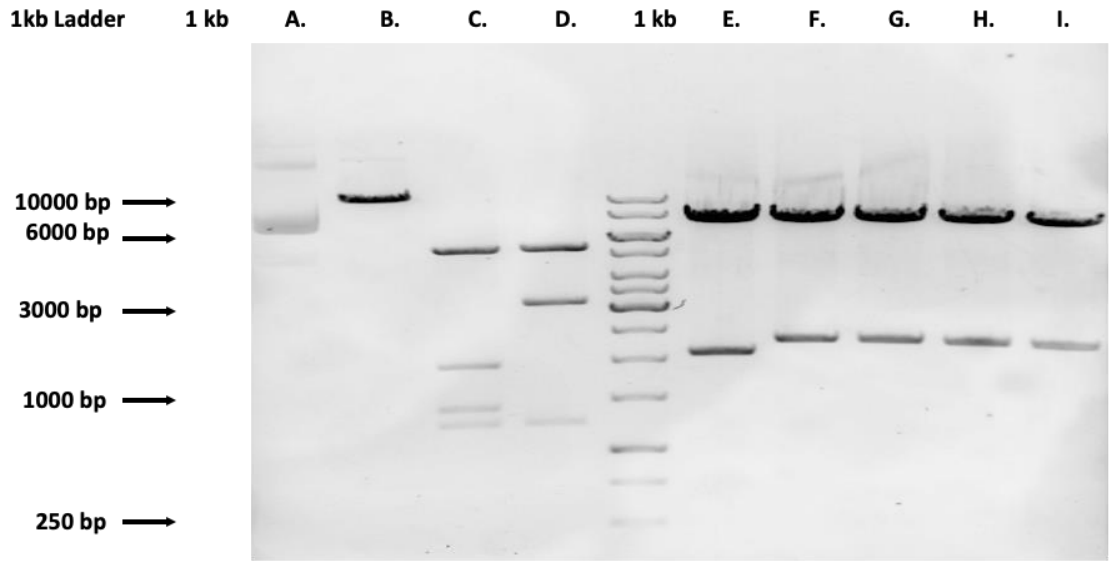


Figure 14: Cloning of BSDR gene and POLD3- 1,and 3 Vector Prep.

Restriction digestions and subsequently ligated products for further cloning. Lanes: A. Undigested POLD3-2, B. POLD3-2 digest with Pvu1(linear), C. POLD3-2 PuroR digest with Pst1, D. POLD3-2 BlastR digested with Pst1, E. POLD3-2 BlastR digested with Kpn1, F.&G. POLD3-1 PuroR digested with Kpn1, H.&I. POLD3-3 PuroR digested with Kpn1.

*BsdR gene (423bp) > PuroR (602bp), New gene cassette 179 bp shorter).

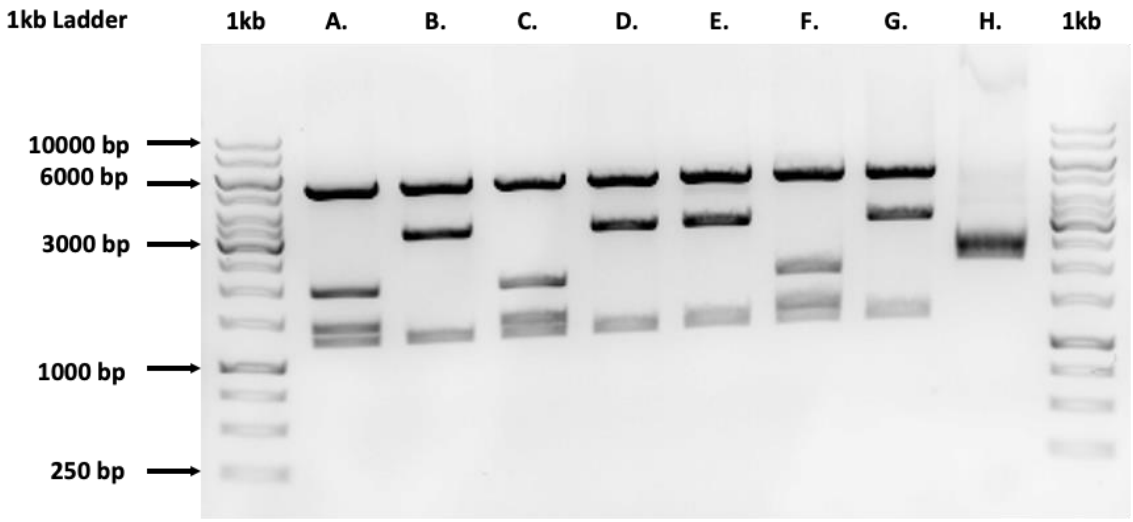


Figure 15: Confirmation of BsdR gene integration via restriction digestion.

Restriction digest to confirm correct integration using Pst1 restriction digest. Lanes: A. POLD3-2 PuroR, B. POLD3-2 BsdR, C. POLD3-1 PuroR, D. POLD3-1 BsdR, E. POLD3-1 BsdR, F. POLD3-3 PuroR, G. POLD3-3 BsdR, H. POLD3-3 BsdR (Failure/Scrapped)

These products were sequenced and subsequently transfected into 406 HeLa cells for knockdown efficiency testing via western blot analysis using anti-POLD3 (Invitrogen #PAS36951) (primary), anti-GAPDH (Abcam #181602) (control), and HRP conjugated secondary anti-rabbit or anti-goat antibodies (Figure 16).

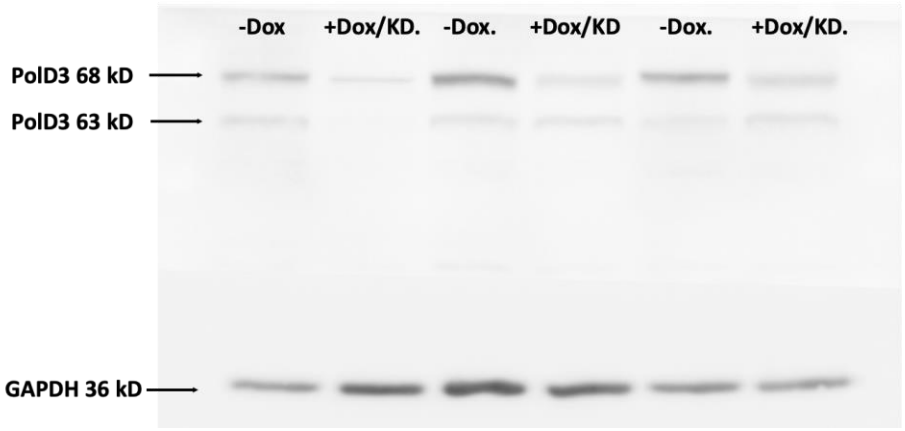
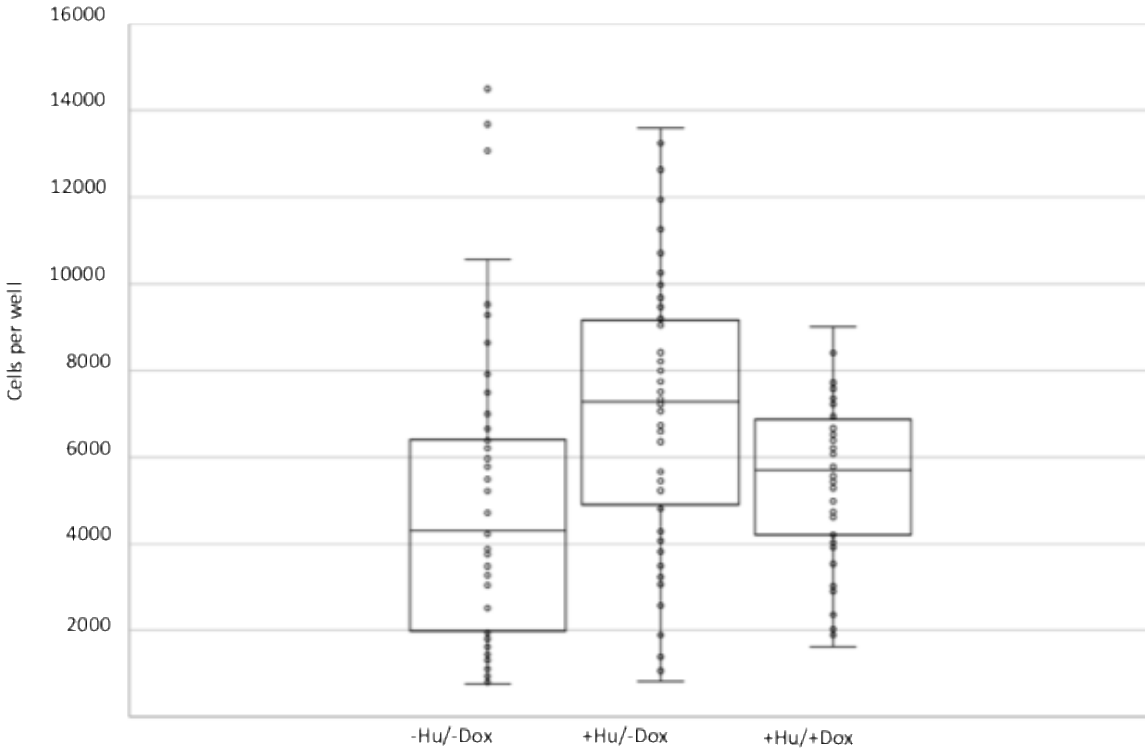


Figure 16: POLD3 Knockdown in 406 HeLa acceptor cell lines.

Knockdown conditions: Cells were incubated in media containing 2 ug/ml Doxycycline, 10% Dox/Tet free FBS, and DMEM for 120 hours with media changed every 24. POLD3 is expressed as two iso-forms, p63 and p68 (50). These were both detected by anti-POLD3 (Invitrogen #PAS36951) at a 1:10,000 dilution. Anti-GAPDH (Abcam #181602) at a 1:10,000 dilution was used as a loading control due to Actin possessing a similar electrophoretic mobility.

Using one of these shRNA producing vectors above (Figure 16), a DF2 (CTG)102 mer cell line (Figure 12a) and methods previously demonstrated to cause double strand breaks we induced replication stress. These breaks were hypothesized to be repaired using homology mediated repair pathways due to microsatellite homology and the single-ended double strand break that these microsatellites can produce (Leffak, 2017). The repair of this break caused mutations in the negatively selectable thymidine kinase gene downstream of a microsatellite known to cause breaks, these mutations increased survivability during exposure to ganciclovir due to inactivation of the TK gene. When exposed to hydroxyurea and POLD3-shRNA, survivability was reduced with respect to control levels (Figure 17). This was quantitated using resazurin which is reduced to resorufin by aerobic respiration of metabolically active cells and can be used as an indicator of cell viability. With this we can conclude that if BIR is as mutagenic as others have demonstrated (Deem, et al., 2011), and is dependent on POLD3 (Kononenko, Ebersole, Vasquez, & Mirkin, 2018), then it must be playing a role in the repair of double strand breaks being caused by the presence of the CTG repeat and hydroxyurea exposure. This is due to BIR also requiring the same form of break for strand invasion (single ended double strand) that we are causing using this cell line. This also agrees with previously published findings that demonstrate BIR's requirement for POLD3 (Kononenko, Ebersole, Vasquez, & Mirkin, 2018).



F-Test Two-Sample for Variances		
	BIR1701POLD3HU	*BIR1701POLD3HUDOX*
Mean	7042.534063	5523.365749
Variance	9152942.094	2917713.153
Observations	60	60
df	59	59
F	3.137026025	
P(F<=f) one-tail	1.01434E-05	
F Critical one-tail	1.539956607	

t-Test: Two-Sample Assuming Unequal Variances		
	BIR1701POLD3HU	*BIR1701POLD3HUDOX*
Mean	7042.534063	5523.365749
Variance	9152942.094	2917713.153
Observations	60	60
Hypothesized Mean Difference	0	
df	93	
t Stat	3.387007012	
P(T<=t) one-tail	0.000518553	
t Critical one-tail	1.661403674	
P(T<=t) two-tail	0.001037106	
t Critical two-tail	1.985801814	

Figure 17: Cell Survivability via 96 Well Resazurin Assay.

Mean cell survival per well post knockdown, hydroxyurea treatment and ganciclovir exposure measured using 96 well resazurin assay demonstrating reliance on POLD3 for ganciclovir resistance.

2. To further examine the effect of replication polarity on the stability of a homopurine/ homopyrimidine microsatellite, cells were created for future assays containing a homopurine/homopyrimidine 88 bp (Pu/ Py) mirror repeat.

After demonstrating our selectable cell lines value in the analysis of microsatellite instability and implicating the $(CTG)_{102}$ microsatellite in BIR, a mirror repeat was integrated into a DF2 vector similar to the one in the previous assay (Figure 12a) (Figure 17). This helped set up future investigation of other microsatellites and error-prone repair mechanisms such as BIR. It was decided to clone the previously used PKD1 88 bp repeat found in intron 19 from on chromosome 16. This homopurine/homopyrimidine (Pu/ Py) mirror repeat was cloned from a cell line previously shown to display mutations under replication stress (Gadgil, et al., 2020) and replaced the $(CTG)_{102}$ microsatellite in our selectable DF2 plasmid construct (Figure 12 d & e). Plasmids were made using standard PCR, sequential digestion and ligation. These cells were made utilizing a 406-acceptor cell line containing an FRT site compatible with the FRT site in the plasmid described previously and utilizing FLP recombinase in a method similar to Cre-Lox-P integration as outlined in (Figure 11). Plasmids made for a previous DNA damage experiment (Gadgil, et al., 2020) that contained the Pu/ Py 88 bp repeat present in the PKD1 gene were used as template in a PCR reaction utilizing. Primers were designed that would allow for a sequential double digestion of the PCR product as well as vector backbone and ligation into a multiple cloning site using Swa1 and Pac1 (Figure 18) Plasmid integration was tested in multiple colony PCR reactions which demonstrated integration of correctly sized cloned fragment (Figure 19). Followed by PCR using diagnostic primers demonstrating correct fragment orientation (Figure 16, 17).

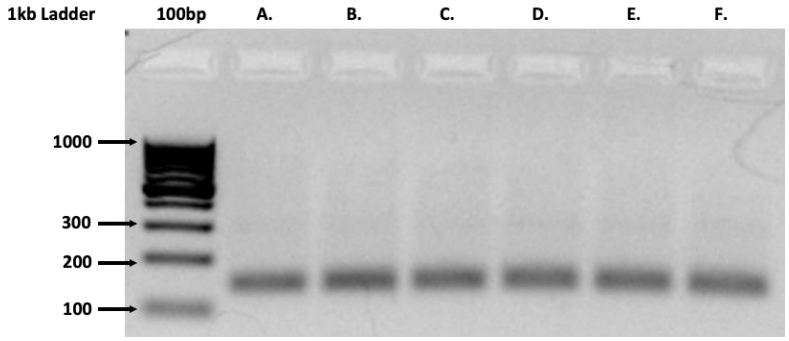


Figure 18: TTF /TTR Cloning using Pac1 Swa1 primer sets.

Lanes A. through F; Pac1 and Swa1 digested GXL PCR products of DF/ Myc (Pu/ Py) 88bp "TTF" plasmid using primers generating the needed homology to utilize the Swa1 and Pac1 sites present in the DF2 MCS plasmid and reorient the 88bp Pu repeat into the opposite strand forming a "TTR" cell line.

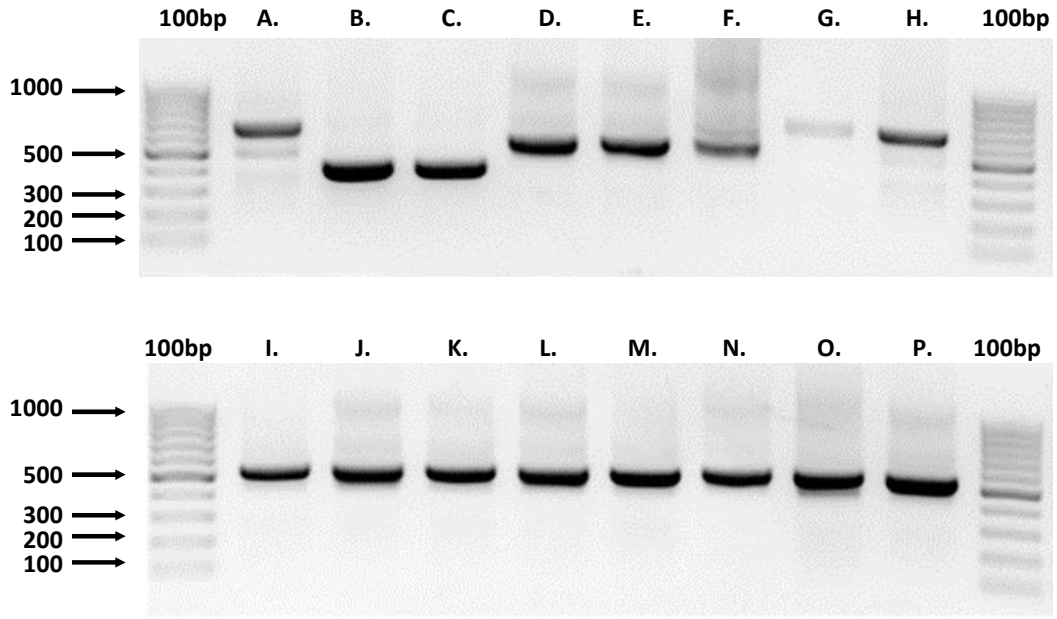


Figure 19: DF2 / MCS Integration testing using MCS Primers

PCR using primers designed to produce products containing sequence from in between the multiple cloning site.

Lanes: A. DF2/MCS/CTG100mer, B. UY6 MCS plasmid containing no 88bp Py/ Pu mirror repeat, C. DF2 MCS plasmid containing no 88bp Py/ Pu mirror repeat, D. through P. are plasmids containing either the TTF or TTR configuration of the 88bp Py/ Pu mirror repeat.

To ensure correct Pu/ Py microsatellite orientation a set of primers was employed that would produce products dependent on orientation. When placing the cloned poly-purine repeat in the forward or leading strand the primers would be oriented in the correct direction producing a product. When flipped to place the poly-pyrimidine side of the mirror repeat in the leading strand they will not. As demonstrated in Figure 20, this was the case and positive integration of both the TTF (Pu) leading and TTR (Py) leading strand orientations were later confirmed by Sanger sequencing. These plasmids were successfully transfected into 406 acceptor cells using the standard FRT/FLP recombination protocols requiring co-transfection with the FLP-Recombinase expressing plasmid pOG44 as described above and in Figure 11. Genomic DNA was prepped from these cell cultures then tested via PCR and Sanger sequencing to insure correct genomic integration. These cells were examined by flowcytometry on a BD Acuri C6 and found to demonstrate a yellow dual fluorescent flow pattern consistent with what we would expect when both eGFP and dTomato are both intact and being constitutively expressed as seen in Figure 21 and 22.

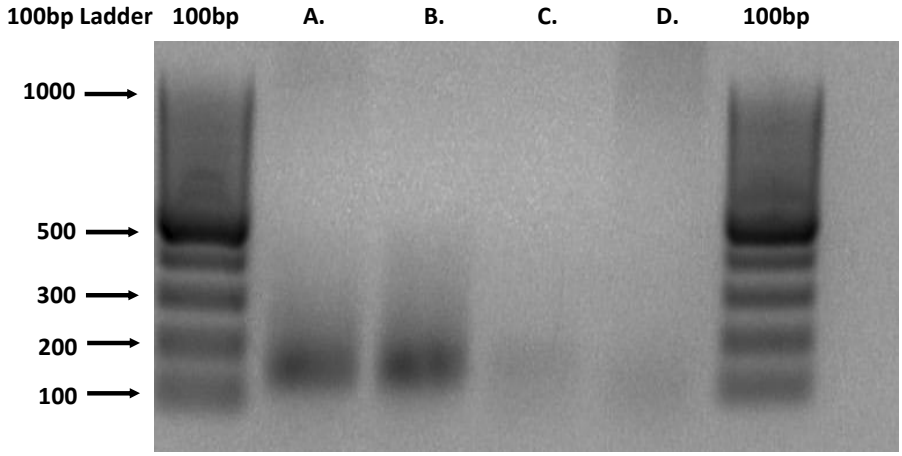


Figure 20: DF2 Integrant Orientation Conformation PCR

PCR of DF2 MCS 88bp Py/Pu mirror repeat Plasmids in "TTF and "TTR" orientation using diagnostic primers designed to give a product or not based on orientation/direction of the cloned fragment. Lanes: A. and B., two separate cloned "TTF"DF2 MCS plasmids; C. and D., Two different cloned "TTR" DF2 MCS plasmids.

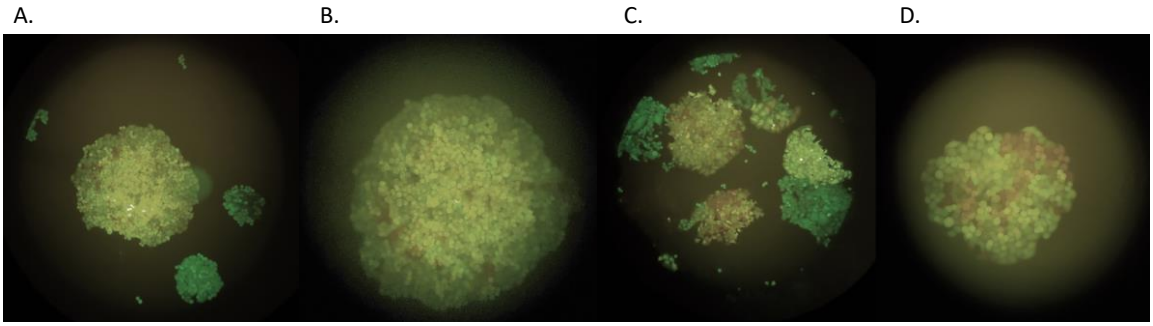


Figure 21: Fluorescent microscopy of "TTF" and "TTR" (Pu/Py) 88pb mirror repeat integrated DF2/Myc/CMV/MCS cell lines

Panel: A. and B. demonstrating "TTF" cell lines showing a dual fluorescent "yellow" pattern, C. and D. demonstrating "TTR" cell lines "yellow" dual fluorescent profile.

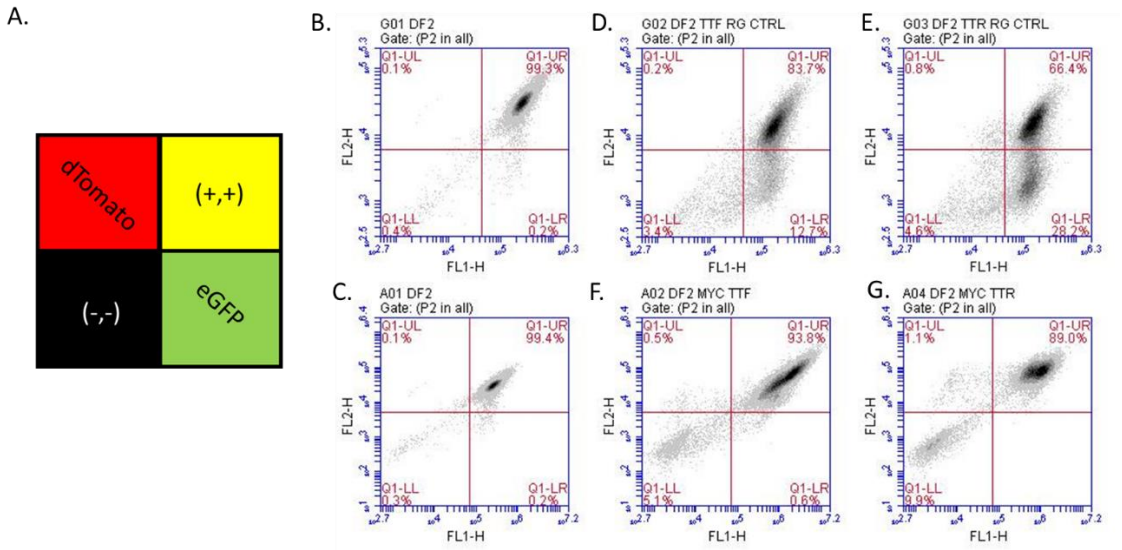


Figure 22: Flowcytometry of "TTF" and "TTR" (Pu/Py) 88bp mirror repeat integrated DF2/Myc/CMV/MCS cell lines

Panel A: corresponding color diagram, Panel B&C: DF2 cell line with no repeat present, Panel D: "DF2 TTF" cell line with demonstrating dual fluorescent profile, Panel E: "DF2 TTR" cell line demonstrating dual fluorescent profile, Panel F: (New)DF2 Myc TTF, G: (New) DF2 Myc TTR.

3. What is the pattern of mutagenesis of the homopurine/homopyrimidine (Pu/Py) G4 forming mirror repeat under replication stress?

Cell lines previously demonstrating telomestatin (TMS) sensitivity and containing an 88 bp Pu repeat in the leading strand for synthesis were used in a DNA damage assay and after treatment with TMS, i-PCR was carried out on their genomic samples. Our previous publications show that the presence of this repeat in the lagging strand orientation can lead to translocations, and these translocations can be detected using a form of genomic library prep known as inverse PCR (Barthelemy, Hanenberg, & Leffak, 2016). Inverse PCR detects translocations by using frequently cutting restriction enzymes and known sequences present on the invading strand between these cut sites as primer targets. These primers face outward and when digested products are ligated together post dilution these primers are then able to generate a product that can be sequenced (Figure 22).

Cells treated with 0.5 uM TMS were harvested and genomic DNA was quantified. The I-PCR was carried out following Q5 Polymerase (NEB®) protocols including a "touchdown" annealing step. A portion of this reaction was run on a 0.8% agarose EtBr gel to ensure expected iPCR products were present (Figure 24). The i-PCR samples were pooled and sent to GeneWiz® for Pac-Bio® long read sequencing.

This was also accompanied by traditional PCR (Figure 24) of this template to send the full length dTomato gene and the full length eGFP gene for sequencing as well. Alignments were performed on these long single molecule sequence reads using NGMLR, a of Burrows - Wheeler Aligner called BWA-MEM, Sniffles and our predicted seq as reference in place of a genome. Together these tools were able to help us assess read depth, coverage, alignment, mutation rates and locations, as well as to identify

possible translocations. BWA-MEM paired with NGMLR makes for a fast and accurate alignment software package for mapping low-divergent sequences against a large reference genome and is recommended for high-quality queries as it and is more robust in addressing data generated from Illumina sequencing based on a previous short-read aligner NGM extended with a new convex gap-cost scoring model to align long-reads across Structural Variant breakpoints (Li & Durbin, 2010) . This also uses Sniffles and successively scans the alignments to identify all types of SVs employing a novel SV scoring scheme to exclude false SVs based on the size, position, type and coverage of the candidate SV (Sedlazeck, et al., 2018). Additional data analysis for all experiments used Excel for simple statistical analysis and rudimentary alignment. By doing so we were able to assess mutations upstream and downstream and locally to the Pu/Py 88 bp microsatellite. It was concluded that there was a significantly higher rate of mutation in dTomato containing PCR products than in eGFP PCR products as well as the i-PCR product that contained the microsatellite (Figure 25). Analysis also demonstrated that there was an expansion or contraction of 1-8bp occurring in >50% of the reads reads covering a poly thymidine repeat located at the beginning of an Alu site located in the dTomato PCR product (Figure 26) This might have artificially elevated the incidence of mutation as Alu elements have been shown to be prone to mutations such as these.

Further analysis was performed on these long iPCR sequence reads generated by Pac-Bio sequencing using a hybrid genome incorporating our ectopic site derived circles iPCR products sequence as well as the human genome Grch-38 as a reference.

Sequence alignment was analyzed via NGMLR and Sniffles was employed to search for structural variants. Several structural variants were found including multiple hits in the

gene DENND3. On its own DENND3 is not associated with diseases attributed to microsatellite instability this result does however demonstrate that microsatellites might be cause translocations globally with respect to the genome as this translocation is non-allelic due to our construct's integration site being on the 18th Chromosome and not the 8th where this DENND3 translocation was found.

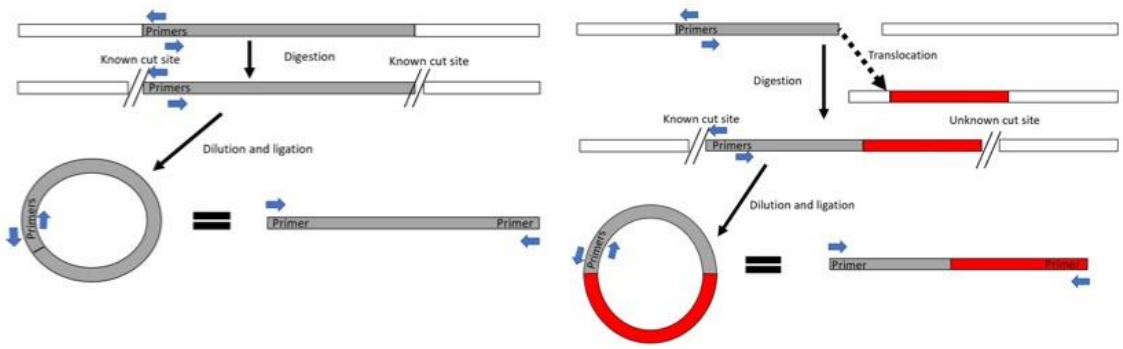


Figure 23: Inverse PCR (iPCR) Mechanism

When template (plasmid or genomic DNA) is ligated together after digestion and dilution primers once designed to face away from each other and back to back now face towards each other as seen here in the circularized product. This circularization is due to a higher probability of intra-strand ligation at a high enough dilution. If translocations occur, sequence might be found between these two primers due to the repetitive nature of the restriction enzyme chosen and because no restriction site is found in between the two primers post ligation.

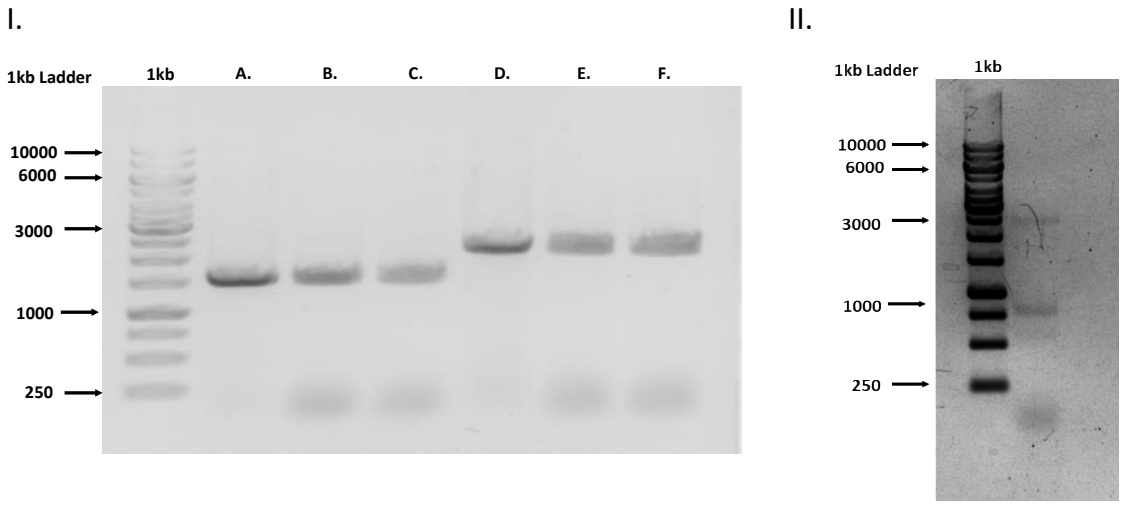


Figure 24: Inverse, eGFP or dTomato Touch down PCR:

Touch down PCR of "TTR" plasmids and Cell line containing the 88 bp poly Pyrimidine mirror repeat in the lagging strand which have been shown to break and accumulate mutations under replicative stress. Products containing eGFP gene containing eGFP and the LTR with a predicted size of 1.5 kb: A. TTR plasmid, B. TTR CTRL Genomic DNA, C. TTR + TMS treatment Genomic DNA, Products containing hpgk Promoter, dTomato gene, and 1 Alu element with a predicted size of 2.2 kb: D. TTR plasmid, E. TTR CTRL Genomic DNA, F. TTR + TMS treatment Genomic DNA.

II. i-PCR of DF2 MYC TTR cell line containing 3 ALU elements treated with TMS previously demonstrating instability with TMS and HU treatment.

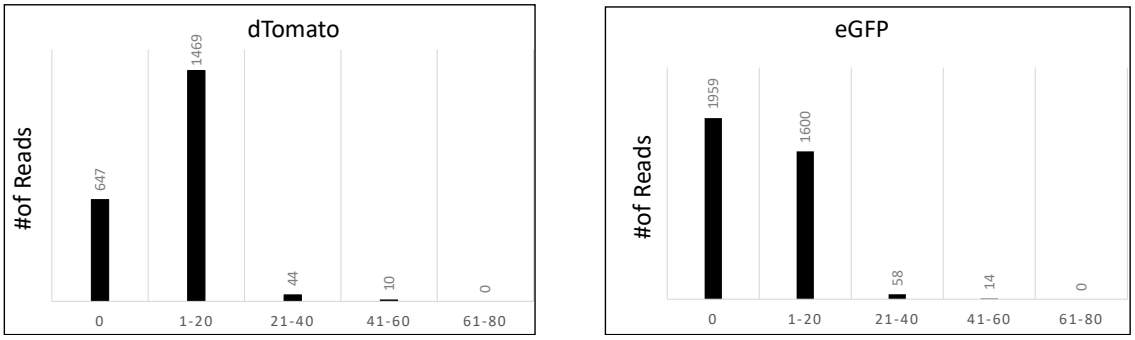


Figure 25 : Mutation rates in largest reads from dTomato and eGFP PCR products

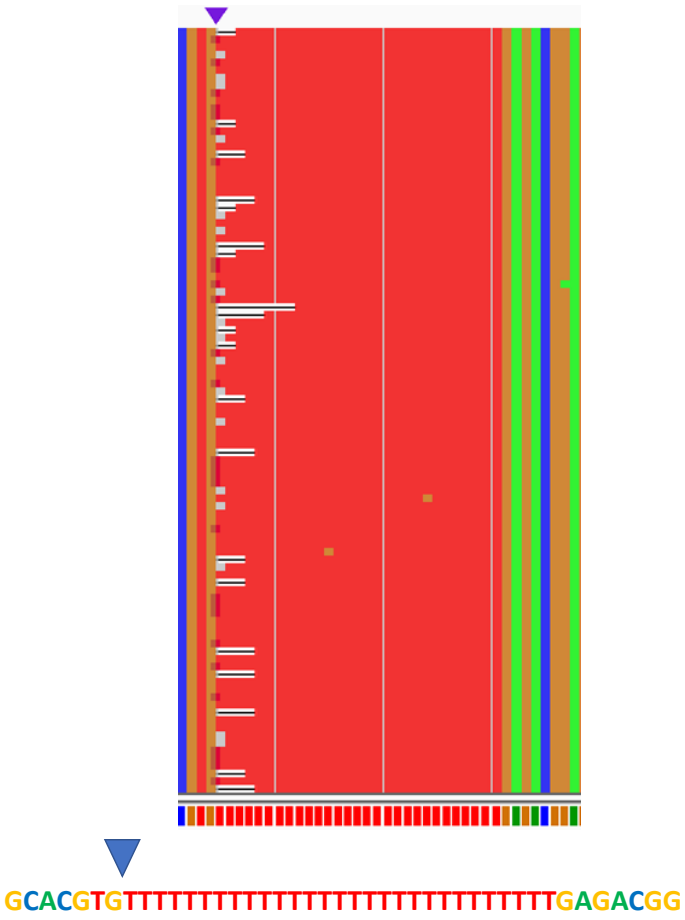


Figure 26: Expansions and contractions at ALU microsatellite.

Off target pol-pyrimidine tract found int the beginning of an ALU element displaying expansions and contractions 243 bp downstream of the dTomato fluorescent reporter and 2598 bp up stream of the PKD1 insert. Deletions varied from 1 to 8 bp while expansions varied anywhere from 2 to 7 bp, reads possessing these mutations comprised 51/100 reads surveyed. Of these reads 28 Poly thymidine expansions and 23 thymidine deletions. This is in contrast to sequencing data of the eGFP gene, possessing no Alu element this sequence only contained 11 insertions at a random distribution of locations throughout the reference sequence the largest of which was 4 bp in length.

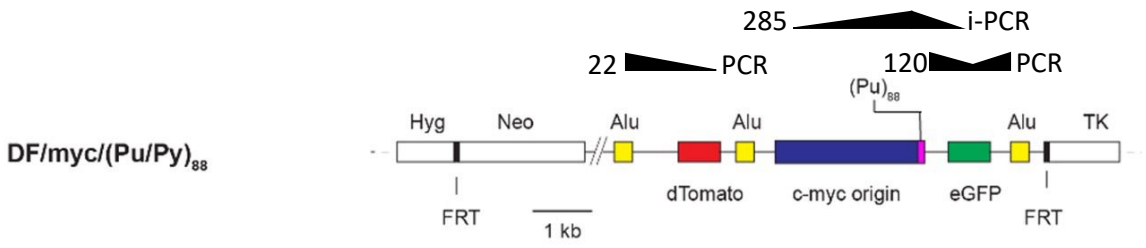


Figure 27: Individual Mismatch number and relative locations present in eGFP and dTomato and iPCR products using TMS treated DF2 Myc TTR treated cells.

V. Discussion

With these results we can concluded that mutations we see under replication accumulate in a thymidine kinase gene downstream of the ectopic (CTG)102 microsatellite in the presence of replication stress (Hu) and lead to ganciclovir resistance in our cell line. This means that by knocking down polymerase deltas 3rd subunit we are abrogating mutagenic potential of BIR thus decreasing cell survivability in the presence of our selectable resistance marker. Without break induced replication and it's mutagenic effects, cells do not accumulate enough mutations in the selectable TK gene. These cells retain the ability to metabolize ganciclovir and will die or senesce in its presence due to fork rescue by another method and proper thymidine kinase expression.

To determine the effect of replication polarity on the stability of a homopurine/ homopyrimidine microsatellite cell lines were created for further examination of the PKD1 homopurine/homopyrimidine (Pu/Py) G4 forming mirror repeat. This 88 bp repeat was placed with either the purine (Pu/ TTR) or pyrimidine (Pu/ TTF) rich strand in the lagging strand for synthesis. It was also shown that these cell lines were stably integrated and capable of retaining this dual fluorescent profile for some time.

While both demonstrate the dual fluorescent profile required for upcoming assays increased instability under endogenous replication stress was observed when the Pu 88 pb repeat (TTR) is found in the lagging strand replication template and either passaged for many cycles or transfected without florescent colony selection which is consistent with findings from (Liu, et al., 2012). Lastly, after assaying cell lines previously created for the study of the PKD1 Pu/Py G4 mirror repeat we found a pattern of mutagenesis

attributable to the homopurine/homopyrimidine PKD1 (Pu/Py) G4 forming mirror repeat in it's Purine rich lagging strand orientation (Figure 25, 26, 27). PacBio long read sequencing demonstrated translocations do accumulate after treatment with replication stressors such as telomestatin and while the locations don't specifically correlate to a disease per se one can conclude that indeed the mere ability to translocate, synapse and form gene fusions is deleterious to genome stability. A poly-pyrimidine tract found at the beginning of an Alu element also displayed expansions and contractions 243 bp down-stream of the dTomato fluorescent reporter and 2598 bp up stream of the PKD1 insert. Deletions varied from 1 to 8 bp while expansions varied anywhere from 2 to 7 bp, reads possessing these mutations comprised 51/100 reads surveyed. Of these reads 28 Poly thymidine expansions and 23 thymidine deletions (Figure 25). These mutations were not seen in the sequence containing eGFP. This eGFP sequence only contained 11 insertions at a random distribution of locations throughout the reference sequence the largest of which was 4 bp in length. This difference in mutation rate can be explained by the presence of the Alu element in the dTomato PCR product and not the eGFP PCR product. After a more extensive alignment was preformed this was noticeable visually by a greater density of mutations being found in the eGFP products with respect to the dTomato product excluding the Alu site. Further work should be performed on this data using a hybrid reference genome that excludes this Alu element when only assessing point mutations, expansion and contractions. This along with previously generated data might be used to further develop patterns of mutagenesis that can then be attributed to diseases associated with microsatellites and replication instability.

VI. References

Annelies de Klein, A. G. (1982). A cellular oncogene is translocated to the Philadelphia chromosome in chronic myelocytic leukaemia. *Nature*, 765-767.

Anoopkumar-Dukie, S., Carey, J. B., Conere, T., O'Sullivan, E., van Pelt, F. N., & Allshire, A. (2014). Resazurin assay of radiation response in cultured cells. *The British Journal of Radiology*.

Barthelemy, J., Hanenberg, H., & Leffak, M. (2016). FANCJ is essential to maintain microsatellite structure genome-wide during replication stress. *Nucleic Acids Research*, 6803-6816.

Beadle, G. W., & Tatum, E. L. (1941). Genetic Control of Biochemical Reactions in Neurospora. *PNAS*, 499-506.

Boboila, C. J. (2010). Alternative end-joining catalyzes robust IgH locus deletions and translocations in the combined absance of ligase 4 and Ku 70. *Proceedings of the National Academy of Sciences of the United States of America*, 3034-3039.

Carvalho, C. M., & Lupski, J. R. (2016). Mechanisims underlying structural variant formation in genomic disorders. *Nature Reviews Genetics*, 224-238.

Cooper-Roth, T. (2010). The Effects Thalidomide Embryonic Development. *The Embryo Project Encyclopedia*, Acessed 2021.

Deem, A., Keszthelyi, A., Blackgrove, T., Vayl, A., Coffey, B., Mathur, R., . . . Malkova, A. (2011). Break-induced replication is highly inaccurate. *PLOS Biology*.

Drucker, L., Uziel, O., Tohami, T., Shapiro, H., Radnay, J., Yarkoni, S., . . . Lishner, M. (2003). Thalidomide Down-Regulates Transcript Levels of GC-Rich Promoter Genes in Multiple Myeloma. *Molecular Pharmacology*, 415-420.

Filippo, J. S., Sung, P., & Klein, H. (2008). Mechanism of Eukaryotic Homologous Recombination. *Annual Review of Biochemistry*, 229-257.

Gadgil, R. Y., Romer, E. J., Goodman, C. C., Shin-ya, K., Hanenberg, H., & Leffak, M. (2020). Replication stress at microsatellites causes DNA double-strand breaks and break-induced replication. *Journal of Biological Chemistry*, 15378-15397.

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA Molecules up to several hundred kilobases. *Nature Methods*, 343-345.

Hader, D.-P., & Sinha, R. P. (2002). UV-induced DNA damage and repair: a review. *Photochemical & Photobiological Sciences*, 225-236.

Her, J., & Bunting, S. (2018). How cells ensure correct repair of DNA double-strand breaks. *Journal of Biological Chemistry*, 10502-10511.

Kanno, S.-i., Kuzoka, H., Sasao, S., Hong, Z., Lan, L., Nakajima, S., & Yasui, A. (2007). A novel human AP endonuclease with conserved zinc-finger-like motifs involved in DNA strand break responses. *EMBO*, 2094-2103.

Kim, J. C., & Mirkin, S. M. (2013). The balancing act of DNA repeat expansions. *Current Opinions in Genetics & Development*, 280-288.

Kim, M.-Y., Vankayalapati, H., Shin-ya, K., Wierzba, K., & Hurley, L. H. (2002). Telomestatin, a Potent Telomerase Inhibitor That Interacts Quite Specifically with the Human Telomeric Intramolecular G-Quadruplex. *Journal of the American Chemical Society*, 2098-2099.

Kononenko, A. V., Ebersole, T., Vasquez, K. M., & Mirkin, S. M. (2018). Mechanisms of genetic instability caused by (CGG)n repeats in an experimental mammalian system. *Nature Structural & Moleculat Biology*, 669-676.

Kramara, J., Malkova, A., & Osia, B. (2018). Break-Induced Replication: The Where, The Why, and The How. *Trends in Genetics*, 518-531.

Leffak, M. (2017). Break-induced replication links microsatellite expansion to complex genome rearrangements. *BioEssays*.

Lewis, T. W., Barthelemy, J. R., Virts, E. L., Kennedy, F. M., Gadgil, R. Y., Wiek, C., . . . Leffak, M. (2019). Deficiency of the Fanconi anemia E2 ubiqitin conjugase UBE2T only partially abrogates Alu-mediated recombination in a new model of homology dependent recombination. *Nucleic Acids Research*, 3503-3520.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler Transform. *Bioinformatics*.

Liu, G., Chen, X., Bissler, J. J., Siden, R. R., & Leffak, M. (2010). Replication-dependent instability at (CTG)•(CAG) repeat hairpins in human cells. *Nature Chemical Biology*, 652-659.

Liu, G., Myers, S., Chen, X., Bissler, J. J., Sinden, R. R., & Leffak, M. (2012). Replication Fork Stalling and Checkpoint Activation by a PKD1 Locus Mirror Repeat Polypurine-Polypyrimidine (Pu-Py) Tract. *Journal of Biological Chemistry*, 33412-33423.

Lu, G., Duan, J., Shu, S., Wang, X., Gao, L., Guo, J., & Zhang, Y. (2016). Ligase I and ligase III mediate the DNA double-strand break ligation in alternative end-joining. *Proceedings of the National Academy of Scinces of the United States of America*, 1256-1260.

Malkova, A., Naylor, M., Yamaguchi, M., Ira, G., & Haber, J. E. (2020). RAD51-dependent break-induced replication differs in kinetics and checkpoint responses from RAD51-mediated gene conversion. *Molecular and Cellular Biology*.

Mason, P. A., & Cox, L. S. (2012). The role of DNA exonucleases in protecting genome stability and their impact on ageing. *AEG*, 1317-1340.

Mateos-Gomez, P. A.-D. (2015). Mammalian polymerase theta promotes alternative NHEJ and suppresses recombination. *Nature*, 254-257.

Mirkin, S. M. (2007). Expandable DNA repeats and human disease. *Nature*, 932-940.

Mladenova, V., Mladenov, E., & Iliakis, G. (2019). Necessities in the Processing of DNA Double Strand Breaks and Their Effects on Genomic Instability and Cancer. *Cancers*, 1671.

Nelson, D. L., & Cox, M. M. (2017). *Lehninger's Principles of Biochemisty 7th Edition.* New York: W.H. Freeman Company.

Nowell, P. C., & Hungerford, D. A. (1960). Chromosome studies on normal and leukemic human leukocytes. *Journal of the National Cancer Institute*, 85-109.

Pannunzio, N. R., Watanabe, G., & Lieber, M. R. (2018). Nonhomologous DNA end-joining for repair of DNA double-strand breaks. *Journal of Biological Chemistry*, 10512-10523.

Rocha, C. R., Silva, M. M., Quinet, A., Cabaral-Neto, J. B., & Menck, C. F. (2018). DNA repair pathways and cisplatin resistance: an intimate relationship. *Clinics*.

Roots, R., Kraft, G., & Gosschalk, E. (1985). The formation of radiation-induced dna breaks: The ratio of double-strand breaks to single-strand breaks. *International Journal of Radiation Oncology, Biology, and Physics*, 259-265.

Rubnitz, J., & Subramani, S. (2021). The minimum amount of homology required for homologous recombination in mammalian cells. *American Socitey for Microbiology*.

Ruff, P., Donnianni, R. A., Glancy, E., Oh, J., & Symington, L. S. (2016). RPA Stabilization of Single-Stranded DNA Is Critical for Break-Induced Replication. *Cell Reports*, 3359-3368.

Saini, N., Ramakrishnan, S., Elango, R., Ayyar, S., Zhang, Y., Deem, A., . . . Malkova, A. (2013). Migrating bubble during break-induced replication drives conservative DNA synthisis. *Nature*, 389-392.

Sallmyr, A., & Tomkinson, A. E. (2018). Repair of DNA double-strand breaks by mammalian alternative end-joining pathways. *Journal of Biological Chemistry*, 10536-10546.

Sedlazeck, F. j., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., Haeseler, A., & Schatz, M. C. (2018). Accurate detection of complex structural variations using single molecule sequencing. *Nature Methods*, 461-468.

Seenisamy, J., Rezler, E. M., Powell, T. J., Tye, D., Gokhale, V., Joshi, C. S., . . . Hurley, L. H. (2004). The Dynamic Character of the G-Quadruplex Element in the c-MYC Promoter and Modification by TMPyP4. *Journal Of The American Chemical Society*, 8702-8709.

Setlow, R. B., Swenson, P. A., & Carrier, W. L. (1963). Thymine Dimers and Inhibition of DNA Synthesis by ultraviolet Irradiation of Cells. *Science*, 1464-1466.

Shrivastav, N., Li, D., & Essigmann, J. M. (2010). Chemical biology of mutagenesis and DNA repair: cellular responses to DNA alkylation. *Carcinogenesis*, 59-70.

Singh, A., & Xu, Y.-J. (2016). The Cell Killing Mechanisims of Hydroxyurea. *Genes*.

Sotiriou, S. K., Kamileri, I., Lugli, N., Evangelou, K., Da-Re, C., Huber, F., . . . Halazonetis, T. D. (2016). Mammalian RAD52 Functions in Break-Induced Replication Repair of Collapsed DNA Replication Forks. *Molecular Cell*, 1127-1134.

Stahl, M. M. (1958). The replication of DNA in Escherichia coli. *Proceedings of the National Academey of Sciences of the United States of America*, 671-682.

Sung, P. (2018). Introduction to the Thematic Minireview Series: DNA double-strand break repair and pathway choice. *Journal of Biological Chemistry*, 10500-10501.

Wright, W. D., Shah, S. s., & Heyer, W.-D. (2018). Homologous recombination and the repair of DNA double-strand breaks. *Journal of Biological Chemistry*, 10524-10535.

Yilong Li, N. D. (2020). Patterns of somatic structural variation in human cancer genomes. *Nature*, 112-121.