A NOVEL SYSTEM FOR DETECTION OF DNA DOUBLE STRAND BREAKS AND REPAIR IN HUMAN CELLS

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

By

Todd Warren Lewis
B.S., Wright State University, 2010
M.B.A., Wright State University, 2015

2017
Wright State University

Michael Leffak, Ph.D.
Dissertation Director

Mill W. Miller, Ph.D.
Director, Biomedical Sciences Ph.D. Program

Robert E. W. Fyffe, Ph.D.
Vice President for Research and Dean of the Graduate School

Committee on Final Examination

Michael Leffak, Ph.D.

Steven J. Berberich, Ph.D.

Yong-jie Xu, Ph.D.

David Cool, Ph.D.

Mill W. Miller, Ph.D.
ABSTRACT

Lewis, Todd Warren Ph.D., Biomedical Sciences Ph.D. program, Wright State University, 2017. A novel system for detection of DNA double strand breaks and repair in human cells

Imperative to genomic stability is the ability of the cell to repair damaged DNA which can occur from numerous endogenous byproducts of metabolism or exogenous components from the environment. The Fanconi anemia pathway is a DNA repair mechanism used by human cells to resolve multiple forms of DNA damage including interstrand crosslinks (ICL). Fanconi anemia (FA) is an autosomal recessive inherited disorder characterized by genome instability, developmental abnormalities, cancer predisposition and bone marrow failure. FA is attributed to a mutations in at least 18 genes (FANCA-FANCT) that play a concerted role in DNA repair. FANCT is the latest discovery in the FA pathway and is a UBE2T ubiquitin conjugase. An FA patient with biallelic germline mutations in FANCT presented with the classical symptoms of FA with the exception of hematopoietic indications. Both mutations, a maternal duplication and a paternal deletion of exons 2-6, were AluY mediated. Upon further evaluation it was determined that the FA patient had a reversion of the maternal duplication back to a WT allele which restored FANCT function in the hematopoietic lineage. The genomic reversion was also attributed to an Alu mediated recombination (AMR) event. The factors that led to the AMR event were unknown and provided reason to investigate further. The promise of harnessing the mechanism to utilize a genomic duplication reversion could be of importance for therapeutic interventions.

No current human cell line models were sufficient to address what factors are involved in the FANCT reversion, so a novel Dual Fluorescence (DF) model system was created in a HeLa cell line to test the hypothesis that Alu mediated homology directed
repair is sufficient to explain genetic reversion of a partially duplicated FANCT locus. The DF model system emulates elements of the FANCT locus using partial intronic sequences and Alu elements found in the patient’s duplicated FANCT locus. Key to the model system design is the incorporation of TOM and eGFP fluorescent marker genes and an I-SceI recognition site into the ectopic locus. I-SceI is a rare endonuclease found in yeast, with no known recognition site found in humans and when expressed in the DF cells will cause a targeted DNA double strand break. The fluorescent marker genes are separated and flanked by Alu elements. After an I-SceI induced DSB, it is possible that AMR will occur and eliminate one of the fluorescent marker genes, thereby shifting fluorescence of the cell. This shift in fluorescence can be detected by microscopy or flow cytometry. To test whether Alu mediated homology directed repair is sufficient to explain genetic reversion of a partially duplicated FANCT locus, multiple DF cell lines were created that vary in; location of the I-SceI recognition site, number of Alu elements and the presence of repetitive DNA sequences.

Flow cytometry of the DF cells after an I-SceI mediated DSB, reveals that recombination of the proximal Alu elements is the most likely repair outcome. Further evaluation of DF cell line recombinants by Flow Assisted Cell Sorting (FACS) and DNA sequencing disclose that the model system can differentiate between homology and non-homology directed repair. To test if Alu mediated homology directed repair can facilitate a genetic reversion, canonical DNA repair proteins in homology directed repair (HDR) and non-HDR pathways were impaired. The results of this experiment provide evidence that only HDR pathways support the genetic reversion. Upon evaluation of a FANCT null cell line in the DF model and the well-established U2-OS GFP model, both models
indicate FANCT is involved in homology directed repair. The unique DF model system was able to determine that FANCT is also involved in non-HDR. These experiments provides evidence for the novel finding of FANT’s role in HDR and non-HDR outside of ICL resolution via the FA pathway. Another aspect to examine when assessing AMR causality is what factors may lead to the initial DNA break occurrence. There is evidence in multiple diseases that an AMR event was caused by a DNA break at a microsatellite sequence. To address the question of how repetitive DNA sequences influence Alu mediated homology directed repair, the DF cell lines were modified to include multiple microsatellite motifs. The results of an I-Scel induced break juxtaposed to a microsatellite sequence reveal that both homology and non-homology directed recombination events are significantly impaired, presumably by the formation of DNA secondary structure. The results of this project suggest that Alu mediated homology directed repair is sufficient to explain genetic reversion of a partially duplicated FANCT locus. Additionally, the development of the DF cell lines provide a novel model system for detection of DNA double strand breaks and repair in human cells.
# Table of Contents

## INTRODUCTION

- Fanconi anemia promotes genome instability ................................................................. 1
- Alu elements possess mutagenic potential ................................................................. 5
- Microsatellite sequences are hotspots for recombination ........................................... 10
- DNA double strand break repair mechanisms ......................................................... 14
- Development of a novel DNA repair reporter system ............................................. 17

## MATERIALS AND METHODS

- Stable integration of DF constructs ............................................................................. 22
- Expression of I-Sce1 endonuclease in human cells ....................................................... 24
- siRNA treatment ........................................................................................................... 24
- Small molecule inhibitor treatments ............................................................................ 26
- Single cell sorting and PCR .......................................................................................... 26
- Flow cytometry ............................................................................................................. 27
- Western blotting ........................................................................................................... 28
- Aphidicolin and hydroxyurea treatments .................................................................... 29
- Statistical analysis ....................................................................................................... 29

## RESULTS

### I. Development of a novel model to emulate reversion of a genetic duplication in Fanconi anemia

- Successful creation of a Dual Fluorescence model system in human cells ........... 30
- Recombination is not biased by placement of DSB ................................................... 31
- Presence of upstream Alu element affects recombination pattern ......................... 43
FACS analysis and PCR of DF2 cell line reveals unique patterns of recombination ................................................................. 49

Summary ........................................................................................................................................................................ 54

II. DNA repair proteins modulate reversion in FANCT model .......... 57
   CtIP and RAD51 promote reversion of FANCT duplication ............... 57
   DNA-PK does not promote FANCT reversion ................................ 60
   ATM affects both HDR and non-HDR events .............................. 62
   FANCT is involved in HDR and non-HDR events ........................... 64
   Summary ........................................................................................................................ ........................................ 70

III. HDR is reduced when a DSB occurs adjacent to a DNA secondary structure forming sequence ........................................ 70
   Creation of model cell lines .............................................................. 71
   DNA repair is not detected in replication challenged model cell lines .... 72
   Secondary structure forming sequences adjacent to a DSB modulate HDR. 82
   Summary ........................................................................................................................ ........................................ 84

DISCUSSION .............................................................................................................................................................. 89

   Novel DF model system can emulate the genomic reversion displayed in FANCT patient cells ........................................ 89
   Analysis of the DF2 cell line reveals unique patterns of recombination ..... 91
   Impairment of DNA repair proteins CtIP and Rad51 confirm red quadrant recombinants are due to an HDR event .................. 92
   Inhibition of DNA-PK confirms that double negative and eGFP quadrant recombinants are primarily derived from non-homology dependent repair events ............................................................. 93
   DF system is capable of detecting homology dependent and non-homology dependent repair events simultaneously ................................................................. 94
The repair pattern of the double positive recombinants was not determined under DNA repair inhibition conditions .................................................. 95

FANCT is involved in homology dependent and non-homology dependent DNA repair.................................................................................. 96

DNA repair is not detected in replication challenged DF3 microsatellite model cell lines ................................................................. 98

Secondary structure forming sequences influence DNA repair patterns in human cells ........................................................................... 101

FUTURE DIRECTIONS .............................................................................................................. 105

CONCLUSION ......................................................................................................................... 106

REFERENCES ........................................................................................................................ 107
List of Figures

Figure 1: Genome instability in Fanconi anemia patient .................................................. 3
Figure 2: FA proteins and repair activation................................................................. 4
Figure 3: Model for patient reversion of partially duplicated FANCT allele............. 6
Figure 4: Microsatellite sequences capable of DNA secondary structure formations... 12
Figure 5: Select DNA DSB homology directed repair mechanisms......................... 16
Figure 6: Schematic of recombination in U2-OS DR-GFP cell lines......................... 18
Figure 7: Representation of how the model system reversion emulates patient reversion
........................................................................................................................................ 20
Figure 8: Creation of stable ectopic cell lines.............................................................. 23
Figure 9: Peak I-SceI protein expression is at 24 hours post transfection.............. 25
Figure 10: Model cell lines with only Tomato or eGFP.............................................. 32
Figure 11: Model cell lines to determine DSB location bias in recombination............ 33
Figure 12: Fluorescent microscopy reveals a shift in fluorescence in experimental cell
lines after I-SceI expression.......................................................................................... 36
Figure 13: Flow cytometry determination of recombination by shift in fluorescence ... 37
Figure 14: A delay in flow cytometry measurement is required................................. 38
Figure 15: Genomic recombination occurs within 48-72 hours.............................. 39
Figure 16: PCR analysis of DF1, DF2 and DF3 cell lines displays distinct recombination
patterns....................................................................................................................... 41
Figure 17: Location of DSB does not affect repair pattern after I-Sce expression........ 42
Figure 18: DF4 and DF5 cell lines emulate patient WT allele.................................... 44
Figure 19: PCR analysis of DF4 and DF5 cell lines demonstrate distinct recombination patterns .......................................................... 46

Figure 20: DF4 and DF5 recombination patterns after I-Sce1 expression ....................... 47

Figure 21: Loss of Alu1 changes the pattern of recombination: .................................. 48

Figure 22: FACS sorting and PCR of double positive quadrant .................................. 51

Figure 23: FACS sorting and PCR of red quadrant .................................................. 52

Figure 24: FACS sorting and PCR of double negative quadrant ............................... 55

Figure 25: FACS sorting and PCR of green quadrant ............................................. 56

Figure 26: Flow cytometry of siCtIP treated DF2 cells ........................................ 59

Figure 27: Flow cytometry of RAD51i treated DF2 cells ...................................... 61

Figure 28: Flow cytometry of DNA-PKi treated DF2 cells ................................... 63

Figure 29: Flow cytometry of ATMi treated DF2 cells ......................................... 65

Figure 30: FANCT is involved in HDR .................................................................. 67

Figure 31: FANCT maybe involved in HDR and non-HDR events ......................... 68

Figure 32: Comparable percentage of GFP+ recombinants in FANCT null model systems ................................................................................................................. 69

Figure 33: Creation of repeat containing model cell lines ..................................... 73

Figure 34: Repeat motif model cell lines used ..................................................... 74

Figure 35: Treatment of DF3 Myc cell line with replication inhibitors ................... 76

Figure 36: Treatment of DF3 Myc CAG84 cell line with replication inhibitors ....... 77

Figure 37: Treatment of DF3 Myc CTG100 cell line with replication inhibitors ....... 78

Figure 38: Treatment of DF3 Myc TF cell line with replication inhibitors ............. 80

Figure 39: Treatment of DF3 Myc TR cell line with replication inhibitors ............. 81
Figure 40: Replication inhibitors induce recombination in DF2 CTG<sub>100</sub> cell line .......... 83

Figure 41: TNR sequences juxtaposed to an I-Sce1 DSB reduces HDR.......................... 85

Figure 42: Triplex forming sequences juxtaposed to an I-Sce1 DSB reduces HDR........ 86
List of Tables

Table 1: Alu mediated recombination derived mutations in tumor suppressor genes ..... 9
ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my mentor Dr. Michael Leffak for the knowledge and support he has provided during my time in the lab. I have developed immensely as a scientist and critical thinker under his guidance. I would also like to thank the members of my dissertation committee: Dr. Mill Miller, Dr. Steven Berberich, Dr. Yong-jie Xu and Dr. David Cool. The committee’s suggestions and critical guidance has truly made me a better scientist.

To current and former members of the Leffak lab, I would like to offer my sincerest gratitude for the help and camaraderie that you have provided me. I would like to especially thank Dr. Yanzhe Gao for his help in developing my techniques as a burgeoning scientist. I would like to thank the entire Biochemistry and Molecular Biology Department for their encouragement and support. This includes former and current Chairs; Dr. Steven Berberich, Dr. Michael Leffak and Dr. Madhavi Kadakia. I also would like to thank Shannon Norton for the support and encouragement she offered on a daily basis. I would like to extend a very special thanks to the Biomedical Sciences Program including Dr. Gerald Alter, Dr. Mill Miller and Karen Luchin for their overwhelming support. A special thank you to Dr. Courtney Sulentic, Dr. Eric Romer and other members of the Sulentic lab who provided my first opportunity to conduct research.

Finally, I would like to thank all my family and friends who have supported my decision to return to academic studies. It is with your love and support that I am able to pursue my passion of helping others through research.
DEDICATION

To my parents Georgia and Jim, my soul mate Daphne

“Learn from yesterday, live for today, hope for tomorrow. The important thing is not to stop questioning”

-Albert Einstein
INTRODUCTION

Imperative to genomic stability is the ability of the cell to repair damaged DNA, which can occur from numerous endogenous byproducts of metabolism or exogenous components from the environment (Abbotts et al., 2014; Guirouilh-Barbat et al., 2014). DNA damage can directly cause single strand breaks (SSB) and double strand breaks (DSB) or breaks could arise from DNA repair processing of adducts, lesions or replication collapse (Gent et al., 2001; Johnson & Jasin, 2001; Rothstein et al., 2000). DSB’s are particularly detrimental because they can lead to genomic rearrangement, tumorigenesis and cell death if not repaired correctly and efficiently.

Fanconi anemia promotes genome instability

Fanconi anemia (FA) is an autosomal recessive inherited disorder characterized by genome instability, developmental abnormalities, cancer predisposition and bone marrow failure (Kalb et al., 2006). FA is attributed to mutations in any of at least 18 genes (FANCA-FANCT) that play a concerted role in DNA repair (Wang & Smogorzewska, 2015). FA patients are particularly sensitive to DNA interstrand crosslinks (ICLs), which covalently link DNA strands and block replication and transcription (Auerbach, 2009). For clinical diagnosis, patient cells will have increased chromosome breakage if subjected to an ICL agent such as mitomycin C (MMC) or diepoxylbutane (DEB) when compared to healthy individuals, due to the inability to effectively repair the damage (Shimamura et al., 2002). Chromosome abnormalities
include mitotic chromosome breaks, chromatid breaks, acentric fragments and chromatid exchanges and contribute to the clinical presentation of the disease (Figure 1) (Cirkovic et al., 2011). FANCT, an E2 ubiquitin conjugase (UBE2T), was recently classified as the 18th FA protein member when it was found deficient in an FA patient (Hira et al., 2015).

It is known that after the FA core complex recognizes an ICL, it initiates an E1-E3 ubiquitination chain that includes FANCT and FANCL (E3 ubiquitin ligase) (Figure 2). FANCL will ubiquitinate FANCD2 as part of the ID complex (FANCI and FANCD2), thereby activating the ID complex to interact with other HDR repair proteins and resolve the ICL (Kee & D’Andrea, 2010; Takata et al., 2006). Interestingly, some Fanconi anemia proteins are not used exclusively for just ICL repair, but are also involved in other types of repair such as homology directed repair (HDR) (Patel, 2007). It is currently unknown what the function of FANCT is outside of the FA pathway, but it could additionally play a role in HDR or non-HDR pathways.

Fanconi anemia typically presents with growth abnormalities and anemia, but a unique FA phenotype was revealed when fibroblasts from a 16 year old FA patient 100166/1 were analyzed (Rickman et al., 2015; Virts et al., 2015). Curiously, patient 100166/1 displayed most of the typical clinical presentations of FA, that include developmental and growth abnormalities, but the patient unexpectedly did not display bone marrow failure and anemia in the hematopoietic cell lineage. Examination of the genotype in the patient’s non-hematopoietic cells revealed that one allele is partially duplicated (maternal) and the other allele contains a partial deletion (paternal). FA is an autosomal recessive disease so the fact the patient can no longer produce a functional FANCT protein in these tissues leads to the phenotype. When observing the genotype in
Figure 1: Genome instability in Fanconi anemia patient. A variety of chromosome abnormalities appear when Fanconi anemia patient cells are challenged with DEB an ICL agent. (A) Chromosome break, (B) Chromatid break, (C) Acentric fragment, (D) Chromatid exchanges (Cirkovic et al., 2011)
**Figure 2: FA proteins and repair activation.** Within the FA pathway the FA core complex will recognize DNA damage and initiate the ubiquitination cascade of E1, FANCT, and FANCL. This will elicit a monoubiquitination of the ID complex: FANCD2 and FANCI. Once the ID complex is ubiquitinated it recruits HR proteins to complete the DSB intermediate step.
the hematopoietic lineage it was determined that the partial duplication had reverted back to wild type (WT) and produced a functional FANCT protein that enables amelioration of the anemia phenotype. Detailed examination of the parental and patient genotypes suggested that the maternal duplication/reversion and paternal deletion of the FANCT locus were due to Alu mediated recombination (AMR) events involving AluY elements (Figure 3).

**Alu elements possess mutagenic potential**

An Alu element is from the family of repetitive elements that are considered transposable elements under the classification Short Interspersed Nuclear Element (SINE) and consist of ~300 bp. A transposable element is known as a mobile genetic element due to its ability to transfer genetic information to other parts of the genome. An Alu element is a Class I retrotransposon that is initially transcribed from DNA to RNA by RNA Pol III, and then transcribed back to DNA by reverse transcriptase. SINEs do not encode for reverse transcriptase, instead using reverse transcriptase produced from a trans-acting Long Interspersed Nuclear Element (LINE), and therefore are non-autonomous (Dewannieux et al., 2003).

Alu elements are primate specific, occurring prevalently throughout the human genome with copy numbers exceeding 1 million (Batzer & Deininger, 2002; Häsler & Strub, 2006; Lander et al., 2001). There are variations in Alu sequences that have been classified into ~30 subfamilies, with the AluY subfamily being the most active for retrotransposition in the hominoid lineage (Hormozdiari et al., 2011). Conserved throughout eukaryotes, Alu elements can induce genomic instability either by interrupting a gene directly during transposon insertion, or by facilitating aberrant
Figure 3: Model for patient reversion of partially duplicated FANCT allele. The numbered gray boxes represent exons. The yellow boxes indicate AluY elements. (A) The mother has one partially duplicated allele and the father one partially deleted allele. Both parents have a functional WT allele and do not display the phenotype of FA, an autosomal recessive disease. (B) The progeny (FANCT patient) displays the FA phenotype in all tissues except hematopoietic cells due to a maternal duplication and paternal deletion. (C) In the patient’s hematopoietic cell lineage the maternal duplication was reverted back to a functional WT allele due to Alu mediated recombination.
A

maternal wild type

paternal wild type

maternal duplication

paternal deletion

B

patient duplication

patient deletion

C

patient non-hematopoietic cells-duplication

patient hematopoietic cells-reversion
recombination following a DNA double strand break (DSB) due to sequence homology of nearby Alu elements. Alu mediated recombination (AMR) can occur when there is a DNA break between two Alu elements, followed by a recombination between the upstream Alu and downstream Alu, thereby facilitating a genomic deletion of the intervening DNA sequence. Inappropriate Alu element recombination serves as a facilitator of at least 33 diseases and 16 cases of cancer that are attributed to unequal homology directed repair following a DSB with flanking Alu elements (Deininger & Batzer, 1999).

This project will focus on direct Alu repeat elements, considering that inverted Alu repeat elements have been selected against over time due to extreme genomic instability (Lobachev et al., 2000). With insight gained from examination of tumor suppressor gene mutations, it has become apparent that Alu mediated recombination (AMR) has a major impact pertaining to loss of function in many DNA repair proteins (Table 1). The results of this study support that AMR events are not isolated incidents, and may be involved in the impairment of other DNA repair genes. It should be noted that the density of Alu elements within a locus directly correlates to the incidence of AMR as represented by the increased number of mutations in the Alu dense MSH2, BRCA1, and VHL genes, in comparison to the relatively less dense MLH1 gene (Belancio et al., 2010; Mazoyer, 2005; van der Klift et al., 2005).

Concurring with the Alu density finding, another study analyzed the human genome for AMR deletion events and deduced that 75% of deletion events occurred when the Alu elements were less than 1kb apart. The likelihood of Alu mediated recombination also decreases as the distance between Alu regions increases. In addition,
Table 1: Alu mediated recombination derived mutations in tumor suppressor genes. Number of unique patient mutations attributed to recombination of Alu elements in select tumor suppressor genes (column 1). The disease a mutation is implicated in (column 2). Type of tissue the mutation is observed in (column 3). (Adapted from Belancio et al, 2010)

<table>
<thead>
<tr>
<th>Gene</th>
<th>Disease</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>10x MLH</td>
<td>HNPCC</td>
<td>Germ line</td>
</tr>
<tr>
<td>32x MSH2</td>
<td>HNPCC</td>
<td>Germ line</td>
</tr>
<tr>
<td>30x VHL</td>
<td>VHL</td>
<td>Germ line</td>
</tr>
<tr>
<td>15x hCAD</td>
<td>Hepatoma</td>
<td>Somatic</td>
</tr>
<tr>
<td>5x MYB</td>
<td>T-ALL</td>
<td>Somatic</td>
</tr>
<tr>
<td>7x MLL1</td>
<td>AML</td>
<td>Somatic</td>
</tr>
<tr>
<td>23x BRCA1</td>
<td>Breast cancer</td>
<td>Germ line</td>
</tr>
</tbody>
</table>

Table 1: Alu mediated recombination derived mutations in tumor suppressor genes. Number of unique patient mutations attributed to recombination of Alu elements in select tumor suppressor genes (column 1). The disease a mutation is implicated in (column 2). Type of tissue the mutation is observed in (column 3). (Adapted from Belancio et al, 2010)
it was found that 490 of the 492 AMR deletions studied were due to Alu elements in direct orientation (Sen et al., 2006). These data suggest that AMR is not an isolated occurrence, especially when the Alu elements are in parallel orientation. Therefore, it would be valuable to create a model system within human cells that has the potential to identify factors involved in AMR.

While AMR primarily presents as deletions there is also the possibility of translocation or expansion events. A partial tandem duplication attributed to AMR is found in a subset of acute myeloid leukemia (AML), trisomy 11 and Fanconi anemia patients (Strout et al., 1998; Virts et al., 2015). FANCA is a subclass of FA in which both alleles are mutated with either a point mutation, frameshift, deletion or expansion that leads to a non-functional protein. Interestingly several cases have been identified in which one allele has reverted back to a wild type allele in the lymphohematopoietic cell line by AMR and corrected the FA phenotype (De Rocco et al., 2014; Morgan et al., 1999). This functional correction is quite significant because the reversion could provide a proliferative advantage to the reverted cells and thus repopulate the bone marrow (Waisfisz et al., 1999). While it is hypothesized that a genetic reversion in some diseases are due to Alu mediated homology directed repair, it is not always clear why the initial break occurred.

**Microsatellite sequences are hotspots for recombination**

Another aspect to examine when assessing AMR causality, is what factors may lead to the initial DNA break occurrence. Once a DSB is made and repaired via AMR, the resulting recombination product will not provide details of break origin because it will recombine on the Alu elements and lose the intervening sequence resulting in a
deletion and discarding where the initial break occurred. While there are a number of possibilities that could lead to a DSB there are also some predictable sequence dependent regions where a DSB is more likely to occur. Common fragile sites are frequently areas of genomic rearrangement, and are often identified when observing a metaphase spread the fragile sites will appear as constrictions or gaps when cells are exposed to aphidicolin (Schwartz et al., 2005). While common fragile sites are thought to occur in all individuals, rare fragile sites are not as common (Glover et al., 1984; Zlotorynski et al., 2003). Some fragile site loci can be attributed to an expanded microsatellite DNA sequence such as trinucleotide repeat sequences (TNR’s), G-rich DNA or triplex forming sequences (TFS) capable of forming secondary structures (DNA hairpins, triplexes or quadruplexes) that can induce replication fork stalling and fork collapse if not properly repaired (Ma et al., 2012; Sutherland, 2003).

While TNR’s are one type of microsatellite sequence that contains a repeating sequence of three nucleotides, other microsatellite sequences can consist of a repeating sequence of ~1-10 nucleotide sets. TFS’s can be an uninterrupted microsatellite repeat sequence or have a few interruptions within the repeat tract. For simplicity sake, TFS will be referred to as a microsatellite sequence throughout this document. TFS’s form secondary structures where a ssDNA stretch of homopyrimidines “fall back” on a duplexed section of the motif and through Hoogsten bonding form a secondary structure known as a triplex (Figure 4). During replication, a single strand break in the template strand can develop into a double strand break once the replication fork reaches the break. The nascent strand would stop at this point and result in a DSB (Branzei & Foiani, 2010; Dalgaard, 2011; Mirkin & Mirkin, 2007). To experimentally induce replication
Figure 4: Microsatellite sequences capable of DNA secondary structure formations. (A) CAG trinucleotide repeats that have conformed to a hairpin structure. (B) Single stranded homopyrimidine strand can Hoogsteen base pair to the duplexed DNA and form a triplex structure. (C) GCC trinucleotide repeats can form a G-quadruplex structure.
dependent DNA breaks, replication inhibitors such as aphidicolin or hydroxyurea are used to slow down the replicative polymerases. By specifically inhibiting the polymerases it can initiate uncoupling of the replisome from the replicative CMG (CDC45, GINS, MCM 2-7) helicase, considering that the helicase is not kinetically affected by these replication inhibitors. This, in turn, will produce long stretches of ssDNA that have an increased potential for damage and breaks (Chan et al., 2012). Once the nascent strand encounters the single strand break, it is effectively converted into a double stranded break (Caldecott, 2008; Petermann & Helleday, 2010). Combining a hypersensitive TNR or TFS locus with the use of low dose replication inhibitors will increase the abundance of ssDNA, thereby exacerbating the secondary structure forming ability of the DNA sequence. The in vivo formation of DNA secondary structures can increase the likelihood of genomic instability (Yang et al., 2003).

There is evidence in multiple diseases that an AMR event was caused by a DSB initiated at a microsatellite sequence. Amongst subgroups of FANCA, Birt-Hogg-Dube Syndrome (BHDS) and Glanzmann Thrombasthenia (GT) patients; microdeletions and microinsertions were found within 5 nt. of tetranucleotide CCTG/CAGG flanking motifs indicating hotspots within the gene, in addition to a GAA/TTC repeat tract in another subset of FANCA patients (Benhammou et al., 2011; Gregory et al., 2001; Levran et al., 1997; Li & Bray, 1993). These breaks were repaired by an alternate mechanism that did not result in AMR but rather a frameshift that allows the breakpoint origin to be tracked. If AMR had proceeded, the intervening sequence would be lost and the breakpoint undetectable, as presented in other patients of FANCA, BHDS, and GT originating from AMR. This displays the potential for specific motifs CCTG/CAGG, GAA/TTC (also
implicated in microsatellite instability attributed to hairpin and triplex formations respectively) to induce genomic instability by inducing AMR in three independent diseases: FANCA, BHDS and GT. Once a DNA double strand break occurs in the intervening sequence between Alu elements, there are several possible homology directed repair mechanisms that could account for the AMR.

**DNA double strand break repair mechanisms**

The two primary forms of DSB repair are non-homologous end joining (NHEJ) and homologous recombination (HR) (Pierce et al., 2001). Homologous recombination is primarily a fidelitous process that predominates during S/G2 phases when the sister chromatid is available to use as a template, and requires broken chromatid strand invasion into the sister chromatid based upon sequence homology (Saleh-Gohari & Helleday, 2004). For strand invasion to occur, the DSB ends must undergo resection of DNA 5’ ends which occurs primarily in S/G2 phases of the cell cycle. The resection of the DNA 5’ ends of the DSB will expose DNA 3’ tails, that are subsequently used for strand invasion of the sister chromatid (X. Li & Heyer, 2008; Symington & Gautier, 2011). In contrast, NHEJ religates the broken DNA ends and does not require significant DNA end processing but usually results in the loss or addition of a few nucleotides, which makes it a potentially mutagenic process (Davis & Chen, 2013).

After a DNA double strand break, the protein kinase ATM will act as a transducer to activate effectors of DNA damage repair in both the HR and NHEJ pathways (Marechal & Zou, 2013; Wang et al., 2013). Competition between HR and NHEJ that can influence DSB outcomes, and a number of DNA repair proteins have been implicated to influence the pathway choice, ultimately with resection being the decisive point.
between HR and NHEJ pathways (Chapman et al, 2012; Jain et al., 2009). For the HR pathway, there is growing evidence that C-Terminal Binding Protein 1-Interacting Protein (CtIP) initiates the first step in the DNA resection process required for homology directed repair (Sartori et al., 2007; You et al., 2009). Intermediate to the homologous repair process, RAD51 recombinase is required for the free 3’ DNA tail to search for a homologous DNA sequence and invade the sister chromatid to use as a repair template (Baumann & West, 1998). In contrast, non-homologous repair does not require resection of DNA ends. Ku proteins first recognize and bind to non-resected DNA ends and facilitate the binding of DNA-dependent protein kinase (DNA–PK). Once bound to the DNA, DNA-PK will recruit additional DNA processing and repair proteins required for the NHEJ process (Pawelczak et al., 2011).

Classical HR and NHEJ are the predominating forms of DNA repair in the genome, but when a DSB occurs near repeated sequence homologies alternate repair pathways may be utilized (Huertas, 2010; Liang et al., 1998). When repeated sequences, such as an Alu element in direct repeat orientation, are in close proximity flanking a DSB the invading strand could invade the second repeat of the sister chromatid and lose the intervening sequence between repeats (Kataoka et al., 2013; Witherspoon et al., 2009). This is a mutagenic process termed unequal homologous recombination (uHR), and is defined as intrachromosomal if it is within the same allele, and intermolecular if it uses the sister chromatid for repair (Figure 5).

Another possible mutagenic outcome is an intrachromosomal event termed single strand annealing (SSA), which is intramolecular and does not use the sister chromatid. For SSA, once resection has exposed the region of homology on the opposite DNA
Figure 5: Select DNA DSB homology directed repair mechanisms. Rectangles represent Alu elements on DNA strands. All HDR pathways presented are DNA 5’ end resection dependent. (A) Precise homologous repair invades the sister chromatid and provides faithful repair. (B) Unequal homologous repair results in loss of intervening sequence. (C) Single strand annealing is an intrachromosomal recombination and results in loss of intervening sequence.
strand, it aligns and reanneals, possibly forming 3’ flaps that are subsequently excised from the recombinant product (Fishman-Lobell et al., 1992; Jain et al., 2009). The resulting recombination products of SSA and uHR can appear the same since the intervening sequence between the Alu’s will be deleted and the sister chromatid is an exact copy of the other strand before damage (Kim et al., 2001; Vidal et al., 2002). To help understand the underlying DNA break and repair factors involved in Alu mediated homology directed repair, it would be extremely valuable to have a model system that can emulate an Alu mediated genetic reversion in human cells.

**Development of a novel DNA repair reporter system**

A widely accepted reporter assay for determination of homology directed repair uses a genetically engineered cell line containing the DR-GFP construct (Moynahan et al., 2001; Nakanishi et al., 2011). In this model a full length of the GFP gene is located approximately 3.7 kb upstream of a truncated GFP gene. The upstream GFP gene has an I-Sce1 recognition sequence placed into the GFP reading frame and introduces a stop codon, whereas the downstream GFP fragment is 5’ and 3’ truncated, therefore no GFP protein is produced. I-Sce1 is rare yeast endonuclease that has no known recognition site in the human genome and will induce a DNA double strand break at a specific 18 bp DNA sequence. After I-Sce1 expression in the DR-GFP system, an I-Sce1 induced DSB in the upstream GFP fragment can use the downstream GFP fragment as donor template to repair the break by HDR after 5’ resection has occurred (Figure 6). Since the downstream GFP fragment (template for repair) does not possess an I-Sce1 recognition site (with a stop codon) the functional WT GFP will be restored and GFP positive cells can be detected by flow cytometry. There are several drawbacks to using the DR-GFP
Figure 6: Schematic of recombination in U2-OS DR-GFP cell lines. DR-GFP construct contains an upstream full length GFP coding fragment interrupted with an I-Sce1 recognition site that contains a stop codon. In addition, there is an iGFP fragment 3.7 kb downstream which contains an internal GFP sequence that is missing the 5’ and 3’ ends of the coding sequence. After a DSB the broken 3’ DNA end in the upstream element can use the coding sequence in the iGFP fragment as a donor template and restore the GFP to a functional state.
system as a model for the genetic reversion observed in the patient. First, the designed regions of homology are based on an extraneous GFP DNA sequence and not Alu elements. Second, the system can only detect HDR events; to ascertain insight into other forms of DNA repair, additional assays would have to be performed. Thirdly, the DR-GFP model only contains 2 segments of homology (GFP), whereas the duplicated patient locus contains 3 segments of homology (Alu).

To specifically address factors involved in the hypothesized Alu mediated FANCT genetic reversion, a novel Dual Fluorescence (DF) model system was created by our lab to emulate elements of the FANCT endogenous locus (Figure 7). The DF cell line model differs from the DR-GFP system by expressing TOM and eGFP as the default state that will lose one or both fluorescent markers if mutagenic recombination occurs. The primary expected outcome is an HDR event, with recombination facilitated by the Alu elements proximal to the DSB and thereby shifting fluorescence. In contrast, the DR-GFP model begins with no fluorescence as default and then will shift to green if an HDR event occurs. The key advantage of the DF model system is that, like the endogenous FANCT locus, it contains Alu elements as segments of DNA homology. Additionally, the new DF model system has the ability to detect differing recombination outcomes with only one highly sensitive assay.

This project utilized the novel DF model system to support the hypothesis that Alu mediated homology directed repair is sufficient to explain genetic reversion of a partially duplicated FANCT locus. Specifically, I wanted to determine the following aims: i) What are the patterns of Alu mediated homology directed after an inducible DNA
Figure 7: Representation of how the model system reversion emulates patient reversion. (A) Partial duplication of the maternal FANCT allele (B) The ectopic site in the DF2 cells emulate the partial duplication in the patient’s maternal allele. (C) Possible recombination product in the DF2 cell line after an I-Sce1 directed DSB. (D) Reversion of partial duplication to wild type FANCT allele.
double strand break, ii) Which DNA repair pathways are involved in DSB resolution, and iii) How do repetitive DNA sequences influence Alu mediated homology directed repair.
MATERIALS AND METHODS

Stable integration of DF constructs

The Flp recombinase system is used as a method of delivery to stably integrate a donor DNA construct into an acceptor locus at a specific location in the genome. Our lab has previously constructed the HeLa/406 acceptor line that contains a single integrated ectopic FLP Recombinase Target (FRT) site (Figure 8). The FRT site is specifically targeted by FLP recombinase to cleave the FRT sites of the chromosome and a donor plasmid simultaneously, and integrate the donor plasmid into that exact location (chromosome 18p11.22). To aid in selection of successful recombinants, the FRT is flanked by hygromycin phosphotransferase (upstream) and thymidine kinase (downstream) markers. Before insertion of donor DNA within the FRT site, the cells are hygromycin resistant and ganciclovir sensitive. Once there is successful insertion of donor DNA into the FRT site, it will disrupt the TK reading frame and cells are no longer sensitive to ganciclovir. The acceptor cell line was tested by Southern blot for verification of single insertion of the FRT site (Malott & Leffak, 1999). The acceptor cell line is co-transfected with pOG44 vector (Invitrogen) that expresses the FLP recombinase and one of the experimental donor DF constructs at a 10:1 pOG44 / DF construct molar ratio. Within the DF donor construct is a promoterless neomycin resistance gene that forms a Hyg-Neo fusion gene after successful insertion into the acceptor cell line. Twenty-four hours after transfection the cells are placed on selection of 400 ug/ml
Figure 8: Creation of stable ectopic cell lines. Acceptor cell line is co-transfected with the ectopic (DF) construct and pOG44. The pOG44 plasmid expresses the FLP recombinase enzyme and then facilitates the recombination that will incorporate the ectopic construct into the acceptor cell line. The integrated construct shown here is a reference DF construct, more detail is presented in following sections.
geneticin (Gibco) for 10 days. For the following 8 days, cells are placed in 20 uM ganciclovir (Sigma) for 48 hours and then switched to geneticin for 48 hours, and continued for 2 cycles. All DF cell lines have been created and cloned out for a stable integrant. Additionally, PCR was conducted with primers in the acceptor only and donor only sites so that the PCR product will cross the insertion junction to verify correct integration as determined by the presence of PCR product and PCR product size. A total of 12 stable cell lines was generated for experimental use in this project.

**Expression of I-Sce1 endonuclease in human cells**

DF cells in a 6-well tissue culture plate (Falcon) containing ~6 x 10^5 cells/well were trypsinized and transfected with Lipofectamine 2000 (Invitrogen) and 8 ug I-Sce1 plasmid as per manufacturer’s protocol. Dulbecco’s Modified Eagle’s Medium (DMEM) supplemented with 20% Fetal Bovine Serum (FBS) was replaced after 24 hours after transfection to remove any lingering transfection complexes. The I-Sce1 plasmid produces an HA tagged form of the I-Sce1 endonuclease that is recognized by anti-HA antibody for protein detection. A time course optimization for expression was performed by immunoblotting, and it determined peak expression of I-Sce1 endonuclease was at 24 hours post transfection and undetectable by 72 hours post transfection (Figure 9). Cells were grown for 8 days after HA-Sce1 transfection and passaged accordingly until harvested for flow cytometry.

**siRNA treatment**

DF cell lines in a 6-well tissue culture plate (Falcon) containing ~4 x 10^5 cells/well were trypsinized and transfected with Lipofectamine 2000 (Invitrogen) and 100 nM (final concentration) of small interfering RNA (siRNA) pool (an equal molar
Figure 9: Peak I-Sce1 protein expression is at 24 hours post transfection. The western blot indicates that I-Sce1 expression is greatest between 24-48 hours.
mixture of 4 different siRNAs targeting the same transcript) (Hs_RBBP8 SI1027416) targeting CtIP, as per manufacturer recommendations. Control experiments were performed in parallel using a non-targeting All-Stars siRNA (Qiagen SI03650318). At 24 hours post siRNA transfection, cells were trypsinized and transfected with Lipofectamine 2000 (Invitrogen) and 8 ug I-Sce1 plasmid as per manufacturer recommendations.

**Small molecule inhibitor treatments**

Small molecule inhibitors were used to inhibit the functions of some canonical repair proteins. Rad51 inhibitor (Rad51, B02 Sigma SML03664 10 uM), DNA-PK inhibitor (Nu7026, Selleckchem S2893 10 uM), ATM/Kinase inhibitor (caffeine, Sigma C0750 2 mM), ATM inhibitor (Ku60019, Sigma SML1416 1 uM). The small molecule inhibitors were added at the same time of I-Sce1 transfection. A total of 6 x 10⁵ cells/well of a 6 well plate were trypsinized and transfected with Lipofectamine 2000 (Invitrogen), 8 ug I-Sce1 plasmid, inhibitor as per manufacturer recommendations. The inhibitors were used for 3 days after transfection to inhibit the targeted proteins throughout I-Sce1 expression.

**Single cell sorting and PCR**

DF2 cell lines were subjected to I-Sce1 transfection and allowed 8 days of post transfection recovery for full fluorescence separation. A total population of (~5 million cells) was then prepared for cell sorting by centrifugation at 300 x g for 3 min @ 4 °C. The pellet was then re-suspended in 1 ml FACS buffer (Hank’s Balanced Salt Solution, 25 mM HEPES, 1 mM EDTA, 1% BSA and 2% FBS) and filtered through a 35 μm cell strainer tube (Falcon 352235). Cell sorting was performed at Cincinnati Children’s
Medical Hospital (CCMH) Research Flow Cytometry Core (RFCC). Sorting was completed on a BD FACS Aria II flow cytometer with two 96 well tissue culture plates (single cell per well) (Corning CL 3595) for each of the four color quadrants, in addition to a polyclonal pool of cells from each color quadrant. Cells were allowed to propagate and any well with zero or multiple colonies was excluded from further downstream experiments. Single cell clones from each well were later transferred to 10 cm tissue culture dishes (Corning 430167). Once the 10 cm dishes were approximately 80% confluent, the cells were harvested, resuspended in DMEM with 10% FBS and 10% DMSO, then stored in a -80 °C freezer, with a small amount of cells from each dish being used for DNA isolation. DNA isolation was performed using an EZNA tissue DNA isolation kit (Omega Bio-Tek D3396-02) and that purified DNA used in PCR assays.

PCR was performed using Lac-F (5’- CTTCAAATCCGACCCGTAGA-3’) and TK-R (5’- GTAAGTCATCGGCTCGGGTA-3’) forward and reverse primer set. PrimeSTAR GXL polymerase (R050A) per manufacturer instructions for a 50 ul reaction using 120 ng template. Cycling conditions were as follows: denature 98 °C 10 sec, anneal 57 °C 15 sec, extension 68 °C 1 min for 30 cycles. PCR products were then purified with EZNA cycle pure kit (Omega Bio-Tek D6492-02) and 20 ul of each purified product was electrophoresed on a 1% ultra-pure agarose gel (Invitrogen 16500-100) to verify size of PCR product.

**Flow cytometry**

Flow cytometry was performed on the BD Accuri C6 machine. 5 x 10^5 adherent cells were trypsinized (Gibco) and centrifuged at 300 x g for 3 minutes. Supernatant was removed and the cell pellet was resuspended in 300 ul PBS. The cells were centrifuged
for another 3 minutes at 300 x g. Supernatant was aspirated and cell pellet resuspended in 200 ul PBS. The number of events was set to 20,000 for gated cells, with flow speed on medium setting. Color compensation was set at correcting FL2 by subtracting 6.3% of FL1. The color compensation was determined empirically to minimize spectral overlap using the DF6 and DF7 cell lines that produce only red fluorescent protein or green fluorescent protein respectively. A dot plot was used for the FSC-A vs SSC-A plot with Y-axis (0-16,777,215) and X-axis (0-16,777,215) on linear scale to determine gating out of debris. A density plot on log scale with FL2-H Y-axis (300-200,000) and FL1-H X-axis (500-2,000,000) was used for defining the color quadrants. The color quadrants were determined empirically with the DF6 (red only) and DF7 (green only) cell lines, including color compensation. X-axis division is at 10^{4.3} and Y-axis division at 10^{3.7} which enables separation of fluorescent colors into defined quadrants.

**Western blotting**

DF2 treated or untreated cells were lysed with Mammalian Protein Extraction Reagent (M-PER) (Thermo Scientific 78501) and protease inhibitor cocktail (Sigma P8340). Cleared whole cell lysates were electrophoresed on an SDS-PAGE gel and transferred to a polyvinylidene fluoride (PVDF) membrane (ThermoFisher Scientific 88518), membranes were probed overnight using a 1:1000 dilution of primary antibody against the HA tag on I-Sce1 (Santa Cruz sc-7392 (F7)), and mouse β-actin (Sigma-Aldrich A5441). After incubation in corresponding horseradish peroxidase (HRP) conjugated secondary antibody (1:2000) for 2 hours at RT, membranes were then imaged on a Fuji LAS-3000 using ImageReader.
Aphidicolin and hydroxyurea treatments

Aphidicolin (0.2 uM) and hydroxyurea (0.2 mM) were the replication inhibitors chosen for the experiments performed on the cell lines containing secondary structure forming sequences. For these experiments, the replication inhibitors were constantly applied for seven days and the cells were then allowed to recover for four days, before harvesting and testing for flow cytometry results.

Statistical analysis

For statistical analysis of the flow cytometry experiments the Wright State University Statistics Center was consulted. All of the data are expressed as mean±SEM. It was determined that a paired T-test with Bonferroni correction was the most appropriate method when comparing the four flow cytometry quadrants. P < 0.0125 was used as the criterion for significance.
RESULTS

I. Development of a novel model to emulate reversion of a genetic duplication in Fanconi anemia

Successful creation of a Dual Fluorescence model system in human cells

To help understand the underlying factors involved in AMR, I set out to develop a cell line that emulates a simplified locus that has encountered AMR. To create the model cell lines, five experimental DF plasmid vectors were designed and cloned. The vectors are stably integrated into a HeLa cell line (occurring in more than 74,000 PubMed abstracts), and recent examination of the HeLa genome determined that it is relatively stable with mutation rates akin to normal tissue, thus providing a high quality reference genome for research (Adey et al., 2013). The DF cell line designs are based upon the FA patient that had the partial duplication reversion in the hematopoietic cell lineage. The Alu elements and partial introns used in the constructs are based on the sequence of that patient where intron 1 is present only once but intron 6 is present twice since it is part of the duplication. The design incorporates the use of a rare yeast endonuclease I-Sce1 that has no known recognition site in the human genome. The constructs are designed with an I-Sce1 18 bp recognition site so that when I-Sce1 is expressed in cells containing the DF constructs it will cause a DSB. Another key aspect of the DNA constructs is the use of fluorescent protein marker genes, so that expression within the cells can be assayed via flow cytometry. Two control constructs, DF6 and
DF7, were designed without any patient sequence or I-SceI restriction sites and these were used as positive controls for gating and compensating the flow cytometry experiments because they contain only a red or a green fluorescent marker gene respectively (Figure 10).

**Recombination is not biased by placement of DSB**

DF1, DF2, and DF3 are structured similarly and express both Tomato and eGFP, with DF1 being a control without an I-SCE restriction site. The individual constructs are meant to mimic the maternal partial duplication (Figure 11). While it is evident that reversion of the duplication to WT occurred in the FANCT patient, the mechanism is unclear. The reversion back to WT in the patient could happen by a break between the first and second Alu’s or a break between the second and third Alu’s followed by recombination. An aim of this experiment is to observe if the partially duplicated FANCT reversion is biased (more likely to occur) depending on whether the DSB occurs between either the Alu 1/2 segment or the Alu 2/3 segment of the model ectopic locus.

Each fluorescent marker DNA sequence has its own promoter and poly-A signal for independent transcription. The promoter and poly-A signal for each fluorescent marker gene is also different to eliminate sequence homology that could be used for recombination. Characteristic to all DF 1-3 constructs; in order from 5’ to 3’ are an Alu - Intron 1 - Tomato - Alu - Intron 6 - eGFP - Alu - Intron 6, the difference between DF2 and DF3 is simply the placement of the I-SceI restriction site. Each Alu segment is 311 bp in length and each intervening sequence (IVS) is a partial intron consisting of the first 180 bp of the FANCT patient’s intron. The distance between Alu1/Alu2 is 1715 bp, Alu1/Alu3 is 3710 bp and Alu2/Alu3 is 1678 bp. DF2 was constructed with the I-SceI
Figure 10: Model cell lines with only Tomato or eGFP. These two cell lines do not contain any Alu elements, intronic sequences or I-Sce1 recognition site and were used to empirically determine gating of RFP vs GFP quadrants in flow cytometry analysis. (A) DF6 cell line contains only the Tomato fluorescent marker. (B) DF7 cell line contains only the eGFP fluorescent marker.
Figure 11: Model cell lines to determine DSB location bias in recombination.

(A) Schematic of maternal partial duplication in FANCT allele. Arrows are drawn in the patient genotype indicating where the DSB is induced in the experimental DF cell lines meant to mimic reversion in the patient. (B) DF1 cell line is a control that does not harbor an I-Sce1 recognition site and no DSB should occur when I-Sce1 is expressed. (C) DF2 cells have the I-Sce1 recognition site next to eGFP that is located between the second and third Alu’s. (D) DF3 cell lines possess an I-Sce1 recognition site that is adjacent to Tomato and located between the first and second Alu’s.
restriction site 5’ to the eGFP, so if there is bidirectional 5’ DNA resection and Alu2/Alu3 recombine, it would eliminate the eGFP signal and only Tomato would be produced. DF3 has the I-Sce1 restriction site 5’ of the Tomato marker and when I-Sce1 is expressed and bidirectional 5’ DNA resection occurs, then Alu1/Alu2 can recombine to induce loss of the Tomato marker and the cell will fluoresce green. Another possibility is that Alu1/Alu3 could recombine and both Tomato and eGFP will be lost, then the cell would be double negative for fluorescence. These would be the major expected recombination patterns, but other exotic recombinations and translocations could also exist. DF1 should not change fluorescence when I-Sce1 is expressed due to a lack of a restriction site for I-Sce1, therefore it can be used as a control for off-target effects of I-Sce1 on fluorescence.

If the I-Sce1 induced DSB facilitates recombination leading to a loss of fluorescent markers, it can be assessed by fluorescence microscopy. The outcome of DNA repair is a mixed population of cells with various recombination patterns. The different recombination patterns can be detected by microscopy as indicated by a shift in fluorescence from double positive (green plus red = yellow). No change in fluorescence occurred in the DF1 cell line that does not have an I-Sce1 recognition site and that cell population is primarily double positive which appears as yellow. The DF2 had a shift to red (consistent with AMR between Alu2 and Alu3), and to a smaller extent to green (consistent with AMR between Alu1 and Alu2) while also retaining some yellow. The DF3 cell lines has a large shift to green with some red, while also possessing some yellow color. Another possibility is loss of both fluorescent color markers (consistent
with AMR between Alu1 and Alu3 or long range non-HDR) and this can also visualized in both the DF2 and DF3 cell lines after I-Sce1 expression (Figure 12). These results indicate that the model is capable of recapitulating the AMR found in the FANCT patient.

Fluorescence microscopy is sufficient for a quick assay to determine recombination patterns but a more sensitive and quantitative method is desirable. Flow cytometry is sensitive and accurate enough to compare ratios of fluorescence between experimental groups. The TOM (red) fluorescence intensity will be on the FL2 channel along the Y-axis and the eGFP (green) fluorescence intensity will be on the FL1 channel along the X-axis. The measurable fluorescent data output will be divided into four quadrants: double positive = +GFP/+TOM (DP), double negative = -GFP/-TOM (DN), RFP = -GFP/+TOM (Red), GFP = +GFP/-TOM (Green) (Figure 13).

A flow cytometry time course experiment was performed on the DF2 cell line to ascertain how long of a recovery period was needed for DNA repair and the subsequent shift of cells into defined fluorescent quadrants. The recovery time course experiment reveals that the number of recombinants in each quadrant does not change much from days 5-8, but rather form tighter groups (Figure 14). In contrast to the flow cytometry results, the genomic rearrangement occurs at an earlier time point as determined by a time course PCR. The results from PCR indicate that a majority of recombination events occur within 48 hours after I-Sce1 transfection (Figure 15).

DF1, DF2 and DF3 cell lines were transiently transfected with I-Sce1 and allowed to recover for 8 days, at which time cells were harvested and subjected to PCR and flow cytometry analysis. PCR was used to analyze recombination patterns in DF1, DF2 and
Figure 12: Fluorescent microscopy reveals a shift in fluorescence in experimental cell lines after I-Sce1 expression. (A) DF1 cell line remains double positive for TOM and eGFP (yellow). (B) DF2 cell line has a significant shift from double positive to red. (C) DF3 cell line shifts significantly from double positive to green.
Figure 13: Flow cytometry determination of recombination by shift in fluorescence. Cell lines without I-Sce1 transiently expressed will be primarily double positive and in the top right quadrant. Cells that have lost green fluorescence will shift from top right quadrant to top left quadrant. Cells that no longer produce red fluorescence will move down to bottom right quadrant from top right. Cells that are absent of both red and green fluorescence will shift from the top right quadrant down to bottom left.
Figure 14: A delay in flow cytometry measurement is required. Map of the DF2 ectopic insert (Top). A flow cytometry time course experiment from 1 day to 8 days post transfection with the I-Sce1 expression vector.
Figure 15: Genomic recombination occurs within 48-72 hours. Agarose gel of PCR products from DF2 cells post I-Sce1 transfection using the Lac-F and TK-R PCR primers.
DF3 cell lines to determine if the recombination products are of expected size (Figure 16). Using a forward primer (Lac-F) that binds upstream of the first Alu in the ectopic sequence and reverse primer (TK-R) that binds the TK domain in the acceptor cell the full length ectopic site 5373 bp. If there is a DSB and recombination of the proximal Alu’s, the PCR product will be approximately 3347 bp. Recombination of distal Alu’s will produce a PCR product of 1352 bp. When I-Sce1 is expressed in the DF2 and DF3 cells it is apparent that there are two smaller recombination products that equate to the hypothesized AMR products of 3347 bp and 1352 bp, in addition to the full length size 5373 bp PCR product. This is a heterogeneous population so it is expected to observe both size recombination products. Expression of I-Sce1 in the DF1 cells should not produce a smaller PCR product since it not does not contain an I-Sce1 recognition site, and the PCR results indicate this is valid.

Flow cytometry determined, as previously hypothesized, the DF1 cell line did not change the pattern of fluorescence, with around 99% of total cells double positive (Figure 17). If AMR occurs after an I-Sce1 DSB in the DF2 and DF3 cell lines, then a shift in cell population fluorescence is expected from double positive to red (DF2) or green (DF3). The flow cytometry results reveal that after an I-Sce1 DSB, the DF2 cell population contains 53.1% double positive and 33.8% red fluorescent cells, while the DF3 cell population contains 54.1% double positive and 30.2% green fluorescent cells. These data reveal nearly identical AMR recombination patterns between the DF2 and DF3 cell lines.

The combination of microscopy, flow cytometry, and PCR analysis of DF1, DF2 and DF3 cell lines establishes evidence that the DF system is valid for detection of DSB
Figure 16: PCR analysis of DF1, DF2 and DF3 cell lines displays distinct recombination patterns. (A) Map of cell line constructs. (B) (Lanes 2,4,6) Cell lines without I-Sce1 expression, displaying only the full length PCR product. (Lane 3) is the DF1 control cell line without an I-Sce1 recognition site. Recombination products can be observed at the expected sizes after transient I-Sce1 expression in DF2 and DF3 cell lines (lanes 5,7). (C) Maps of the type of recombination patterns expected which correlate to PCR products of DF2 and DF3 cells after an I-Sce1 induced DSB when using the Lac-F and TK-R PCR primers.
Figure 17: Location of DSB does not affect repair pattern after I-Sce expression.

At the top are cell line maps and flow cytometry pattern expectations. (A) The DF1 cell line did not exhibit a shift in fluorescence. (B) DF2 cells predictably migrate significantly towards red and double negative. (C) DF3 cells shift towards green and double negative.
recombination. Additionally, the flow cytometry results indicate that there is not an AMR recombination bias due to the location of the DSB between different Alu elements.

**Presence of upstream Alu element affects recombination pattern**

A pending question needing to be addressed was how DSB recombination patterns would compare between the partial duplication model system (DF1, DF2 and DF3) and a model system that mimics the WT allele. Therefore, two additional cell lines were created that model the WT allele (Figure 18). DF4 is a control that does not contain an I-SceI recognition site and DF5 is a construct that harbors the I-SceI recognition site. The DF4 mimics the DF1 cell line and DF5 emulates the DF2 cell line constructs with the exception of the upstream Alu and IVS 1 removal. DF2 ectopic pattern from 5’ to 3’ is an Alu – IVS 1 - Tomato - Alu – IVS 6 - I-SceI - eGFP - Alu – IVS 6, while the DF5 is Tomato - Alu – IVS 6 - I-SceI - eGFP - Alu – IVS 6. The hypothesis is that the upstream Alu element present in the DF2, but not DF5 may lend an extra opportunity for HDR after a DSB and 5’ DNA end resection that will affect recombination patterns.

PCR was used to analyze recombination patterns in DF4 and DF5 cell lines to determine if the recombination products are of expected size. Using the Lac-F and TK-R primer set, the full length ectopic site PCR product will be 4882 bp. If there is recombination of the Alu elements after the DSB then the expected recombinant size is 2887 bp. In the DF4 cell line only the full length PCR product is detectable and the transient expression of I-SceI did not generate any additional recombination products. In the DF5 cell line without I-SceI expression only the full length band is observed. When I-SceI is expressed in the DF5 cells there are two PCR products observed that equate to the hypothesized AMR product of 2887 bp and the full length size of 4882 bp (Figure
**Figure 18: DF4 and DF5 cell lines emulate patient WT allele.**  
(A) FANCT patient WT allele.  
(B) DF4 cell line is the control that does not contain an I-Sce1 recognition site.  
(C) DF5 is the construct that possesses an I-Sce1 recognition site where an I-Sce1 induced DSB can occur. Both constructs contain only two Alu elements akin to the FANCT patient WT allele.
Additionally, PCR products from the DF6 and DF7 cell lines were electrophoresed on the same agarose gel to verify if the PCR products are the hypothesized size of 2600 bp (Figure 19).

Flow cytometry was also used to detect recombination following an I-Sce1 induced DSB in the DF4 and DF5 cell lines. The DF4 does not contain an I-Sce1 recognition site and displayed no difference in the double positive population of cells (~96%) after I-Sce1 was transiently expressed. In contrast, the DF5 (which contains an I-Sce1 recognition site) cell line exhibited significant recombination after the I-Sce1 induced DSB with 56.4% in the double positive cell population. Interestingly, the DF5 recombination product percentage was higher in the double negative population (22.6%) than the red population (14.4%) (Figure 20).

Now that recombination patterns have been established for DF4 and DF5 cell lines (2 Alu’s), it is pertinent for comparison to the DF2 cell line (3 Alu’s) to address if the additional upstream Alu in the ectopic locus affects DSB recombination. The number of double positive cells after an I-Sce1 induced DSB is comparable at 53.1% for DF2 and 56.4% for DF5. The striking observation between the two cell lines is the difference in the red quadrant 33.8% for DF2 vs 14.4% for DF5 (Figure 21). It appears the majority of cells have shifted from the red quadrant to double negative in the DF5 cell line (22.5%) when comparing to the DF2 cells (9.7%), even when accounting for background differences. Based on the PCR analysis the data indicates that the red quadrant is linked to homology directed repair. When comparing recombination patterns of the (2 Alu) DF5 cells and the (3 Alu) DF2 cells, the evidence suggests that loss of the upstream Alu1 changes the pattern of recombination. One possible explanation for this phenomena is
Figure 19: PCR analysis of DF4 and DF5 cell lines demonstrate distinct recombination patterns.  (A) Map of cell line constructs.  (B) Cell lines without I-Sce1 expression, displaying only the full length PCR product (Lanes 2,4).  DF4 is the control cell line without an I-Sce1 recognition site after I-Sce1 expression (Lane 3).  Recombination products can be observed at the expected sizes after transient I-Sce1 expression in the DF5 cell line (lane 5).  DF6 and DF7 fluorescent control cell lines are at expected size (Lanes 6,7).  (C) Maps of the type of recombination patterns expected which correlate to PCR product of the DF5 cells after an I-Sce1 induced DSB when using the Lac-F and TK-R PCR primers.
Figure 20: DF4 and DF5 recombination patterns after I-Sce1 expression. At the top are cell line maps and flow cytometry pattern expectations. (A) DF4 control cell line does not exhibit recombination due to lack of I-Sce1 recognition site. (B) DF5 cell line displays considerable recombination after an I-Sce1 induced DSB.
Figure 21: Loss of Alu1 changes the pattern of recombination: At the top are cell line maps and flow cytometry pattern expectations. (A) DF2 and (B) DF5 cells before (top panel) and after I-Sce1 expression (bottom panel).
that the upstream Alu present in the DF2 cell line provides an extra opportunity for Alu mediated recombination that could restore the red marker gene. In contrast, the DF5 may continue resection and subsequent recombination occur on the ectopic FRT sites or another long range recombination event that eliminates both fluorescent marker genes and primer binding sites.

**FACS analysis and PCR of DF2 cell line reveals unique patterns of recombination**

To further investigate what type of DNA repair mechanism each quadrant represents, flow assisted cell sorting (FACS) was utilized. FACS is capable of single cell per well sorting for each quadrant. DF2 cells that have been subjected to I-Sce1 expression were allowed to fully separate and recover. Then the whole population of mixed recombinant cells was used for FACS. Placing approximately one cell per well into 96 well tissue culture plates, each of the four quadrants was gated out individually and sorted. Standard Sanger sequencing of the PCR products from each color quadrant yielded uninterpretable results because of the mixture of recombinant possibilities from each quadrant in the sample. By using the FACS method, each individual cell can proliferate into a population that will yield pure genotypes that then can be subjected to PCR and sequencing. This will provide information into what type of repair mechanism would most likely account for the recombinants in each quadrant. The Lac-F and TK-R primer set was used for PCR of FACS clones.

The double positive quadrant still retains both fluorescent markers as detected by flow cytometry. I hypothesize that after an I-Sce1 induced DSB the recombinants present in this quadrant could be produced by: perfect HR, precise re-ligation of break, not cut by
I-Sce1 or non-homologous end joining (NHEJ). If the proposed mechanisms are involved, then the FACS clones from the double positive quadrant PCR should all be the same size with the exception of NHEJ which may be slightly smaller or larger. An agarose gel of PCR products from the double positive quadrant FACS clones reveals that 8 out of 10 clones produced a PCR product of the expected full length size of 5373 bp. Clones 4 and 15 are slightly smaller possibly due to NHEJ repair. Sequencing revealed that the full length PCR products are the exact initial sequence without any deletions. For clones 4 and 15, there were some smaller deletions biased toward the eGFP marker (Figure 22).

The red quadrant has lost the eGFP fluorescent marker and now only contains the TOM marker. The most reasonable explanation for this genomic recombination is through intramolecular single strand annealing (SSA) or intermolecular unequal homologous recombination (uHR). With SSA, there would be recombination between the proximal Alu’s 2 and 3 on the same chromosome which eliminates the TOM fluorescent marker. Another possibility is uHR which uses the sister chromatid as template for repair but the upstream Alu 2 on the broken chromatid could use the downstream Alu 3 as template for repair and skip over the intervening sequence thereby eliminating eGFP on the repaired chromatid. Both of these repair mechanisms ultimately lead to a deletion and thereby a reduction in the size of the PCR product. The expected size of the PCR product after recombination is 3347 bp. PCR of the red quadrant FACS clones produced products in all 10 samples with no full length size products. All 10 out of 10 clones were the correct size of 3347 bp for proximal recombination of Alu 2 and Alu 3 and sequencing confirmed this was the result (Figure 23).
Figure 22: FACS sorting and PCR of double positive quadrant. (A) Map of ectopic locus in DF2 cell line with primer binding locations included (blue arrows). (B) Agarose gel of PCR products from double positive quadrant FACS clones.
**Figure 23:** FACS sorting and PCR of red quadrant. (A) Map of ectopic locus in DF2 cell line with primer binding locations included. (B) Agarose gel of PCR products from the red quadrant FACS clones. All clones produced a PCR product and 10 of 10 were the expected proximal recombination size. No clones produced a full length progenitor size PCR product.
The FACS clones from the double negative quadrant did not display any fluorescence and are presumed to have lost both TOM and eGFP fluorescent markers. The most straightforward explanation is that after a DSB at the I-Sce1 recognition site there is distal recombination between Alu 1 and Alu 3, thereby eliminating the genomic sequence in between that contains the fluorescent markers. Another possibility is that resection extended past both the TOM and eGFP sites and a form of microhomology end joining (MMEJ) has occurred and recombined in the same fashion as NHEJ. The latter explanation would be consistent with the double negative quadrant in the DF5 cells when the upstream Alu1 is not present and therefore offers less opportunity for HDR. The PCR results indicate that the majority of recombinants are not due to recombination of Alu 1 and Alu 3. Out of 8 clones, none produced a full length or proximal Alu recombination size PCR product. Three clones did not produce a PCR product (clones 3, 29, 40), two produced distal Alu recombination size products (clones 25, 46), three clones revealed unexpected size products (clones 27, 36, 44). Sequencing indicated that clones 25 and 46 did undergo distal Alu mediated recombination. Sequencing from the unexpected PCR products appear to be due to translocation events (Figure 24).

Clones generated from FACS flow cytometry of the green quadrant have lost the TOM DNA sequence but yet retained the eGFP sequence. By an unknown mechanism, eGFP is retained while TOM is lost, which is quite unexpected considering the I-Sce1 DSB is proximal to the eGFP. It suggests a possible unidirectional resection toward the TOM from the I-Sce1 mediated DSB origin or possibly an acentric chromosome fragment that contains TOM that is subsequently degraded. There is not a hypothesized size product because using the canonical repair pathway there is no clear path of
recombination that will lead to these products. Of twelve clones, none produced a full
length size PCR product (5373 bp). Only one clone produced a PCR product at the
proximal Alu recombination size (3347 bp). One clone did not produce a PCR product
(clone 27), four produced products around 3 kb (clones 26, 29, 31, 35), three clones seem
to migrate together near 4.2 kb (clones 25, 36, 40) and three other clones produced various
size products (Clones 3, 37, 39) (Figure 25). Sequencing reveals that for the PCR
products around 3 kb, there was upstream Alu 1/2 recombination in addition to some loss
of sequence between the I-SceI site and the eGFP marker. For the other various size
clones it appears that unexpected rearrangements have occurred.

Summary

This novel DF model system was developed to provide a tool that could lend
insight into which factors play a role in the AMR reversion in the FANCT patient. The
results of the experiments presented in this section provide evidence that the DF model
system can emulate the reversion of a FANCT duplication, in addition to being used as a
possible diagnostic tool for DNA recombination. It was observed that placement of a
DSB on either side of Alu 2 in the maternal duplication models provides the same
opportunity for reversion back to WT without bias. Additionally, when comparing the
effects of a DSB in context of the duplication model (DF2) vs WT model (DF5) it was
observed that the upstream Alu in the duplication model promotes HDR. When
considering known DNA repair pathways it can be hypothesized what type of
recombination may be involved in the cells from each quadrant, but further testing helps
to confirm our hypothesis. By using techniques such as flow cytometry, PCR of
polyclonal recombinant population and FACS single cell clone PCR, the results support
Figure 24: **FACS sorting and PCR of double negative quadrant.** (A) Map of ectopic locus in DF2 cell line with primer binding locations included (blue arrows). (B) Agarose gel of PCR products from the double negative quadrant FACS clones.
Figure 25: FACS sorting and PCR of green quadrant. (A) Map of ectopic locus in DF2 cell line with primer binding locations included (blue arrows). (B) Agarose gel of PCR products from the double negative quadrant FACS clones.
the previously stated assumptions on the type of repair that is occurring in the cells from each quadrant. Those conclusions are: double positive quadrant is a mixture of homology-dependent and -independent repair, red quadrant is homology dependent repair, green quadrant reflects homology-independent repair and the double negative quadrant is homology-independent (primarily) and -dependent repair.

**II. DNA repair proteins modulate reversion in FANCT model**

To aid in the understanding of what type of DNA repair is involved in the FANCT reversion, use of the DF2 cell line as a DNA repair reporter system can help elucidate which DNA repair pathways are involved in the DSB resolution. By impairing specific proteins involved in either HDR, non-HDR or both forms of repair, any change in repair pathway after an I-Sce1 induced DSB can be detected in the DF2 cell line by using flow cytometry to detect a modulation in fluorescence in the same quadrant when comparing the experimental condition to control condition (RTC).

**CtIP and RAD51 promote reversion of FANCT duplication**

The reversion of the duplicated allele in the FANCT patient’s hematopoietic cell lineage was hypothesized to be a homology directed repair event using nearby AluY elements as a source of homology. The prevailing theory for AMR is that a break occurs, followed by 5’ resection of the DNA ends which exposes 3’ ssDNA tails that can then seek homology either intra- or inter-molecularly (Kataoka et al., 2013). The required resection is governed by CDK dependent protein recruitment which is only active during S/G2 phases when the sister chromatid donor template is available. In vivo experiments provide evidence that DNA resection can continue for up to 3.5 kb in both directions from the DSB, which is within the limits of the DF model system to allow for
recombination of proximal or distal Alu elements (Ferretti et al., 2013; Tomimatsu et al., 2014; Zhou et al., 2014).

I chose to inhibit CtIP and RAD51 that act in different steps of homology directed repair to aid in determining the role of HDR in the FANCT genetic reversion. C-Terminal Binding Protein 1-Interacting Protein (CtIP) is at the very upstream segment of the homology directed repair pathway that begins with resection of DNA 5’ ends (Takeda et al., 2007). CtIP is a CDK dependent positive regulator of resection that interacts with the MRN complex (Wang et al., 2013). Previous analysis of PCR and FACS sequencing of the DF2 cell line recombinants reveal that the red quadrant recombinants are most likely due to HDR. If CtIP is required for resection, then inhibition of CtIP in the DF2 model system should decrease the number of events in the red quadrant. siRNA was used to target and inhibit CtIP in the presence of I-Sce1 expression in DF2 cells to assay for differences in DNA recombination (Figure 26). Results from this experiment reveal that differences in the red quadrant (64% RTC) and the double positive quadrant (124% RTC) are significant when comparing siCtIP samples to siCtrl samples. The double negative and green quadrants were not significantly different.

RAD51 is involved in an intermediate step in homology directed repair after the DSB and 5’ DNA resection (Khanna & Jackson, 2001). RAD51 will displace RPA from 3’ ssDNA ends and perform its primary function of strand invasion and identification of complementary homologous sequences in the sister chromatid (Sung & Robberson, 1995). The expectation is that by inhibiting RAD51, the cell cannot efficiently repair the I-Sce1 directed DSB by homology directed repair in DF2 cells, which results in fewer
Figure 26: Flow cytometry of siCtIP treated DF2 cells. (A) Map of ectopic locus in DF2 cell line. (B) Flow cytometry results of siCtrl (left) and siCtIP (right) DF2 cells 8 days post I-Sce1 transfection. (C) Diagram illustrates the percentage difference in fluorescent recombinants. (D) Western blot demonstrating CtIP knockdown. Percentages represent the number of events in siCtIP cells relative to siCtrl cells per quadrant. All of the data are expressed as mean±SEM. Percentages in red were determined to be statistically significant by using a paired t-test with Bonferroni correction. P < 0.0125 vs control.
events within the red quadrant. Inhibition of RAD51 was achieved by using the chemical inhibitor RAD51 B02, that disrupts binding to DNA and nucleoprotein filament formation (Alagpulinsa et al., 2014; Huang & Mazin, 2014), with parallel expression of I-Sce1. The results in this experiment indicate that there is significant difference between the red quadrant (49% RTC) and the double positive quadrant (136% RTC) when comparing the control to the experimental RAD51 inhibition (Figure 27). There were not statistically significant differences in the double negative or green quadrants.

The knockdown of CtIP and the inhibition of RAD51 function results in a decrease in the red quadrant, which indicates that CtIP and RAD51 promote the reversion of the FANCT duplication to WT by homology directed repair.

**DNA-PK does not promote FANCT reversion**

DNA-dependent protein kinase (DNA-PK) also works towards the upstream portion of DNA double strand break repair pathway, contributing to non-homologous end joining (NHEJ). After a DSB there is some competition by the MRN complex (HDR) and Ku70/80 (NHEJ) for broken DNA ends (Lamarche et al., 2010; Tomita et al., 2003). After Ku70/80 binds broken DNA ends, the DNA-PK catalytic subunit is activated and will help recruit Artemis, DNA polymerases μ and λ, XRCC4, DNA ligase IV, and XLF to the break for completion of NHEJ activities (Pawelczak et al., 2011). Nu7026 is a potent ATP competitive inhibitor of DNA-PK (Niazi et al., 2014; Willmore et al., 2004) and was used in conjunction with I-Sce1 on the DF2 cell line. Unlike the CtIP and RAD51 inhibition experiments, the DNA-PK inhibitor treatment did not affect the red quadrant. There was a significant decrease of cells present in the double negative (65% RTC) and green quadrants (76% RTC), in addition to an increase in the double positive
Figure 27: Flow cytometry of RAD51i treated DF2 cells. (A) Map of ectopic locus in DF2 cell line. (B) Flow cytometry results of control (left) and RAD51i (Rad51 B02 10 uM) treated (right) DF2 cells after I-Sce1 expression. (C) Diagram illustrates the percentage difference in fluorescent recombinants. Percentages represent the number of events in RAD51i cells relative to control cells per quadrant. All of the data are expressed as mean±SEM. Percentages in red were determined to be statistically significant by using a paired t-test with Bonferroni correction. P < 0.0125 vs control.
quadrant (120% RTC) (Figure 28). The lack of a statistically significant change in the red quadrant when impairing DNA-PK suggests that DNA-PK is not involved in the homology directed reversion of the FANCT duplication to WT, but rather promotes DSB repair in a homology independent manner.

**ATM affects both HDR and non-HDR events**

Ataxia Telangiectasia Mutated (ATM) protein exhibits protein kinase activity. In addition to checkpoint control, ATM is also significantly involved in repair of DNA double strand breaks. ATM is needed for full activation of DNA-PK activity to efficiently recruit DNA termini processing proteins required for NHEJ completion (Chen et al., 2007; Jiang et al., 2015). ATM is also required for successful repair of DSB by homologous recombination. After the DSB, ATM is needed to phosphorylate members of the MRE11-RAD50-NBS1 complex (MRN) that initiate the cascade of events to repair the DSB by homology directed repair (Gatei et al., 2000; Lavin et al., 2015; Uziel et al., 2003). Caffeine is an inhibitor of ATM protein kinase activity (Blasina et al., 1999) and was used in conjunction with transient expression of I-Sce1. Ku-60019 is a potent inhibitor of ATM, that ablates autophosphorylation and ATM activation (Golding et al., 2009), and was additionally used because of its high selectivity to help further differentiate ATM’s role in DSB repair.

Flow cytometry analysis of the caffeine treated DF2 cells results in a decrease of the red (76% RTC), double negative (56% RTC), and green (73% RTC); while double positives increased (133% RTC). In the KU-60019 inhibitor treated DF2 cells, the results are red (50% RTC), double negative (64% RTC), and green (80% RTC) as these quadrants display a decrease in percentages. Meanwhile, the double positive quadrant is
Figure 28: Flow cytometry of DNA-PK\textsuperscript{i} treated DF2 cells. (A) Map of ectopic locus in DF2 cell line. (B) Flow cytometry results of control (left) and DNA-PK\textsuperscript{i} (Nu7026 10 uM) treated (right) DF2 cells after I-Sce1 expression. (C) Diagram illustrates the percentage difference in fluorescent recombinants. Percentages represent the number of events in DNA-PK\textsuperscript{i} cells relative to control cells per quadrant. All of the data are expressed as mean±SEM. Percentages in red were determined to be statistically significant by using a paired t-test with Bonferroni correction. P < 0.0125 vs control.
increased (149% RTC) (Figure 29). The trends appear to be the same between the caffeine and the KU-60019 samples with the exception of the red quadrant in the Ku-60019 treated cells which has a significantly lower percentage than the caffeine treated cells, possibly attributed to Ku-60019 being more potent inhibitor of ATM. The results displayed in this section suggest that ATM promotes the reversion of the FANCT duplication to WT by homology directed repair. Additionally, it is observed that ATM also affects homology independent repair as displayed by the reduction in cells present in the double negative and green quadrants after inhibition of ATM function.

**FANCT is involved in HDR and non-HDR events**

The Fanconi anemia repair pathway is composed of at least 18 proteins (FANCA-FANCT), whose concerted effort can repair some forms of DNA damage. For proteins to be defined in the Fanconi anemia pathway, the patient cells are subjected to DNA interstrand cross-linking agents such as MMC or cisplatin and assayed for chromosome instability when compared to healthy patients (Pinto et al., 2009; Stoepker et al., 2015). FANCT, a UBE2T ubiquitin conjugase, is known to be required for resolution of interstrand crosslinks as part of the FA pathway, but participation in other DNA repair pathways is not yet known.

To determine the effect of FANCT depletion on DNA recombination after an I-Sce1 induced DSB, experiments were performed in both the U2-OS DR-GFP and DF3 cell lines. Additionally, these experiments will evaluate how well the newly developed DF model system can determine HDR when compared to the well-established DR-GFP system by comparing the shift in GFP fluorescence as detected by flow cytometry. The Hanenberg lab developed stable cell lines which eliminated the function of endogenous
Figure 29: Flow cytometry of ATMi treated DF2 cells. (A) Map of ectopic locus in DF2 cell line. (B) Flow cytometry results of control (left), ATMi (right) treated DF2 cells after I-Sce1 expression. (C) Diagram caffeine (left) and Ku-60019 (right) illustrates the percentage difference in fluorescent recombinants relative to control cells per quadrant. All of the data are expressed as mean±SEM. Percentages in red were determined to be statistically significant by using a paired t-test with Bonferroni correction. P < 0.0125 vs control.
FANCT by using CRISPR technology. All results in the U2-OS cell lines are after I-Sce1 expression under different experimental conditions. The term “no treatment” is simply the naïve state without CRISPR knockout or complementation. The other 6 experimental conditions have Cas9 nuclease stably integrated and constitutively expressed. The G2.2 is an sgRNA that targets the endogenous FANCT locus and “knocks out” the protein. Complementation experiments were performed with stable integration of a functional FANCT or the control (empty vector) EV plasmid which is the backbone in the functional FANCT plasmid (Figure 30). Values represent the percentage of GFP+ cells in experimental (columns 2-7) relative to control (RTC) naïve cell lines (column 1). The results indicate that there is significant reduction in HDR when compared to control (column 1) for FANCT null or FANCT null complemented with (empty vector) EV (columns 2&3). When FANCT null cells were complemented with FANCT (column 4) the HDR efficiency was restored to near normal levels. The CAS9 lentivirus (LC) alone (column 5), LC + EV (column 6) and LC + FANCT (column 7) cell lines were used as controls to assure results are due to only the FANCT knockout.

The DF3 cell line was analyzed by flow cytometry after I-Sce1 expression and the overall pattern of recombination was different between the FANCT+ and FANCT null cell lines (Figure 31). There was not a significant difference in the red quadrant RTC. In contrast, there is significant difference when observing the double positive, double negative and green quadrants RTC. When comparing the results of the DR-GFP system with the DF3 cell lines the GFP+ percentages are relatively close to each other (Figure32) when compared to their relative controls. This indicates that the DF3 recombination system is comparable to the widely used DR-GFP system in terms of HDR that is
**Figure 30: FANCT is involved in HDR.** Values represent the percentage of GFP+ in experimental cells relative to control (RTC) naïve (column 1) U2OS cell lines.

There is significantly less HDR in the FANCT null (column 2) and FANCT null + EV (column 3) than in all other experimental cell lines. The G2.2 FANCT knockout was then complemented by FANCT and HDR is restored to near naïve conditions (column 4). The presence of the CAS9 alone (column 5) or complemented with EV or FANCT (columns 6 & 7) did not reduce HDR levels relative to the naïve cell line. The data are presented as mean±SEM. Paired t-test was used for statistical analysis. * = P ≤ 0.05.
Figure 31: FANCT maybe involved in HDR and non-HDR events. All results are after an I-Sce1 induced DSB break in DF3 cells. (A) Map of DF3 construct. (B) Results of each of four quadrants comparing the DF3 FANCT null cell lines vs. FANCT+ control per quadrant. The data are presented as mean±SEM. Paired t-test was used for statistical analysis. * = P ≤ 0.05, ** = P ≤ 0.01.
Figure 32: Comparable percentage of GFP+ recombinants in FANCT null model systems.  
(A) Western blot demonstrating CRISPR knockout of UBE2T (FANCT) in DF3 and U2-OS DR-GFP cells.  
(B) Summary slide displaying the percentage of GFP+ cells RTC in the well-established U2-OS DR-GFP system (column 1) is comparable to the novel DF3 system (column 2) in FANCT null cells.
detected after anI-SceI induced DSB break. In addition, the DF3 model system has the capability to monitor other forms of DNA repair undetected in the DR-GFP system, such as: SSA, NHEJ and MMEJ.

**Summary**

The DF cell lines were constructed to provide insight into what DNA repair factors are involved in the FANCT reversion in the patient. This system can now be expanded past the question of what happens in the FANCT patient and be used to assess DNA repair more generally. Assessing the effects of inhibiting DNA repair proteins in the DF3 cell line reveals that each quadrant has the ability to report different types of repair outcomes, when compared to previous findings in literature. Comparison of the well-established DR-GFP HDR reporter to the DF3 cell line reveals nearly the same amount of reduction in HDR efficiency when FANCT is not present (~40% reduction). The DF3 cell lines have the novel ability to identify additional forms of repair, when compared to the DR-GFP system, that can be attributed to the system design.

**III. HDR is reduced when a DSB occurs adjacent to a DNA secondary structure forming sequence**

Microsatellites are repetitive DNA sequences that have the potential to be mutagenic and promote genomic instability (Bhargava & Fuentes, 2010; Brandström et al., 2008; Mirkin & Mirkin, 2007b). FANCA, BHDS and GT diseases all have a subset of patients where the mutagenesis can be attributed to a large AMR. Additional patients in each of the three diseases presented a different molecular etiology where a small insertion or deletion occurred around a microsatellite motif sequence located between
two Alu sequences (Benhammou et al., 2011; Gregory et al., 2001; Levran et al., 1997; Li et al., 1993). It is therefore reasonable to question if microsatellite sequences have the potential to induce DNA damage and/or affect DNA break repair. Previous studies have concluded that microsatellite sequences have the potential to expand or contract when the replicative polymerases are stalled with aphidicolin (Aph) or hydroxyurea (HU), but it is not definitive that the instability is due a microsatellite induced break in humans (Kim, Pytlos, & Sinden, 2006; Sinden et al., 2002). In this chapter, DF3 cell lines containing DNA secondary structure forming sequences were created to help address the microsatellite induced break and repair modulation questions. The DF microsatellite system provides a novel way of testing for microsatellite induced damage and repair in human cells by a shift in fluorescence. The experiments in this section will determine how repetitive DNA sequences influence Alu mediated homology directed repair.

Creation of model cell lines

This project utilized CTG/CAG trinucleotide (TNR) or a polypurine-polypyrimidine (Pu-Py) DNA sequence as template for secondary structure forming motifs. Our lab confirmed that an ectopic cell line containing a microsatellite with 102 CTG/CAG repeats and an upstream c-Myc replicator resulted in instability when challenged with replication inhibitors. Nascent strand abundance assays reveal that a select c-Myc sequence can initiate replication (Malott et al., 1999). The proximity of the c-Myc replicator to the TNR was also found to be imperative for instability in previous studies, therefore it will be included in this experimental framework (Liu, Chen, et al., 2012).
The DF3 construct was modified by adding the c-Myc replicator upstream of the first Alu sequence. Subsequently, the secondary structure forming sequences were placed into the original I-Sce1 recognition site that is kept intact at the 3’ end of insert (Figure 33). The sequences used in the TNR model system are a 100 CTG and an 84 CAG triplet repeats. The length of both of these repeats are large enough to have a propensity for instability (Brook et al., 1992). The secondary structure forming sequences used in the triplex model system are an 88 bp mirror repeat that is an asymmetric homogenous polypurine (TR) or polypyrimidine (TF) sequence. The construct name indicates which sequence (CTG, CAG, TF or TR) is used as template for lagging strand synthesis relative to the c-Myc origin. This allows for determination if replication polarity may also affect repeat instability (Figure 34).

**DNA repair is not detected in replication challenged model cell lines**

One reason that the specific ectopic repeat sequences were placed into the DF3 cell line system was to determine if there will be a modulation in fluorescence above the baseline c-Myc only construct. A modulation could potentially occur either spontaneously or induced by replication inhibitors. If the repeat sequence induces a DNA break and results in deletion of the DNA sequence containing a fluorescent marker, it can be detected by flow cytometry. There was not a detectable spontaneous drift of the control cell line fluorescence over the course of 3 months’ time.

In addition to surveillance of spontaneous fluorescence shift, all five model cell lines (Myc only, CTG100, CAG84, TR, and TF) were treated with 0.2 uM aphidicolin or 0.2 mM hydroxyurea replication inhibitors to investigate if a DNA break and subsequent recombination is detected. The replication inhibitors were constantly applied to cells for
Figure 33: Creation of repeat containing model cell lines. (A) The previously described DF3 cell line was modified by the addition of (B) a c-Myc core replicator core and a DNA sequence motif possessing the propensity of forming a DNA secondary structure (CTG$_{100}$ is used in this example).
Figure 34: Repeat motif model cell lines used. All cell lines in this chapter have a c-Myc core replicator integrated into the ectopic sequence and the nomenclature represents what sequence is in the lagging strand when replicated from the c-Myc origin. (A) Myc only contains only the c-Myc core replicator without any repeat sequence. (B) CTG\textsubscript{100} contains 100 CTG triplets. (C) CAG\textsubscript{84} contains 84 CAG triplets. (D) TR contains an 88 bp homopurine sequence and (E) TF contains an 88 bp homopyrimidine sequence.
seven days and then given four days after replication inhibitor treatment to recover before the flow cytometry was performed. The four days post treatment recovery allows the cells to return to the normal replication cycle and provides time for separation of cells into flow cytometry quadrants based upon cell fluorescence. The Myc only cell line is to be used as a baseline since it does not contain a downstream ectopic repeat sequence. The results demonstrate that in the presence of aphidicolin or hydroxyurea there is not a considerable shift in fluorescence from the double positive (red + green) population (Figure 35), indicating the absence of replication stress-induced breakage and recombination when compared to untreated cells.

The CTG$_{100}$ and CAG$_{84}$ cells lines were next subjected to the 0.2 uM aphidicolin or 0.2 mM hydroxyurea replication inhibitors. The CTG and CAG triplets in this expanded range are believed to form secondary structures in vivo (Liu et al., 2010; Sinden et al., 2002). The replication inhibitors foster a slowing of the replicative polymerase which facilitates the presence of more ssDNA at the replication fork, since the helicase is not impeded. The increase in ssDNA is predicted to afford more opportunity for the intramolecular secondary structures to form (Yang et al., 2003). The formation of secondary structures can lead to genomic instability by increasing DNA damage and breaks. It is therefore intuitive that in the ectopic cell lines containing the repeat sequences, there would be increased DNA breaks, and recombination that is detectable by a shift in fluorescence. Contrary to this hypothesis, the results did not display a considerable shift in fluorescence of the CTG$_{100}$ and CAG$_{84}$ cells lines when compared to baseline control (Figures 36 and 37). The lack of fluorescence shift may be attributed to an absence of DNA breakage. Alternatively, there could DNA breakage at
Figure 35: Treatment of DF3 Myc cell line with replication inhibitors.

(A) Diagram of ectopic insert present in the DF3 Myc stable cell line  (B) Graphical representation of recombinant fluorescence outcomes in each quadrant.  (C-E) Flow cytometry results of naïve, aphidicolin and hydroxyurea treated cells, respectively.
Figure 36: Treatment of DF3 Myc CAG<sub>84</sub> cell line with replication inhibitors.

(A) Diagram of ectopic insert present in the DF3 Myc CAG<sub>84</sub> cell line  
(B) Graphical representation of recombinant fluorescence outcomes in each quadrant.  
(C-E) Flow cytometry results of naïve, aphidicolin and hydroxyurea treated cells, respectively.
Figure 37: Treatment of DF3 Myc CTG_{100} cell line with replication inhibitors.

(A) Diagram of ectopic insert present in the DF3 Myc CTG_{100} cell line  (B) Graphical representation of recombinant fluorescence outcomes in each quadrant.  (C-E) Flow cytometry results of naïve, aphidicolin and hydroxyurea treated cells, respectively.
the repeated sequence that fails to repair, forming an acentric chromosome (Luzhna et al., 2013; Mateuca et al., 2006) containing no fluorescent marker genes. The fluorescent marker genes would then reside on the centromere side of the break.

Concomitant with the DF3 Myc TNR cell lines, the DF3 Myc TR and TF were also exposed to 0.2 μM aphidicolin and 0.2 mM hydroxyurea replication inhibitors. The expectation is much like the DF3 CNG cell lines, that when the replicative polymerase is slowed down by inhibitors it can facilitate uncoupling of the replication fork, because the helicase is unaffected. This will allow increased ssDNA at the replication fork that has the potential to form intra/inter-molecular DNA secondary structures in vivo. This opens the question of whether the secondary forming structures innately possess the ability to facilitate DNA breaks especially under replicative stress. Additionally, it is unknown how a dsDNA break within a secondary forming sequence, such as a triplex, would affect DNA repair. The DF3 Myc TF and TR cell lines can address these questions by placing the secondary forming sequence between Alu elements. This design allows for detection of a shift in fluorescence by flow cytometry, if recombination of Alu elements has occurred. Contrary to this line of reasoning, the results did not display a considerable shift in fluorescence of the TF and TR cells lines when compared to baseline control (Figures 38 and 39). The reason for the lack of fluorescence shift could be from the same factors outlined previously in the DF3 CNG section, such as formation of an acentric chromosome.

An additional cell line was created to add some possible insight to why a shift in fluorescence was not observed in the DF3 microsatellite cell lines. The DF3 microsatellite cell lines have the microsatellite sequence upstream of both fluorescent
Figure 38: Treatment of DF3 Myc TF cell line with replication inhibitors.
(A) Diagram of ectopic insert present in the DF3 Myc TF cell line  (B) Graphical representation of recombinant fluorescence outcomes in each quadrant. (C-E) Flow cytometry results of naïve, aphidicolin and hydroxyurea treated cells, respectively.
Figure 39: Treatment of DF3 Myc TR cell line with replication inhibitors.

(A) Diagram of ectopic insert present in the DF3 Myc TR cell line (B) Graphical representation of recombinant fluorescence outcomes in each quadrant. (C-E) Flow cytometry results of naïve, aphidicolin and hydroxyurea treated cells, respectively.
marker genes and an unrepaired break would leave both fluorescent marker genes together and thereby still have the potential to express both RFP and GFP. The DF2 CTG\textsubscript{100} cell line was created to place the microsatellite sequence between the fluorescent markers, and an unrepaired break at the microsatellite sequence would leave the potential that only one fluorescent protein is expressed and thus a shift in fluorescence can be observed. Since the ectopic site is targeted by FLP recombinase we are assured this ectopic cell line is in the same orientation and location as the DF3 microsatellite cell lines. The orientation was also confirmed by PCR after integration and selection. The DF2 CTG\textsubscript{100} cell line was subjected to the same regimen of aphidicolin (0.2 uM) and hydroxyurea (0.2 mM) for seven days and then four days recovery. The results were dramatically different than the DF3 microsatellite cell line results. The DF2 cell line is expected to lose eGFP and cells drift from the double positive to the red quadrant. If the proposed theory of acentric chromosome is valid in the DF2 CTG\textsubscript{100} model, the outcome of the replication inhibitor experiments should be a shift to green fluorescence due to the eGFP being still linked to the centromere. The result of replication inhibitor treatment is a dramatic shift toward the green quadrant for both aphidicolin and hydroxyurea treated cells. These results provide supporting evidence toward the formation of an acentric chromosome that will lose the Tom gene, but retain the eGFP gene (Figure 40).

**Secondary structure forming sequences adjacent to a DSB modulate HDR.**

The same five cell lines used in the replication inhibitor studies DF3- (Myc only, CTG\textsubscript{100}, CAG\textsubscript{84}, TR, and TF) will also be used for this experiment. Whilst the previous experiments focused on a spontaneous or replication inhibitor induced break, this experiment utilized the I-Sce1 enzyme to induce a DSB at the I-Sce1 recognition site
Figure 40: Replication inhibitors induce recombination in DF2 CTG\textsubscript{100} cell line.

(A) Map of the DF2 CTG\textsubscript{100} construct. (B) Expected outcomes of flow cytometry quadrants. (C) Control without replication inhibitors. (D) Treatment with 0.2 uM aphidicolin. (E) Treatment with 0.2 mM hydroxyurea.
located immediately adjacent to the secondary structure forming sequence in the ectopic insert. Unlike the replication inhibitor experiments that could produce a DSB anywhere within the secondary structure forming sequence, the I-Sce1 induced break experiments will result in a DNA break at a specific location immediately juxtaposed to the secondary structure forming sequence. The reasoning behind this round of experiments is that the I-Sce1 enzyme will induce a DSB at the recognition site, followed by bi-directional 5’ DNA resection that could be impeded by the secondary structure forming sequence itself or proteins associated specifically to microsatellites regions. This modulation in recombination can be assayed in the DF system when comparing fluorescence of experimental cell lines to the control. The expectation is that there would be fewer GFP+ and double negative recombinants if resection is impeded due to the Alu mediated recombination obstruction. Previous work has demonstrated that structure forming A-rich sequences, from a common fragile site, block repair (Burrow et al., 2010; Wang et al., 2014). The results support the assumption as fewer double negative and GFP+ recombinants are detected in the A84, T100 (Figure 41) and TF, TR (Figure 42) cell lines relative to the DF3 Myc controls. RFP+ recombinants are not expected due to the placement of the I-Sce1 recognition site upstream of the Tomato marker gene.

Summary

The experiments performed in this section aim to determine the influence DNA secondary forming sequences can have on DNA breaks, recombination and repair. The DF3- (Myc, CTG100, CAG84, TR, and TF) cell lines that were subjected to replication inhibitors exhibited no shift in fluorescence when compared to the Myc only control cell line, suggesting that the secondary forming sequences do not possess increased sensitivity
Figure 41: TNR sequences juxtaposed to an I-SceI DSB reduces HDR. Flow cytometry results of the TNR containing cell lines CAG84 (A84), CTG100 (T100) relative to the DF3 Myc control (c-myc only) cell line for the (A) double negative and the (B) GFP+ quadrants. The data are presented as mean ± SEM. Paired t-test was used for statistical analysis. ** = P ≤ 0.01.
Figure 42: Triplex forming sequences juxtaposed to an I-SceI DSB reduces HDR.

Flow cytometry results of the triplex forming sequence containing TF and TR cell lines relative to the DF3 Myc control (c-myc only) cell line for the (A) double negative and the (B) GFP+ quadrants. The data are presented as mean±SEM. Paired t-test was used for statistical analysis. ** = P ≤ 0.01.
to DNA breaks in the presence of replication inhibitors. Upon further evaluation, it is possible that the microsatellite containing DF3 model system design may not be able to detect DNA breaks because of the fluorescent markers’ gene positioning, relative to microsatellite sequences. Indeed, the flow cytometry results of the DF2 Myc T100 cell line with replication inhibitors, suggests that there may be DNA breaks occurring in the DF3 repeat models but that it is not detected due to system design. If the DNA breaks occurring at satellite sequences are recalcitrant to repair, it could result in an acentric chromosome fragment. The DF3 cell line designs have the secondary forming sequences upstream of the fluorescent marker genes, leaving them linked on the centromeric chromosomal segment. The DF2 Myc T100 cell line is designed so that the satellite sequence interrupts the fluorescent marker genes. In this case, after an unresolved DNA break at the satellite sequence, the Tom gene could segregate with an acentric chromosome fragment and only the eGFP gene replicate after mitosis.

A decrease in recombination was observed in the DF3- (CTG_{100}, CAG_{84}, TR, and TF) cell lines when compared to the DF3 Myc control, when extrinsically facilitating a DNA break. For these experiments, I-Sce1 was expressed in the multiple cell lines and targets the I-Sce1 recognition site juxtaposed to the secondary structure forming sequence present in the ectopic segment. This will produce a break that may capable of eliciting recombination and a shift in fluorescence in the model cell lines. The results indicate a decrease in homology and non-homology directed DNA repair within the cell lines containing the secondary forming sequences, when compared to the DF3 Myc control cell line, after an I-Sce1 directed DNA break. These results suggest that in human cells,
DNA secondary structures can inhibit DNA recombination after a DSB near a microsatellite motif.
DISCUSSION

The data presented here suggest that the novel DF model system can be used for detecting DNA repair patterns. By comparing the percentage of HDR repair after a DSB, it became clear that the novel DF system could closely recapitulate the well-established U2-OS DR-GFP system. With impairment of DNA repair proteins involved in decided repair fates, it was possible to indirectly define the recombination outcomes for the DF system flow cytometry quadrants. FANCT, a UBE2T ubiquitin conjugase, was recently discovered to be involved in Fanconi anemia when deficient. This discovery established FANCT’s importance in resolving interstrand crosslinks via the FA pathway, FANCT’s involvement in other types of DNA repair is unknown. The U2-OS DR-GFP system was used to define the effect FANCT has on HDR after an I-Sce1 directed DNA break. FANCT was also impaired in the DF system and yielded quantitatively similar HDR results, in addition to information about other types of DNA repair that FANCT may influence. Lastly, by adding microsatellite sequences to the DF system, it provides a novel model system to test the effects non-canonical DNA structures can exert on either producing DNA strand breaks or influencing which repair pathway is utilized in human cells.

Novel DF model system can emulate the genomic reversion displayed in FANCT patient cells
The primary reasoning for developing the DF model system was to elucidate factors involved in the FANCT genomic rearrangements of hematopoietic cells present in a FANCT patient. The reversion of the FANCT maternal duplication in the FA patient is thought to occur from the homology directed repair of a DNA break that utilizes the nearby Alu elements for recombination. DF (1-3) cell lines were created to emulate the partial duplication found in the maternal allele. It is hypothesized that the partial duplication was reverted back to WT as a result of an HDR event using the Alu elements as template. An outstanding question is whether there is a recombination bias when all three Alu elements are present within the locus. Therefore, an I-Sce1 recognition site was placed downstream of the central Alu in the DF2 cell line and upstream in the DF3 cell line. Since the reversion to WT could occur by a break and recombination on either side of the central Alu our goal was to determine if there is a bias to one side.

The results indicate that there is nearly the same opportunity for a proximal Alu mediated deletion event on either side of the central Alu element when comparing the recombination pattern of the DF2 and DF3 cell lines by flow cytometry. This indicates that the initial break that induced the patient hematopoietic cell back to WT FANCT could have occurred between either Alu1/2 or Alu2/3 and repaired with close to the same efficiency. DF4 and DF5 cell lines were created to determine if the WT deletion in the FANCT patient’s paternal allele occurs with the same efficiency as a deletion in the duplication model. DF2 and DF5 are the exact same ectopic sequence except for removal of the upstream Alu element in the DF5 cell line. Following an I-Sce1 induced DNA break in DF2 and DF3 experimental cell lines, there are approximately 53-56% cells remaining in the double positive population. It would therefore be plausible to expect the
same amount of proximal Alu mediated recombination between the DF2 and DF5 cell lines. Surprisingly, there was less than half the amount of proximal Alu mediated recombination in the DF5 model when compared to the DF2 cell line results. Additionally, there was a sharp increase in the double negative population of the DF5 cells. This would indicate that there is less proximal AMR and a shift towards a presumable non-HDR event when the upstream Alu is not present. In summary, the results of this section provide evidence that a reversion of the nonfunctional partially duplicated maternal allele back to a functional WT allele is more likely than the conversion of a WT allele to a FANCT deleted paternal allele.

**Analysis of the DF2 cell line reveals unique patterns of recombination**

The original hypothesis of what type of DF2 cell line recombinants would be present in each flow cytometry quadrant was that the RFP quadrant would reflect proximal Alu recombination and the double negative quadrant would reflect distal Alu recombination. There were no expectations for the GFP quadrant, as there is no canonical repair mechanism to recombine and lose red fluorescence but not green fluorescence. The remaining double positive quadrant population were expected to comprise of; no DNA break, precise re-ligation, perfect HR repair or NHEJ (Kühne et al., 2003; Lin et al., 1999; Vu et al., 2014). Unexpectedly, there was a small population of green cells detected by flow cytometry. This led to a need to further analyze the different quadrants and obtain a more detailed picture of what type of recombination is present in each quadrant. Sequencing and classical PCR could not be used to answer this question because of the heterogeneity of the cell population. It was decided to use FACS to process the recombinant cell population by single cell sorting for each quadrant. By
using this technique, it was possible to propagate each individual clone and have enough cells for further assays. This made it possible to have pure populations that could then be used for PCR and DNA sequencing to provide insight as to what type of recombination is present in each quadrant.

A random selection of single cell clones from each quadrant were next subjected to PCR analysis to first determine if a PCR product could be detected. The upper PCR primer was located at the upstream portion of the ectopic sequence and the lower PCR primer was located in the TK portion of the acceptor cell line DNA. The PCR results from 10 single cell sorted clones of the red quadrant all produced a product of expected size if proximal Alu recombination had occurred. Sequencing revealed that the recombinants had undergone a proximal Alu mediated recombination. This validates that the DF system can recapitulate the reversion witnessed in the FANCT patient cells. The double positive quadrant contained homology and non-homology directed repair outcomes. For the double negative and green quadrants, there was a mix of wild-type and unanticipated (deletions and inversions) PCR products as detected by DNA sequencing. Thus, although the major pathways of DSB recombination have been deduced, our results show a much greater complexity of recombination outcomes when a single chromosome locus was monitored.

**Impairment of DNA repair proteins CtIP and Rad51 confirm red quadrant recombinants are due to an HDR event**

To further examine factors impacting a genetic reversion, canonical DNA repair proteins that are involved in specific types of repair were disrupted and the modulation of flow cytometry patterns analyzed. To test the effect that homology directed repair plays
in FANCT reversion, it was decided to decrease the activity of two HR proteins, CtIP and Rad51. CtIP is vital for initiation of 5’ DNA resection that is required for homology directed repair and works at the beginning of the HDR process (Polato et al., 2014; Takeda et al., 2007). Rad51 is also vital to HDR and is involved in the middle step of the repair process by assisting with homology search and donor template strand invasion that is required for homologous recombination (San Filippo et al., 2008; Sung et al., 1995). Impairment of either of these proteins should reduce HDR and the expectation is it will reduce the number of recombinants present in the red quadrant after an I-Sce1 directed DNA break. The results reveal that inhibition of both CtIP and Rad51 reduced the number of red quadrant recombinants to 64% and 49% relative to control, respectively. In combination with the PCR and sequencing results of red quadrant recombinants, these results provide evidence that the DF system is capable of detecting the HDR events responsible for the FANCT reversion in patient cells.

**Inhibition of DNA-PK confirms that double negative and eGFP quadrant recombinants are primarily derived from non-homology dependent repair events**

The PCR and sequencing results from the FACS clones for the double negative and eGFP quadrants indicate that those quadrants primarily represent non-HDR events. Inhibition of a DNA repair protein known for its involvement in NHEJ was chosen to further examine the validity of that hypothesis. DNA-PK is immediately recruited to a DSB in vivo (Uematsu et al., 2007). DNA-PK binds the broken DNA ends and then facilitates recruitment of additional DNA processing enzymes to complete repair in a non-homology mediated manner (Davis et al., 2013; Pawelczak et al., 2011b). The result
of inhibiting DNA-PK concomitant with an I-Sce1 directed DNA break was a decrease in double negative and eGFP recombinants to 65% and 76% relative to control. The results of PCR, sequencing results and NHEJ inhibition provide validity that the double negative and eGFP flow cytometry quadrants represent recombinants derived primarily from non-HDR events.

**DF system is capable of detecting homology dependent and non-homology dependent repair events simultaneously**

The previous experiments have displayed the ability of the DF system to detect homology dependent or non-homology dependent events independently, but to reach full potential as model system of DNA recombination it must be capable of detecting both types of repair events simultaneously. The protein kinase ATM was targeted for inhibition as a means of achieving this goal. ATM works soon after a DNA break and is needed for phosphorylation and activation of both CtIP and DNA-PK (Jiang et al., 2015; Wang et al., 2013). For ATM inhibition purposes, caffeine and an ATM specific inhibitor were used. Caffeine inhibits ATM function but it is not just specific to ATM and also affects other protein kinases, such as ATR (Sarkaria et al., 1999). The result of using caffeine to inhibit ATM was a decrease of recombinants in the red, double negative and eGFP quadrants by 24%, 44% and 27%, respectively. Since caffeine is promiscuous to other protein kinases, a very specific inhibitor of ATM was also independently used. That experiment also resulted in a decrease of recombinants in the red, double negative and eGFP quadrants by 50%, 36% and 20%, respectively. While both methods of inhibition resulted in a decrease of both homology dependent and non-HDR events, the more specific inhibition of ATM appears to preferentially affect the red quadrant and
homology directed repair. The greater reduction of non-HDR events in the caffeine experiment could be attributed to an inhibition of another protein kinase outside of ATM that is preferentially involved in non-homologous repair. The results of this experiment confirm that the DF system is capable of detecting homology dependent and non-homology dependent repair events simultaneously.

The repair pattern of the double positive recombinants was not determined under DNA repair inhibition conditions

An issue not yet addressed is why the double positive quadrant recombinants increase after DNA repair protein inhibition regardless of homology dependent or non-homology dependent event targeting. PCR and sequencing result analysis of the double positive quadrant recombinants were performed after an I-SceI directed DNA break, but under normal repair conditions. The expected recombinants in the double positive quadrant are; no break occurred, precise re-ligation with no DNA end processing, perfect HR and limited NHEJ or MMEJ. The PCR and sequencing results indicate that either no break occurred, precise re-ligation with no DNA end processing or perfect HR outcomes predominate with a minority of non-homology dependent events (i.e. NHEJ). It can be assumed that the amount of I-SceI digestion is comparable under different experimental conditions considering that the standard deviations between independent trials were very small which would indicate a consistency of DNA breaks in all experiments. All compared control and treatment conditions were performed with the same batch of enzymes and materials to facilitate consistency. This should ensure that the “no I-SceI induced break” levels remain consistent throughout all experiments.
The ratio of repair outcomes in the double positive quadrant will certainly change depending on what the experimental condition is. Unfortunately, the “no I-Sce1 induced break”, precise ligation without processing and perfect HR will all present with the same genomic configuration. While it was outside the scope of this project, it would be possible to generate some quantitative numbers of homology dependent repair vs non-homology dependent repair if next generation sequencing was employed. FACS of the double positive quadrant would be required, followed by NGS and analysis. An attempt was made to decipher changes in the homology dependent vs non-homology dependent repair of the treatment vs control conditions, by performing PCR of the polyclonal population with primers on each side flanking the I-Sce1 restriction site. After PCR, the products were subjected to in vitro digestion by I-Sce1. The theory is that the NHEJ events would then be resistant to in vitro digestion and the percent change of treatment conditions could be quantified from product intensities on a gel. The results of this experiment were inconclusive. Perhaps an alternative to this method would be single droplet PCR but that technology was not available to the investigator. Due to the uncertainty of the type of DNA repair found in the double positive quadrant, this project primarily focused on the better defined repair outcomes of the red, green, and double negative quadrants.

**FANCT is involved in homology dependent and non-homology dependent DNA repair**

FANCT was recently implicated in the etiology of a Fanconi anemia patient with genomic rearrangements of both FANCT alleles (Virts et al., 2015). The FA pathway has been characterized in the resolution of deleterious interstrand crosslinks (Wang &
Smogorzewska, 2015). FANCT is the E2 ubiquitin conjugase that performs the transfer of ubiquitin to the E3 ubiquitin ligase FANCL that will ubiquitinate the FANC I/D complex (Rickman et al., 2015b). It is currently unknown if FANCT is involved in other DNA repair events outside of the FA pathway. The well-established DR-GFP cell reporter system and the newly defined DF system were both employed in an attempt to answer the question of FANCT’s role in additional DNA repair pathways. The DR-GFP reporter system is widely used as reporter system to assay for homology directed repair (Moynahan & Jasin, 2010). The major drawback of this system is there is only a fluorescent readout for HDR events, not non-homology dependent events. The DF system is capable of defining both HDR and non-HDR events using fluorescence detection by flow cytometry. This experiment was designed with two questions in mind. First, does the DF system compare similarly to the established DR-GFP system in its ability to detect HDR events. Second, is to determine if FANCT is involved in HDR and non-HDR DNA repair by use of the DF system.

To achieve the goal of determining FANCT’s role in DNA repair, both the DR-GFP and DF3 cell lines were subjected to CRISPR mediated knockout of FANCT. For the DR-GFP system HDR was reduced to approximately 60% relative to the control cell line after an I-Sce1 mediated DNA break. HDR was restored to near control levels by adding back a vector expressing FANCT which confirms that the effect displayed was due to loss of FANCT. In the DF3 cell line, the results also revealed a decrease in green fluorescence of approximately 60% relative to the control cell line after an I-Sce1 mediated DNA break. Additionally, the DF system displays an approximate 50%
decrease in double negative recombinants when compared to the control after an I-Sce1 induced DNA break.

The results from this experiment demonstrate several things. First, the DF system is comparable to the well-established DR-GFP system when assessing the ability to detect HDR events. Second, the findings here reveal that FANCT can play a role in homology dependent and non-homology dependent DNA break repair. This repair is independent of the interstrand crosslink resolution function in FA pathway since no crosslinking agents were used.

**DNA repair is not detected in replication challenged DF3 microsatellite model cell lines**

The previous experiments have established the DF system as capable of emulating the reversion pattern witnessed in the FANCT patient. It is unknown exactly where and why the initial DNA break occurred in the FANCT patient, only that the patient had encountered a reversion of the maternal duplication to WT due to Alu mediated recombination in the hematopoietic linage. Genomic reversions have occurred in other Fanconi anemia patients and in some subsets the breakpoint can be narrowed down to a microsatellite motif (Benhammou et al., 2011; Gregory et al., 2001; Levran et al., 1997). Therefore, it would be pertinent to evaluate the effect of a microsatellite sequence within the ectopic site of the DF system. If the microsatellite sequence can initiate a DNA break, then subsequent repair can be monitored by flow cytometry. Previous research has revealed that microsatellite sequences can be a source of instability in the form of DNA expansions or contractions (Usdin & Grabczyk, 2000). While genomic instability of microsatellite sequences are well documented, exactly how the instability occurs in
human cells is not unanimously accepted. It has been posited that instability can occur due to DNA replication, transcription coupled repair or other DNA repair processes (Liu et al., 2009; McMurray, 2010; Nakamori et al., 2011). Replication inhibitors have been used to help induce genomic instability, presumably by increasing single strand DNA at the replication fork that lends more opportunity for non-B DNA secondary structures to form (Liu et al., 2012). With the DF3 TNR and Pu/Py model cell lines it is possible to test if the use of replication inhibitors can elicit a DNA break and recombination event involving the microsatellite sequence. The c-Myc replicator was also added to the cell lines because previous studies have demonstrated that instability is more likely the closer the microsatellite sequence is to a replication origin (Liu, 2010). Additionally, the c-Myc replicator allows for a determination if replication polarity will influence the outcome.

Aphidicolin (0.2 uM) and hydroxyurea (0.2 mM) were utilized as replication inhibitors for this experiment. The low concentrations of the replication inhibitors aphidicolin (Casper et al., 2002; Glover et al., 1984b; Schwartz et al., 2006) and hydroxyurea (Chen et al., 2015) (Petermann & Helleday, 2010b) will slow the replicative polymerases but not arrest replication. Both were constantly in media for seven days and then the cells recovered for four days and harvested for analysis. The results from this experiment did not reveal any shift in fluorescence of the CAG_{84}, CTG_{100}, TR or TF cell lines when compared to the c-Myc only control cell line. Initially, this was an unexpected result considering the amount of literature that substantiate microsatellite sequences as points of instability. Although the genomic rearrangement of microsatellite sequences are rare events in vivo, the hope is that the highly sensitive flow cytometry
assay could detect even small changes. By microscopic evaluation there was marginal additional cell death observed (<5%) in the replication inhibitor treated, microsatellite containing, cell lines when compared to the c-Myc only cell line.

In the replication inhibitor experiments it is unknown where a break may occur within the microsatellite sequence, and there is a possibility that after a DSB the DNA ends will not be “clean” (readily ligatable). The broken DNA ends could be “dirty” and would require more processing enzymes to repair the DNA break (Boiteux & Jinks-Robertson, 2013; Kanaar et al., 2008). Combined with the possibility that the microsatellite sequence can intramolecularly “fall back” on itself once single stranded and form a structured end, even more processing enzymes may need to be involved. This could impede repair and detection to an already rare event in the replication inhibitor experiments.

An alternative hypothesis may be able to explain why no genomic instability was detected in the DF3 microsatellite models after replication inhibitor treatment. The outcome may be influenced by where the centromere is located relative to the ectopic sequence. If the centromere is located downstream of the ectopic sequence then the fluorescent markers could be linked. The end result after an unrepai red DNA double strand break in the microsatellite sequence is the upstream portion (telomeric) of the chromosome could become acentric after mitosis. The downstream portion of the broken chromosome, containing both color marker genes would migrate with the centromere and the cell could appear as double positive because both fluorescent markers are present.

The construction of the DF2 CTG<sub>100</sub> cell line, which placed the microsatellite sequence between the fluorescent markers can be used as tool to elucidate a possible
explanation for the lack of genomic instability detection (Gadgil et al., 2016). When the DF2 CTG\textsubscript{100} cell line was subjected to the same regimen of aphidicolin (0.2 uM) and hydroxyurea (0.2 mM) the results were dramatically different than the DF3 microsatellite cell line results. The DF2 cell line is expected to lose eGFP and the cells would drift from the double positive to the red quadrant. If the proposed theory of acentric chromosome is valid then in the DF2 CTG\textsubscript{100} model, the outcome of the replication inhibitor experiments should be a shift to green fluorescence due to the eGFP being still linked with the centromere. The result of replication inhibitor treatment is a dramatic shift toward the green quadrant, therefore supporting the theory of an acentric chromosome (Figure 40). More validation will need to be performed before it is certain the interpretation of this result is correct. The cells could be FACS sorted and single cell clones grown that could then be used for FISH assays to map the centromere with the Tom and eGFP fragments.

**Secondary structure forming sequences influence DNA repair patterns in human cells**

The results of the DF microsatellite cell line experiments can provide additional depth to the current knowledge of what happens when a DNA break occurs within or adjacent to a microsatellite sequence in human cells. The DF3 and DF2 microsatellite cell line experiments provide evidence to infer that a DNA break within or near a microsatellite sequence may be recalcitrant to repair. The model cell lines were created so that an I-Sce1 recognition site was located immediately downstream of the microsatellite sequences. The hypothesis is that after the I-Sce1 induced DSB and 5’ resection, it will expose ssDNA of the microsatellite sequences that has the possibility to
form secondary structures and structured DNA ends. Just the secondary structure alone could form a physical impediment or the secondary structure could preclude canonical DNA repair proteins from performing their functions. The results of this experiment indicate that secondary forming sequences juxtaposed to a DNA break in human cells can influence recombination patterns. This is consistent with previous work that shows structured DNA can inhibit the processing of DNA double strand breaks (Bikard et al., 2010; Wang & Vasquez, 2014; Wang et al., 2014).

In the trinucleotide repeat cell lines there is significant reduction of recombinants in the double negative and green quadrants after an I-Sce1 induced DNA break of the CAG84 and CTG100 cell lines (68% and 75% green, 57% and 60% double negative) respectively, when compared to the DF3 Myc cell line. The slightly higher decrease in recombinants of the CTG100 cell line may stem from the fact that formation of CTG is favorable to CAG secondary structures due to less steric hindrance of the pyrimidine (T) relative to purine (A) or simply just more repeats in the CTG100 cell line. The decrease in double negative and green quadrant recombinants for either cell line may be explained by the resection process. Among several explanations for this result, one is that once resection has occurred, the exposed ssDNA TNR tract forms secondary structure that impedes repair proteins or the repair process; the second is that when the helicase separates the DNA strands allowing for resection, it offers opportunity for secondary structure formation that can impede the actual resection process and may cause it to abort. Either of these scenarios could interrupt the long range resection process and result in an unrepaired DNA break or restrict to short range resection and NHEJ. Both of these outcomes could result in a double positive cell that retains both fluorescent markers.
The triplex forming TF and TR models also exhibit a reduction in the number of double negative and green quadrant recombinants when compared to the DF3 Myc control cell line, after an I-Sce1 induced DNA break. The TF cell line had a higher reduction of recombinants than any of the experimental cell lines (86% green and 76% double negative) RTC. In comparison, the TR cell line exhibited less of a reduction than the TF cell line, with a reduction of (66% green and 52% double negative) when compared to the c-Myc only control cell line. From a strictly steric hindrance point of view the Py-rich strand should form secondary structure easier than the Pu-rich strand simply due to less bulky structures. In contrast, previous results have displayed the tendency at physiological pH is for the Pu-rich strand to preferentially form DNA secondary triplex and G-quadruplex structures (Inoue, 2005). This theory was substantiated from a replication viewpoint with primer extension and replication blockage experiments (Liu, Myers, et al., 2012). It is unclear exactly what DNA secondary structures the Pu-Py sequence will conform to when a juxtaposed DSB occurs in human cells. While a Py-rich third strand triplex is not thought to occur under physiological conditions with intact DNA strands, it has been observed that after a DSB, the 3’ ssDNA of the Py-rich strand and the 5’ ssDNA of the Pu-rich strand will preferentially fold back to form secondary structures \textit{in vivo} (Jain, Wang, & Vasquez, 2008). It has also been observed that a DSB will contribute to a favorable relaxed DNA supercoiling condition as to promote triplex DNA structures in Saccharomyces cerevisiae (Wang & Vasquez, 2014). The results of these experiments can provide a framework by which the outcome of the TF and TR cell line assays can be explained. After I-Sce1 produces a DSB, bidirectional 5’ DNA resection occurs that will expose the Py-rich (TF) or the Pu-rich
(TR) 3’ DNA sequence. The TF cells would therefore produce a 3’ Py-rich tail that could fall back on the still double stranded duplex DNA forming a triplex structure. This triplex structure could then prevent the recruitment of canonical DNA repair proteins and possibly even disrupt longer range resection. The disruption of longer range resection could potentiate multiple outcomes such as a short range non-HDR event or an unrepaired acentric chromosome. Both of these outcomes would result in the possible retention of both fluorescent markers and present in the double positive quadrant.

While the results of the DF microsatellite and I-Sce1 experiments display a definite modulation in recombination patterns, more assays would need to be performed to tease apart all recombination fates. As alluded to earlier, after I-Sce1 expression, there would be a need to separate the heterogeneous cell population by fluorescent quadrants. Once the recombinant cells are separated by fluorescence, the next step would be to define the different DNA recombination patterns from each quadrant. Single cell sorting by fluorescence followed by DNA sequencing could achieve this aim, but takes considerable time to grow individual cells into colonies in order to generate enough starting material. An alternate, quicker, method would be to perform droplet PCR and sequencing or next generation sequencing on the recombinant cells from each quadrant. The goal would be to generate enough data to provide quantitative statistical relevance to the DNA recombination outcomes.
FUTURE DIRECTIONS

The work presented in this project provides a basis as to what factors may lead to the FANCT genomic reversion presented in the hematopoietic cells of a FANCT patient. The novel DF cell line model system has potential to be a very powerful system to detect what type of recombination is occurring under differing experimental conditions. The experiments performed within this project have helped define the different DNA recombination possibilities for each quadrant but more work needs to be completed to ensure the results are not hyperbolized. It is possible that any decrease of cells within a quadrant, relative to control, could be due to cells within that quadrant dying off during the recovery period. To address this possibility, cell viability assays and next generation sequencing can be performed at different time points throughout the recovery period to further refine what type of recombination results are presented for each flow cytometry quadrant. Since the FANCT reversion occurred preferentially in the hematopoietic cell lineage of the patient, it would be prudent to also modulate levels of hematopoietic specific proteins in the DF cell lines. AID and RAG recombinase are both known to induce recombination in hematopoietic cells, and their activity could possibly potentiate the reversion observed in the FANCT patient. The DF microsatellite cell line models also need to be probed further to determine if the results of the DF3 microsatellite cell lines are due to lack of repeat instability, or if it is just difficult to detect with the design of the system. It is possible that the DNA break is especially difficult to repair and cells will die off or form an acentric chromosome and only present with one fluorescent marker. The DF2 CTG100 cell line results indicate that an acentric chromosome is a definite possibility and cell viability and FISH assays can be performed to address these issues.
CONCLUSION

This project presents data that FANCT’s role in DNA repair expands beyond resolution of interstrand crosslinks via the FA pathway. The experimental DF model system was successful in emulating the Alu mediated recombination found in the genomic reversion of the FANCT patient. The novel DF model system has shown to be equivalent to the well accepted U2-OS DR-GFP model system for detection of homology directed repair and was developed to assist in the interrogation of FANCT’s role in repair. Results from this project also demonstrates the ability of the DF model system to be used as method for detection of both homology and non-homology directed repair. The DF3 microsatellite cell line experiments provides evidence that a DNA break near a microsatellite element can modulate the DNA recombination outcomes. This project expands the current knowledge of FANCT and DNA microsatellite roles in DNA repair fates, in addition to developing a diagnostic tool for simultaneous examination of homology and non-homology directed DNA repair.


