TRANSCRIPTIONAL REGULATION OF SOX5 AND WWOX BY FANCD2

Kathryn Henrikson
University of Rhode Island, khenrikson@uri.edu

Follow this and additional works at: https://digitalcommons.uri.edu/theses

Recommended Citation
Henrikson, Kathryn, "TRANSCRIPTIONAL REGULATION OF SOX5 AND WWOX BY FANCD2" (2021). Open Access Master's Theses. Paper 1965.
https://digitalcommons.uri.edu/theses/1965

This Thesis is brought to you for free and open access by DigitalCommons@URI. It has been accepted for inclusion in Open Access Master's Theses by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons@etal.uri.edu.
TRANSCRIPTIONAL REGULATION OF SOX5 AND WWOX BY FANCD2

BY

KATHRYN HENRIKSON

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN CELL AND MOLECULAR BIOLOGY

UNIVERSITY OF RHODE ISLAND

2021
MASTER OF SCIENCE

OF

KATHRYN HENRIKSON

APPROVED:

Thesis Committee:

Major Professor  Niall Howlett
Claudia Fallini
Angela Slitt

Brenton DeBoef
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2021
ABSTRACT

Fanconi anemia (FA) is a rare genetic disease characterized by an increased risk for bone marrow failure, leukemia, and premature cancers (Alter et al., 2018). The FA pathway is involved in the repair of DNA damage such as stalled replication forks and DNA interstrand crosslinks (ICL) (Feng et al., 2019; Schlacher et al., 2012). It has been previously seen that under conditions of replication stress, FANCD2, a key protein in the FA pathway, binds to and traverses large actively transcribed genes (Okamoto et al., 2018). Using U2OS 3xFLAG cells we seek to answer the question of why FANCD2 is binding to these genes and if by doing so is it acting as a regulator of transcription? To answer this question, we used a candidate gene approach and chose WWOX and SOX5 and observed if under conditions of aphidicolin (APH), there were changes in their protein and transcript levels. We also observed if the absence of FANCD2 would affect their protein and transcript levels. In this study, it was found that under conditions of replication stress there was little change in both the protein and transcript levels of SOX5 and WWOX. Interestingly, the analysis of an RNA sequencing experiment brought to light the potential for FANCD2 to be involved in neuronal development as it was found that there were significant differences in normalized read counts in a subset of neuronal genes in the absence of FANCD2.
ACKNOWLEDGMENTS

I would like to thank my graduate advisor Dr. Niall Howlett. He believed in me and supported me throughout my two years in his lab. His guidance was helpful in my research as well as the writing of this thesis. The opportunity he gave me will help me throughout my career both professionally as well as personally.

I would also like to thank the rest of my thesis committee, Dr. Angela Slitt, and Dr. Claudia Fallini for their thoughtful and supportive feedback as well as their willingness to participate in the process of my master’s thesis. I would also like to thank Dr. Steven Irvine for agreeing to be my committee chair. I want to thank my current lab mate Justin Blaize for his support as well as previous lab members Juan Cantrez-Valez and Eden Francoeur who trained me during my first year of graduate school. I would like to thank Janet Atoyan, Kathleen Tran, Tomali Chakravarty, Dr. Kathryn Ramsey, and Hannah Trautmann, for their guidance and feedback while I learned how to perform qPCR.

Thank you to my family for always supporting me throughout my career. They were on my mind throughout my time at URI and they will continue to be there throughout my career. I also want to thank my fiancé for being a huge support during the last two years that we have lived in Rhode Island away from both our families during a pandemic. I could not have done this without you.
# TABLE OF CONTENTS

ABSTRACT .......................................................................................................................... ii

ACKNOWLEDGMENTS ................................................................................................. iii

TABLE OF CONTENTS ............................................................................................... iv

LIST OF TABLES ........................................................................................................... v

LIST OF FIGURES .......................................................................................................... vi

CHAPTER 1 ..................................................................................................................... 1

   INTRODUCTION .......................................................................................................... 1

CHAPTER 2 ..................................................................................................................... 17

   METHODOLOGY ........................................................................................................ 17

CHAPTER 3 ..................................................................................................................... 31

   RESULTS ................................................................................................................... 31

CHAPTER 4 ..................................................................................................................... 60

   DISCUSSION ............................................................................................................. 60

BIBLIOGRAPHY ........................................................................................................... 73
# LIST OF TABLES

| TABLE | PAGE |
|-------|------|
| Table 1. Okamoto et al., 2018 ChIP-seq top 30 hits of FANCD2 overlap. | 14 |
| Table 2. Cell Culture reagents used and catalog numbers | 24 |
| Table 3. Immunoblotting reagents used and catalog numbers | 25 |
| Table 4. qPCR reagents used and catalog numbers | 26 |
| Table 5. Antibodies, catalog numbers, and antibody working dilutions used in this study | 27 |
| Table 6. qPCR primer sequences, melting temperatures (Tm), and sequence lengths used in this study | 28 |
| Table 7. RNA sequencing heat map of no treatment samples | 58 |
| Table 8. RNA sequencing heat map of 24 h aphidicolin (APH) samples | 59 |
**LIST OF FIGURES**

| FIGURE | PAGE |
|--------|------|
| Figure 1. Identification and coordination of FA pathway in ICL repair. | 3 |
| Figure 2. The FA pathway and replication stress. | 8 |
| Figure 3. FANCD2 displays broad peak over \textit{WWOX}. | 15 |
| Figure 4. FANCD2 displays broad peak over \textit{SOX5}. | 16 |
| Figure 5. Analysis of \textit{SOX5} and \textit{WWOX} gene expression in U2OS cells incubated in the absence or presence of aphidicolin. | 32 |
| Figure 6. Analysis of \textit{GAPDH} primer efficiency using a serial dilution of cDNA. | 35 |
| Figure 7. Analysis of \textit{WWOX2} primer efficiency using a serial dilution of cDNA. | 36 |
| Figure 8. Analysis of \textit{SOX5 2} primer efficiency using a serial dilution of cDNA. | 37 |
| Figure 9. Primer gradients using different quantities of cDNA and melting temperatures. | 40 |
| Figure 10. Optimal melting temperatures determined for new primer sets. | 41 |
| Figure 11. Effect of aphidicolin on SOX5 protein expression. | 43 |
| Figure 12. Effect of aphidicolin on WWOX protein expression. | 45 |
| Figure 13. Effects on SOX5 after siRNA knockdown of FANCD2. | 47 |
| Figure 14. Effects on WWOX after siRNA knockdown of FANCD2. | 49 |
| Figure 15. Analysis of \textit{SOX5} gene expression in U2OS cells incubated in the absence or presence of aphidicolin. | 51 |
| Figure 16. Analysis of \textit{WWOX} gene expression in U2OS cells incubated in the absence or presence of aphidicolin. | 53 |
Figure 17. RNA-seq analysis reveals no significant differential expression in *WWOX* ................................................................. 57

Figure 18. U2OS 3xFLAG cells at grown for either 24 hours or 48 hours with no treatment (NT) added................................................................. 69

Figure 19. RNA sequencing reveals potential link between FANCD2 and neurological development in no treatment samples....................................................... 70

Figure 20. RNA sequencing reveals potential link between FANCD2 and neurological development in aphidicolin treated samples........................................ 71

Figure 21. RNA-seq data displaying 24-hour aphidicolin (APH) differentially expressed genes categorized by gene ontology ...................................................... 72
CHAPTER 1

INTRODUCTION

Fanconi anemia (FA) is a rare genetic disorder that can lead to bone marrow failure, leukemia, and premature cancers (Alter et al., 2018). The median age of onset of FA is seven years but patients may exhibit symptomatic and asymptomatic disease from birth to the age of fifty (Fiesco-Roa et al., 2019). Biallelic mutations in one of the twenty-three FA genes that encode for proteins can lead to FA diagnosis (Fiesco-Roa et al., 2019). FA results from biallelic mutations in any one of the following genes: FANCA, FANCB, FANCC, BRCA2 (FANCD1), FANCD2, FANCE, FANCF, FANCG, FANCI, BRIP1 (FANCJ), FANCL, FANCM, PALB2 (FANCN), RAD51C (FANCO), SLX4 (FANCP), ERCC4 (FANCQ), RAD51 (FANCR), BRCA1 (FANCS), UBE2T (FANCT), XRCC2 (FANCU), MAD2L2 (FANCV), RFWD3 (FANCW) (Milletti et al., 2020). Patients with FA have an increased chance of developing acute myelogenous leukemia and squamous cell carcinomas of the head and neck in comparison to the non-FA population (Scheckenback et al., 2012; Fiesco-Roa et al., 2019). Clinical characteristics of patients with FA include physical abnormalities such as short stature, microcephaly, and abnormal skin pigmentation as well as an overall increased chance of developing cancers (Hays et al., 2014). It is important to note that FA patients should be further subtyped as there is clinical variability amongst the subtypes.

The FA genes are autosomal recessive apart from FANCB which is X-linked, and FANCR which is autosomal dominant. They are involved in the DNA damage
repair pathway which is associated with the mending of DNA interstrand crosslinks (ICL) and stabilization of stalled replication forks (Fig. 1) (Feng et al., 2019; Schlacher et al., 2012). The elimination of DNA damage is necessary for the maintenance of genomic integrity. Inefficient repair of DNA damage can lead to genomic instability increasing the likelihood of cancer commencement and development (Niraj et al., 2019). The common diagnostic test for establishing if someone has FA is by determining if there is cellular hypersensitivity to DNA interstrand crosslinking agents (ICLs) such as mitomycin C (MMC). The exposure of FA cells to this agent results in increased levels of chromosomal aberrations, including chromosomal breaks and radial formations, as well as an increase in cell cycle arrest at the G2/M phase (Kee & D’Andrea, 2012).

The FA protein network can be further divided into groups based on commonalities in function. FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL, and FANCM function as the FA core complex and together with the E2 ubiquitin-conjugating enzyme FANCT/UBE2T, monoubiquitinate FANCD2 and FANCI. FANCL, FANCB, and FAAP100 can also form a sub-catalytic complex that is able to support low levels of monoubiquitination, while FANCA-FANCG-FAAP20 and FANCC-FANCE-FANCF provide support for the chromatin and DNA damage associated with the catalytic module (Huang et al., 2014). Mutations in any one of the eight proteins in the core complex results in loss of function in the FANCD2/FANCI monoubiquitination (Kee & D’Andrea, 2012). FANCI and FANCD2 are paralogs and when monoubiquitinated, form a heterodimer that will be referred to as ID2 (Joo et al., 2011).
Figure 1. Identification and coordination of FA pathway in ICL repair. a,b. Replication forks stall at DNA ICLs and recognized by FANCM-FAAP24-MHF1-MF2 (FAAPs) or UHRF1. Helicase removed by BRCA1 allowing one of the replication forks to converge on the ICL. c. FANCM further assists in the initiation of the ATR-kinase dependent checkpoint. d. FA core complex monoubiquitinates FANCI-FANCD2 (ID2) complex. e. FANCD2 recruit’s endonuclease to ICL site to induce DSB to further be repaired by homologous recombination (HR) (adapted from Niraj et al., 2019).
Recently, Tan et al., 2020 discovered that the monoubiquitination of ID2 promotes protein:protein interaction and helps to stabilize the ID2 heterodimer onto the double stranded DNA (dsDNA). This clamping action has also been seen to only require the monoubiquitination of FANCD2 but not the dimer (Tan et al., 2020). The monoubiquitination of ID2 promotes the assembly of foci at sites of DNA damage in chromatin to further facilitate DNA repair and protect the genome. FANCI along with the core complex are required to produce FANCD2 foci that mark locations of DNA double-stranded breaks (DSBs), stalled replication forks, and R-loops in the nucleus in order to protect nascent DNA from degradation by nucleases (Taniguchi et al., 2002; Schwab et al., 2015). Following the exposure of DNA to damaging agents, and during S-phase of the cell cycle, the core complex will monoubiquitinate FANCD2 at lysine 561 to signal the activation and translocation to nuclear foci (Garcia-Higuera et al., 2001).

Post monoubiquitination, there is a recruitment of effectors that cleave the DNA including FANCQ (XPF) and FANCP (SLX4) (Y. Kim et al., 2011; Stoepker et al., 2011). FANCM-FAAP24-MHF also form a complex which acts upstream of the FA pathway to detect DNA damage as well as initiate signal transduction pathways to promote the monoubiquitination (Nepal et al., 2017). The monoubiquitination leads the BRCA2 protein into chromatin complexes to facilitate the assembly of DNA damage inducible RAD51 nuclear foci (Taniguchi et al., 2002; Schwab et al., 2015). FANCD2 will also oversee the coordination of ICL incision by SLX4-XPF or FAN1 nuclease leading to repair by homologous recombination (H. Kim & D’Andrea, 2012). While it has been established that the monoubiquitination of FANCD2 is essential for
some DNA damage inducer resistance, how FANCD2 does this as well as responds to DNA replication stress is not fully understood.

The replication of the human genome is an intricate process that requires the organized activation and maintenance of replication forks ensuing from many sites of origins of replication during S-phase (Schwab et al., 2015). Precise DNA replication requires numerous factors, including proteins of the FA pathway. The accurate duplication of chromosomes, followed by their even segregation during mitosis is essential for genome stability (Mankouri et al., 2013). Errors that can occur during replication can impact the precise copying of chromosomes and segregation during mitosis. Replication stress is therefore, one of the main sources of genome instability. Sources of replication stress result from both endogenous as well as exogenous sources and can act either locally or globally. Sources can be categorized into alternations of origins firing, impediments to fork progression, conflicts between the DNA replication and transcription machineries, and DNA replication in unfit metabolic conditions (Magdalou et al., 2014).

Collapsed replication forks are a frequent contributor to spontaneous recombination events and genomic instability, both of which are hallmarks of cancer. DNA replication originates at thousands of individual replication origins that form bidirectional replication forks. Prior to S-phase, every origin is prepped by combinations of replication initiation proteins to ready the chromatin for replication (Barlow et al., 2013). After commencing, cells need to achieve balance, accuracy, speed, as well as the utilization and administration of necessary resources for example nucleotides and replication factors to complete replication in a timely manner.
Therefore, eukaryotic cells initiate replication origins intermittently, dividing them into early replicating and late-replicating origins (Zeman & Cimprich, 2014).

DNA replication must also work through parts of the genome that are considered as more difficult to replicate because of an increased probability of breaks and gaps. Regions of the genome known as common fragile sites (CFSs) are more prone to breaks and gaps on metaphase chromosomes likely because of an increase of stalled replication forks at these regions (Debatisse et al., 2012). These regions require the replication fork to progress through CFSs, which causes a reliance on fork-restart mechanisms (Magdalou et al., 2014). When the DNA polymerase has stalled, helicases are continuing to unwind DNA resulting in an accumulation of single stranded DNA (ssDNA) at stalled forks. ssDNA is more unstable and more prone to be acted upon by nucleases (Byun et al., 2005).

During replication stress, regions containing CFSs have paused replication forks and consequently the forks are either at a standstill or moving slower as they move from flanking regions through CFSs to complete replication. This results in the inevitable collision of replication and transcription machineries intensifying genome instability (Debatisse et al., 2012). One result of this collision is the formation of R-loops, which are DNA: RNA hybrids, that result from the nascent transcript reannealing to the corresponding template DNA strand causing the non-template strand to be displaced as ssDNA (Hamperl & Cimprich, 2016). These hybrids can lead to DNA damage because of the exposure of the ssDNA (Sollier & Cimprich, 2015). Previous studies show that the transcription of large genes can require more than one
cell cycle, consequentially resulting in replication-transcription collisions resulting in potential R-loop formations (Hamperl & Cimprich, 2016).

Replication-transcription conflicts are often inevitable as there are many large genes in the genome requiring start-restart replication as well as more than one cell cycle for transcription (Helmrich et al., 2011). Conflicts are unavoidable in the longer human genes because these genes often overlap with CFSs resulting in longer times for transcription. Genes overlapping with CFSs are also replicated late in S-phase and are hotspots for chromosome instability (Le Tallec et al., 2014). This frequent association seen between large genes and transcriptional fragility, suggests a functional relationship, however the mechanisms underlying this association are not fully understood (Le Tallec et al., 2014) Collisions of the two systems can as previously stated result in R-loop formations but can also result from cis structural features such as DNA supercoiling, CFSs, and G-quadruplexes (Fig. 2) (García-Muse & Aguilera, 2016).
Figure 2. The FA pathway and replication stress. While undergoing replication stress, the ATR checkpoint kinase is activated resulting in the monoubiquitination of the FANCD2-FANCI complex by the FA core E3 ligase complex. This causes the replication machinery to be put under pressure and collide with concurrent transcription complex causing transcripts to enter the DNA helix and create R-loops. FANCD2 is recruited to sites of R-loops however, how it is recruited is still unknown (adapted from Okamoto et al., 2019).
As DNA replication continues, it exposes ssDNA because of lagging strand synthesis. These ssDNA are more prone to fold into G-quadruplexes and impede replication fork progression. G-quadruplexes form due to the repetitive nature of the genome and can result in genomic instability (Hänsel-Hertsch et al., 2017). They are non-canonical DNA secondary structures that form because of the interactions of guanines in G-rich sequences where the nucleotides have interacted by Hoogsteen hydrogen bond that have been stabilized by a cation (Maffia et al., 2020).

A frequent hindrance of replication is caused when the replication fork encounters CFSs. These regions have been seen to undergo chromosomal rearrangements in tumors but the correlation between their role in cancer and implementation in replication has only recently been uncovered (Maffia et al., 2020). Studies have seen that these regions in the genome are sensitive to replication stress and the treatment of cells with aphidicolin, a DNA polymerase inhibitor, can cause breaks in metaphase chromosomes (Glover et al., 2017; Helmrich et al., 2011). When replication forks are confronted with obstacles during S phase, normally dormant replication origins are fired to prevent or rescue the instability. These CFS regions can be large with some being greater than 1 Mb in length as seen in a recent study identifying FANCD2 accumulating at large transcriptionally active genes (Okamoto et al., 2018). The Okamoto study found that FANCD2 preferentially bound to sites that were enriched near the center of the genic region of the large genes. Large genes can have a greater distance between dormant origins, increasing the likelihood of replication stress occurring. When replication stress is encountered at CFSs, because
of the short supply of origins, the ability to activate this rescue mechanism is lost (Shima & Pederson, 2017).

Recent studies have investigated if there is a connection between FANCD2 and CFSs. One study has found that FANCD2 has been shown to reposition to large genes that comprise CFSs after replication stress (Fernandes et al., 2021). It has also been seen that FANCD2 forms foci at CFSs during mitosis, marking these areas as damaged (Chan et al., 2009). As a result of the FA pathway involvement in the coordination of replication and transcription by preventing or mending R-loops, a connection forms between the failure to resolve R-loops, CFS replication, and the genomic instability commonly seen in FA patients. In addition, a connection can be formed as FANCD2 has been seen to promote CFS replication by the regulation of R-loop formation (Okamoto et al., 2019). It has also been previously seen that FANCD2 may help promote the replication of CFSs (Madireddy et al., 2016).

Further recent discoveries involving FANCD2 include that FANCD2 has been seen to accumulate on large actively transcribed genes when under conditions of replication stress (Okamoto et al., 2018). A genome-wide chromatin localization of FANCD2 using chromatin immunoprecipitation sequencing (ChIP-seq) analysis was used to identify the binding footprints of FANCD2. This study found that FANCD2 accumulates in the central region of large actively transcribed genes with some of the genes overlapping with known CFSs in cells that had been cultured in the presence of aphidicolin (Okamoto et al., 2018). This study also assembled a list of the top thirty genes of a total of 120, that FANCD2 was found to bind to under conditions of replication stress (Table 1). However, what this study did not find was why is
FANCD2 binding to these regions, and if by doing so it is preventing the transcription in these areas undergoing replication stress.

It has also been recently discovered that many of these large genes are prone to copy number variation (CNV) (Wilson et al., 2015). This is a type of mutation where the gene copy number can vary under certain conditions. Under replication stress, there could be a deletion of one allele and a triplication of the other allele causing the normal number of alleles to increase from two to three. At a locus that is more prone to CNV, there can be considerable variability in copy number and consequently in the levels of transcription of that gene (Wilson et al., 2015).

While there have been many studies of FANCD2 acting as part of the damage response, it has only recently come to light that this protein may have another function of acting as a transcriptional repressor for large actively transcribed genes that are undergoing replication stress. FANCD2 has been seen to bind to large transcriptionally active genes and that because of its role in DNA repair, it is binding to these regions because they are more susceptible to DNA damage. The question remains of what is FANCD2 doing when it binds to these large actively transcribed genes? Is it acting as a protector of the genome by blocking susceptible genes from being transcribed and causing DNA damage?

In this study, we used a candidate gene approach to test the hypothesis: if there are changes in gene expression at the sites that FANCD2 is found to bind to, then is FANCD2 impacting gene expression by binding to these regions? We focus on SOX5 and WWOX in U2OS 3xFLAG cells to answer this question as these two genes encode for proteins found in the top ten of the top thirty genes list and were shown to have
broad FANCD2 peaks in the ChIP-seq data (Figs. 3 & 4). Aphidicolin, as previously stated is a DNA polymerase inhibitor and frequent inducer of replication stress. Through the utilization of aphidicolin as the replication stress inducer, we used immunoblotting as well as quantitative-PCR (qPCR) to analyze both the protein and transcript levels of SOX5 and WWOX. These two genes were chosen as SOX5, is a transcription factor involved in breast cancer regulation, and WWOX, a member of short-chain dehydrogenases and is found to span the common fragile site FRA16D (Driouch et al., 2002; Pei et al., 2014). They were also chosen because both proteins are intracellular making them good targets for immunoblotting using whole cell lysates. We first determined if there was a change in the protein and transcript levels under aphidicolin conditions and hypothesized that when FANCD2 is present and under conditions of replication stress, the amount of protein and transcript levels will decrease. We then used an siRNA to knock down FANC12, aphidicolin as our replication stress inducer, and analyzed the consequences this had on both the protein and transcript levels and hypothesized that the levels will then increase in the absence of FANCD2.

We further analyzed the role FANCD2 may play under replication stress conditions by using ACHT hTERT cells both with and without FANCD2 and both with and without aphidicolin treatment for RNA sequencing. ACHT hTERT cells are a telomerase immortalized primary fibroblast patient cell line. ACHT hTERT FANCG (FANCD2 -/-) lack FANCD2 and ACHT hTERT FANCD2 have FANCD2 complemented back into the line (ACHT hTERT +FANCD2). A patient cell line was used instead of U2OS 3xFLAG cells to observe a more realistic cell line that could be
more similar to an FA patient with a mutation in FANCD2 as opposed to a transformed cell line (U2OS 3xFLAG). This experiment allowed us to view the overall transcriptome to reveal not only if there are a greater number of genes affected by the presence or absence of FANCD2 but also if there is a pattern to the type of gene affected.
Table 1. Okamoto et al., 2018 ChIP-seq top 30 hits of FANCD2 overlap. List of top 30 genes that FANCD2 was found to bind to, the chromosome location of the gene, if the location coincides with a known CFS, and the number of reads under aphidicolin (APH) conditions.

| Top hits order | Chromosome location | Gene | CFS   | Number of reads |
|----------------|---------------------|------|-------|-----------------|
| 1              | 10q23               | NRG3*| -     | 8608            |
| 2              | 16q23               | WWOX | FRA16D| 7929            |
| 3              | 16q23               | CDH13*| -     | 4704            |
| 4              | 9p21                | LINGO2*| FRA9C| 3832            |
| 5              | 7p22                | SDK1*| FRA7B | 3395            |
| 6              | 11q14               | DLG2*| FRA11F| 3302            |
| 7              | 3p14                | FHIT*| FRA3B | 3082            |
| 8              | 3q11                | EPHA6*| -     | 3074            |
| 9              | 3p26                | CNTN4*| -     | 2906            |
| 10             | 12p12               | SOX5*|       | 2241            |
| 11             | 10q11               | PRKG1 | FRA10C| 2091            |
| 12             | 21p11               | miR-3648| -     | 2002            |
| 13             | 7q31                | IMMP2L*| FRA7K| 1937            |
| 14             | 21p11               | miR-3687| -     | 1929            |
| 15             | 5p14                | CDH12*| -     | 1800            |
| 16             | 1p31                | NEGR1*| FRA1L | 1392            |
| 17             | 5p15                | CTNND2*| -     | 1072            |
| 18             | 8q24.3              | TRAPPC9*| FRA8D| 1001            |
| 19             | 20p12               | MACROD2*| -     | 891             |
| 20             | 4p15                | -     | FRA4D | 860             |
| 21             | 1p36                | -     | FRA1A | 797             |
| 22             | 11q13               | -     | FRA11H| 765             |
| 23             | 7q11                | MAGI2*| -     | 636             |
| 24             | 7q11                | AUTS2*| FRA7J | 616             |
| 25             | 13q33               | FAM155A| -     | 497             |
| 26             | 4q35                | FRG2  | -     | 401             |
| 27             | 9p32                | ASTN2*| -     | 380             |
| 28             | 22q13               | -     | FRA22A| 346             |
| 29             | Xq22                | DIAPH2| -     | 339             |
| 30             | 2q22                | THSD7B| -     | 322             |

Genes with (*) are related with autism, neurodevelopmental or psychiatric disorders. The number of ChIP-seq reads found in a gene or in a region that consist of a combined gene of CFS. SOX5 and WWOX have been highlighted.
Figure 3. FANCD2 displays broad peak over WWOX. ChIP-seq data from Okamoto et al., 2018 depicting FANCD2 accumulating in broad peaks over WWOX. U2OS 3xFLAG cells were treated with 0.4 µM aphidicolin (APH) for 24 hours. Anti-FLAG ChIP-seq reads depicting the WWOX gene region under no treatment (NT; yellow) or APH (orange) with peaks shown in blue. Image made using Integrated Genomics Viewer (IGV). Input samples are those have been crosslinked but not immunoprecipitated and are used as the control.
Figure 4. FANCD2 displays broad peak over SOX5. ChIP-seq data from Okamoto et al., 2018 depicting FANCD2 accumulating in broad peaks over SOX5. U2OS 3xFLAG cells were treated with 0.4 µM aphidicolin (APH) for 24 hours. Anti-FLAG ChIP-seq reads depicting the SOX5 gene region under no treatment (NT; yellow) or APH (orange) with peaks shown in blue. Image made using Integrated Genomics Viewer (IGV). Input samples are those have been crosslinked but not immunoprecipitated and are used as the control.
CHAPTER 2

METHODOLOGY

Cell Culture

We received U2OS 3xFLAG cells created by Okamoto et al., 2018 that stably integrated a 3xFLAG on the 3’ end of the FANCD2 coding sequence. U2OS 3xFLAG cells are an osteosarcoma line and were grown in Dulbecco modified Eagle medium (DMEM) supplemented with 10% (vol/vol) fetal bovine serum, 1% (vol/vol L-glutamine, and 1% (vol/vol) penicillin-streptomycin.

ACHT hTERT FANCD2 and ACHT hTERT FANCG (FANCD2/-) cells received from Kalb et al., 2007, were grown in Dulbecco modified Eagle medium (DMEM) supplemented with 18% fetal bovine serum (vol/vol), 1% (vol/vol L-glutamine, and 1% (vol/vol) penicillin-streptomycin. ACHT hTERT cells are human FA-D2 (FANCD2/-) telomerase-immortalized skin fibroblast cells with one cell line having FANCD2 complemented into the cell. Mutations include: Allele 1 c.2444G>A (exon 26), r.2444G>A, p(R815Q), and Allele 2 c.2715+G>A, c.2715+1G>A (IVS28+1G>A), r.2715_2716ins27 (aberrant splicing), and p.(E906Lfs *4) (Kalb et al., 2007).

Immunoblotting

For immunoblotting analysis, cells were washed in phosphate-buffered saline (PBS) and lysed in 2% w/v SDS, 50 mM Tris-HCl, 10 mM EDTA followed by sonication for 10 s at 10% amplitude. Proteins were resolved on NuPage 3-8% w/v Tris-Acetate or 4-
12% w/v Bis-Tris gels (Invitrogen) and transferred to polyvinylidene difluoride (PVDF) membranes. Antibodies and concentrations used are in Table 5.

**Replication Stress assay**

U2OS 3xFLAG cells were plated at 1 x 10^6 cells/mL and grown for 24 hours in 10 cm² dishes in a total of 8 mL DMEM media supplemented with 10% FBS, 1% (vol/vol L-glutamine, and 1% (vol/vol) penicillin-streptomycin. Cells were then treated with 0.4 µM aphidicolin (APH) for 24 or 48 hours. One set of cells were then released from APH treatment and allowed to grow for an additional 24 hours. Cells were harvested 24 hours after APH treatment, 48 hours after APH treatment, and 24 hours after release from APH treatment. ACHT hTERT FANCD2 and ACHT hTERT FANCG cells were also subjected to same experimental conditions. Cells were trypsinized using 0.05 % trypsin for 4 minutes. Media was then added to neutralize the trypsin and cells were spun down for 4 minutes to pellet in 15 mL centrifuge tubes. Media was then aspirated and cells were pelleted by centrifugation and resuspended in PBS.

**siRNA**

U2OS 3xFLAG cells were plated at 1.75 x 10^5 cells per well in a 6 well plate and the following day treated with a 20 nM FANCD2 siRNA (target sequence: AACAGGCCATGGGATACACTTTGA) or an siControl, and lipofectamine 2000 (Thermo Fisher) (Howlett et al., 2005). siControl is a random sequence that does not target a specific gene. siFANCD2 or siControl was left on for 24 hours, removed and allowed to recover for 24 hours. Cells were then incubated in the absence or presence of 0.4
μM APH for 24 hours, 48 hours, or 24 hours plus 24 hours of release to induce replication stress. Cells were harvested 24 hours or 48 hours after APH treatment. The control for this experiment was a no treatment siControl in the presence and absence of APH. The siControl is used as it is an siRNA but does not target a specific gene.

**RNA Isolation**

U2OS 3xFLAG cells were grown in 10 cm² dishes and treated using the replication stress assay as stated above. Cells were trypsinized using 0.05 % trypsin for 4 minutes. Media was then added to neutralize the trypsin and cells were spun down for 4 minutes to pellet in 15 mL centrifuge tubes. Media was then aspirated and cells were pelleted by centrifugation and resuspended in PBS. Using the PureLink RNA Mini Kit (ThermoFisher) RNA was isolated, 80 µL/sample of a ~3 U/µL DNase treatment (ThermoFisher) was applied to the column provided in the PureLink RNA Mini Kit, and RNA was quantified using Nanodrop 2000 (ThermoFisher).

**cDNA Synthesis**

RNA was converted into single-stranded cDNA using the High-Capacity Reverse Transcription Kit (ThermoFisher). 1 µg of RNA was used from each sample. cDNA quantification was done using Nanodrop 2000 (ThermoFisher). 500 ng of cDNA was used in quantitative real-time PCR (qPCR).
qPCR

Abundance of SOX5 and WWOX transcripts were determined relative to GAPDH transcript abundance by qPCR. 5 µL PowerUp SYBR Green master mix (ThermoFisher) was used along with 4µL of 1.25 µM of primers designed by Integrated DNA Technologies (IDT) were used and Origene Technologies. Samples were run on the Roche LightCycler 480 at the URI Genomics and Sequencing Center. 1 µL of a 500 ng stock of cDNA was used in the reaction. Data were analyzed via the \(2^{-\Delta\Delta CT}\) calculation method. Each reaction was run with initial denaturation at 95 °C for 5 minutes, followed by 45 cycles of denaturation at 95 °C for 15 seconds, annealing at 66 °C for 15 seconds, and an extension of 1 minutes at 72 °C. After completion of the last cycle, a melt curve analysis was performed according to the machine’s preset Step and Hold protocol (95 °C for 5 seconds; 65 °C for 1 minute; temperature ramp increment of 0.3 °C up to 95 °C).

Primer Efficiency

Primer efficiency qPCR was performed using 1500 ng, 150 ng, 15 ng, 1.5 ng, and 0.15 ng cDNA for GAPDH, WWOX, SOX5, and FANCD2. The same protocol as used for qPCR was used for the efficiency test with 1 µL of cDNA, 5 µL of PowerUp SYBR Green master mix, and 4 µL of the 1.25 µM stock primer. Primer efficiency was run for 2 different sets of WWOX primers (WWOX 1 and WWOX 2), 2 different sets of SOX5 primers (SOX5 1 and SOX5 2), and 3 different sets of FANCD2 primers (FANCD2 1, FANCD2 2, and FANCD2 3).
Primer Gradient

Primer gradient PCR was performed using temperatures of 54.2, 55.2, 57.8, 59.5, 62.7, and 64.1 °C as well as 59.4, 62.4, 64.3, 66.2, 68.1, and 71 °C. Quantities of cDNA tested include 50 ng, 100 ng, 200 ng, 500 ng, and 1 µg. Initial PCR reaction was the same as used for qPCR with 1 µL of the stock cDNA, 5 µL of PowerUp SYBR Green master mix, and 4 µL of the 1.25 µM stock primer. The PCR gradient was carried out on an Eppendorf vapo.protect Thermocycler with an initial denaturation of 94 °C for 2 minutes, followed by 40 cycles of denaturation for 5 seconds at 95 °C, annealing at gradient temperatures (59.4, 62.4, 64.3, 66.2, 68.1, 71.0 °C), and an extension at 72 °C for 10 seconds.

PCR products were tested on a 2% agarose gel using 1x TBST and 10 µL SYBR SAFE DNA gel stain. 2 µL of 6x gel loading dye was added to the 10 µL reaction and the total 12 µL was loaded onto the gel. Agarose gel was run at 100 V for 50 minutes on a PowerEase 90W by Life Technologies. Gel was imaged using the Bio-Rad imager in the Howlett Lab.

RNA-seq Analysis

ACHT hTERT FANCD2 and ACHT hTERT FANCG (FANCD2-/-) cells were seeded, and the following day treated with 0.4 µM APH for 24 hours, 48 hours, and a no treatment sample in quadruplet. Cells were trypsinized using 0.25 % trypsin for 8 minutes. Media was then added to neutralize the trypsin and cells were spun down for 4 minutes to pellet in 15 mL centrifuge tubes. An ice-cold PBS wash was then applied, and cells were transferred to a 1.5 mL microcentrifuge tube where they were again
spun down in a microcentrifuge set to 4°C for 4 minutes. PBS was aspirated and cells were flash frozen using liquid nitrogen.

24 samples were sent to Genewiz in South Plainfield, NJ, USA. RNA library preparation with polyA selection and HiSeq Sequencing RNA library preparations was also completed at Genewiz. RNA samples were quantified using Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and RNA integrity was checked using Agilent TapeStation 4200 (Agilent Technologies, Palo Alto, CA, USA). RNA sequencing libraries were then prepared using NEBNext Ultra II RNA Library Prep kit for Illumina using the manufacturer’s instructions (NEB, Ipswich, MA, USA). mRNAs were then initially enriched with Oligod(T) beads, and the enriched mRNAs were fragmented for 15 minutes at 94 °C. First and second strand cDNA was then subsequently synthesized, and the cDNA fragments were end repaired and adenylated at the 3’ ends. Universal adapters were ligated to cDNA fragments and followed with an index addition and library enrichment by PCR with limited cycles. The sequencing library was then validated on the Agilent TapeStation (Agilent Technologies, Palo Alto, CA, USA), and quantified by using Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA) as well as by quantitative PCR (KAPA Biosystems, Wilmington, MA, USA). The sequencing libraries were then clustered on a single lane of a flowcell and following clustering, the flowcell was loaded onto the Illumina HiSeq 4000 instrument according to manufacturer’s instructions. Samples were sequenced using a 2x150bp Paired End configuration at 60 million reads per sample. Image analysis and base calling were conducted using the HiSeq Control Software (HCS). The raw sequence data (.bcl files) were generated from the Illumina HiSeq and converted into fastq files. They were then
de-multiplexed using Illumina’s bcl2fastq 2.17 software. One mismatch was allowed for index sequence identification.

**Bioinformatics Workflow**

To analyze the quality of the raw data, the sequence reads were trimmed to remove possible adapter sequences as well as nucleotides with poor quality using Trimmomatic v.0.36. The quality threshold parameters under Trimmomatic were set at 15 for the sliding window size of 4. Data quality \( \geq 80\% \) of bases \( \geq Q30 \) were kept. The trimmed reads were then mapped to the most recent Homo sapiens reference genome that is available on ENSEMBL using the STAR aligner v.2.5.2b. STAR aligner is a splice aligner that can detect splice junctions and will incorporate them to help align the entire read sequences. BAM files were then generated, and unique gene hit counts were calculated by using FeatureCounts from the Subread package v.1.5.2. Only unique reads that fell within exon regions were counted. After the extraction of gene hit counts, the gene hit counts table was used for downstream differential expression analysis. DESeq2 was then used to compare gene expression between the groups of samples. The Wald test was then used to generate p-values and Log2foldchanges. Genes that had adjusted p-values were called as differentially expressed genes for each of the comparisons and a gene ontology analysis was performed on the statistically significant set of genes using the software GeneSCF. The goa_human GO list was then used to cluster the set of genes.
Statistical Analysis

To determine significance, either a single-factor ANOVA or two-factor ANOVA was used to test and post-hoc analysis was performed if significance was found. Student’s two-tailed paired t-test was also performed. P-value was determined using the Bonferroni correction of the p-value for each test (0.05) divided by the number of tests performed.
Table 2. Cell culture reagents used and catalog numbers. The following table includes a list of cell culture reagents and the company that they were purchased from.

| Product Name       | Source       | Catalog Number |
|--------------------|--------------|----------------|
| DMEM 1x            | ThermoFisher | 11965118       |
| L-glutamine        | ThermoFisher | 25030081       |
| Penicillin-streptomycin | ThermoFisher | 15070063       |
| PBS                | ThermoFisher | 10010049       |
| 0.05% Trypsin      | ThermoFisher | 25300054       |
| 0.25% Trypsin      | ThermoFisher | 25200056       |
| FBS                | ThermoFisher | 10437028       |
| Lipofectamine 2000 | ThermoFisher | 11668-027      |
Table 3. Immunoblotting reagents used and catalog numbers. The following table includes the reagents used for immunoblotting and the company they were purchased from.

| Product Name            | Source    | Catalog Number |
|-------------------------|-----------|----------------|
| 3-8% Tris-Acetate Gels  | ThermoFisher | EA0378BOX      |
| 4-12% Bis-Tris Gels    | ThermoFisher | NP0335BOX      |
| NuPage Antioxidant      | ThermoFisher | 17504044       |
| Clarity                 | BioRad    | 1705060        |
Table 4. qPCR reagents used and catalog numbers. The following table includes a list of qPCR reagents and the company that they were purchased from.

| Product Name                  | Source     | Catalog Number |
|-------------------------------|------------|----------------|
| PureLink RNA Mini Kit         | ThermoFisher | 12183018A     |
| SYBR Green                    | ThermoFisher | 4344463      |
| PureLink DNase                | ThermoFisher | 12185-010    |
| cDNA                          | ThermoFisher | 4368814      |
| Gel Loading Purple Dye        | NEB        | B7025S        |
| Quick-load Purple 100 bp DNA Ladder | NEB     | N0551S      |
| SYBR Safe DNA gel stain       | ThermoFisher | S33102      |
| Agarose                       | ThermoFisher | BP160-100   |
| 10x TAE                       | Invitrogen  | 15558-026    |
Table 5. Antibodies, catalog numbers, and antibody working dilutions used in this study. The following table includes the antibodies used for immunoblotting as well as the company they were purchased from, and the dilution used.

| Product Name          | Source          | Catalog Number | Dilution used for ACHT (µL/mL) | Dilution used for U2OS 3xFLAG (µL/mL) |
|-----------------------|-----------------|----------------|-------------------------------|----------------------------------------|
| Rabbit Secondary      | Fisher Scientific | NA9341         | 1:5                           | 1:5                                    |
| Mouse Secondary       | Fisher Scientific | NA931          | 1:5                           | 1:5                                    |
| Anti- SOX5 Primary    | Abcam           | ab94396        | 1:3                           | 1:5                                    |
| Anti- WWOX Primary    | Abcam           | 137726         | 1:6                           | 1:8                                    |
| Anti- Tubulin         | Fisher Scientific | MS581P0        | 1:15                          | 1:25                                   |
| Anti-FANCD2 Primary   | Novus Biologicals | FI17           | 1:50                          | 1:100                                  |
Table 6. qPCR primer sequences, melting temperatures (Tm), and sequence lengths used in this study. The following table includes a list of the primer sequences used for qPCR as well as the melting temperatures, and sequence lengths.

| Gene       | Forward Primer | Reverse Primer | Tm | Sequence Length |
|------------|----------------|----------------|----|-----------------|
| OriGene    |                |                |    |                 |
| WWOX       | TCCTCCAGGATG   | AAGCCAGCATCGC  | 69, 67 | 201             |
|            | TTTTGTCGCCG    | CCAATAGTC      |    |                 |
| OriGene    |                |                |    |                 |
| SOX5       | CTCGGCAAATGA   | ACTGCCAGTTGCT  | 66, 62 | 142             |
|            | AGGAGCAACTC    | GAGTCAGAC      |    |                 |
| GAPDH      | GTCTCCTCTGAC   | CCAATTCGTTGT   | 64, 64 | 113             |
|            | TTCAACAGCGAC   | CATACCAGGAA    |    |                 |
| FANCD2 040521 | GATGGTGAAGAAAG | GGTCTAATCAGAG  | 49, 44 | 75              |
|            | GACGA          | TCATC          |    |                 |
| SOX5 1     | ATTGGGATCTCG   | GCCCTGGGGCTTGT | 61, 58 | 120             |
|            | CTGGAAAG       | ACTTATAG       |    |                 |
| SOX5 2     | GCCCTAACTGTT   | GTCAGCTGATGTC  | 59, 58 | 103             |
|            | TCATCCTTAC     | CCAACTATT      |    |                 |
| WWOX 1     | ACGAGTGCGAGG   | GGTCCAAGTAGGT  | 59, 57 | 104             |
|            | AGATTTG        | GGTCTTTT       |    |                 |
| WWOX 2     | CCTCAGAGCTCC   | GCAGAGCTTGGAC  | 60, 57 | N/A             |
|            | ATCGATTTAC     | CTGTTAT        |    |                 |
| FANCD2 1   | CATCTTCCTCCA   | CGTCTTCTTAGTT  | 60, 58 | 83              |
|                | CATTCCTAC | CAGCCATC |                   |     |
|----------------|-----------|----------|-------------------|-----|
| **FANCD2 3**   | CTGAGGCCATA | GTAGGGAATGTGG  | 60, 60         | 86  |
|                | GAGGAGATTG  | AGGAAGATG  |                   |     |
| **FANCD2      | CGCCAGTTGGTG | CCATGGCTTTAC  | 62, 62         | 91  |
| N-Terminus     | ATGGATAA   | GGAATGA   |                   |     |
CHAPTER 3

RESULTS

Analysis of SOX5 and WWOX gene expression following aphidicolin treatment

A previous study by Okamoto et al., 2018 utilized ChIP-seq and determined that under conditions of DNA replication stress, FANCD2 was found to bind and traverse large actively transcribed genes. However, it was not determined why FANCD2 was binding to these genes. To explore if FANCD2 could potentially be acting as a specific transcriptional repressor of these genes, we used a candidate gene approach to determine if there were changes in the expression of SOX5 and WWOX genes under conditions of replication stress. A candidate gene approach allowed us to choose two genes from the top 30 genes that FANCD2 was found to bind and do an in-depth analysis of WWOX and SOX5.

Cells were exposed to 0.4 µM APH and samples were run in triplicate and three biological samples were tested. RNA was purified from each sample and first-strand cDNA was synthesized immediately after RNA isolation. Three separate qPCR reactions were assembled using the synthesized cDNA, representing three technical replicates. For each reaction GAPDH was used as the control housekeeping gene. Cycle threshold (Ct) values were averaged as well as the standard errors for the analysis. The Ct value is the cycle at which the fluorescence from the amplification exceeds the background fluorescence with each cycle ideally doubling the target DNA. The lower the Ct value, the higher the initial copy numbers of the target.
Ranges of Ct values acceptable for analysis are less than 35 cycles while the recommended amounts of cDNA added to the reaction range from 5-50 ng of cDNA depending on the gene of interest as suggested by Thermo Scientific. Once normalized to GAPDH, the delta Ct values were also averaged for each timepoint. There was a decrease in gene expression for SOX5 between the no treatment sample and 24 hours of APH and 48 hours of APH (Fig. 5a). Gene expression then increased in the 24-hour release sample. While there were changes seen in gene expression, the changes were not significant. Standard error bars show the amount of variability within the samples between all replicates (Fig. 5a).

For WWOX, fold change analysis of the gene expression showed that there is a minor decrease in gene expression in the 24-hour and 48-hour samples in comparison to the no treatment, and an increase in the release sample but the difference is not significant (Fig. 5b). Bonferroni correction was applied by dividing 0.05 by the number of replicates (27).
Figure 5. Analysis of SOX5 and WWOX gene expression in U2OS cells incubated in the absence or presence of aphidicolin (APH).

U2OS cells were incubated in the absence (No Treatment) or presence of 0.4 μM APH for 24 hours, 48 hours, or 24 hours plus 24 hours of release. Error bars represent the standard errors of the mean calculated from twenty-seven technical replicate measurements from three independent biological replicate experiments. Statistical significance was determined using student’s two-tailed paired t-test. *P < 0.001 and Single-Factor ANOVA. a. SOX5 gene expression. Bonferroni correction has been applied. b. WWOX gene expression.
**SOX5, WWOX, and GAPDH qPCR primer efficiency analysis**

Our results suggested little change in the levels of expression of *SOX5* and *WWOX* under the conditions tested. Our Ct values for *SOX5* and *WWOX* qPCR analysis also suggested that our qPCR conditions might not be optimal. Upon consulting with our colleagues and technical support at Thermo-Scientific, we decided to thoroughly analyze the efficiency of our qPCR primers using varying amounts of quantified cDNA input. Each gene had two sets of primers designed that were tested along with the housekeeping gene, *GAPDH*. To test the efficiency of the primers, a standard curve of cDNA was created using quantities of 1500 ng, 150 ng, 15 ng, 1.5 ng, and 0.15 ng of cDNA.

Primers were run using the same protocol as our normal qPCR and standard curves were created for each primer set with each cDNA quantity run in duplicate. Analysis included observing the Ct values of the different quantities of cDNA and how similar amplification curves were between duplicates.

Figure 6 shows the amplification curves for *GAPDH*, *SOX5*, and *WWOX*. The amplification curves show that the duplicate samples are very consistent and as the quantity of cDNA decreases, the Ct value at which the amplification curve starts increases (Fig. 6, 7, 8). The *GAPDH* standard curve shows that duplicate samples are very close together and for all five quantities of cDNA, there was amplification. The Ct values for *GAPDH* range from 15 at 1500 ng of cDNA to 28 at 0.15 ng of cDNA (Fig. 6). The *WWOX* curve again had amplification for all five quantities of cDNA however, the number of cycles for amplification increased in comparison to the *GAPDH* graph. The Ct values for *WWOX* range from 22 at 1500 ng of cDNA to 35 at
0.15 ng of cDNA (Fig. 7). For SOX5 there was no amplification for the 0.15 ng cDNA and the 15-ng amplified late at the 35th cycle the cycle limit for analysis. The Ct values for SOX5 range from 25 at 1500 ng of cDNA to 35 at 1.5 ng of cDNA (Fig. 8).
Figure 6. Analysis of GAPDH primer efficiency using a serial dilution of cDNA. qPCR was performed using either 1500 ng (blue lines), 150 ng (red lines), 15 ng (green lines), 1.5 ng (purple lines), or 0.15 ng (grey lines) with samples being run in duplicate. 1 µL of cDNA was added to 5 µL of SYBR Green master mix, and 4 µL of a 1.25 µM stock primer. Amplification curves display how similar the duplicate samples are by how close together the same color curves are. As the amount of input cDNA decreases, the number of cycles increases.
Figure 7. Analysis of *WWOX2* primer efficiency using a serial dilution of cDNA. qPCR was performed using either 1500 ng (blue lines), 150 ng (red lines), 15 ng (green lines), 1.5 ng (purple lines), or 0.15 ng (grey lines) with samples being run in duplicate. 1 µL of cDNA was added to 5 µL of SYBR Green master mix, and 4 µL of a 1.25 µM stock primer. Amplification curves display how similar the duplicate samples are by how close together the same color curves are. As the amount of input cDNA decreases, the number of cycles increases.
Figure 8. Analysis of SOX52 primer efficiency using a serial dilution of cDNA. qPCR was performed using either 1500 ng (blue lines), 150 ng (red lines), 15 ng (green lines), 1.5 ng (purple lines), or 0.15 ng (grey lines) with samples being run in duplicate. 1 µL of cDNA was added to 5 µL of SYBR Green master mix, and 4 µL of a 1.25 µM stock primer. Amplification curves display how similar the duplicate samples are by how close together the same color curves are. As the amount of input cDNA decreases, the number of cycles increases.
Optimization of primer annealing temperatures using gradient PCR analysis

To continue our strategy of qPCR analysis optimization, we next performed gradient PCR analysis to determine the optimal annealing temperatures for our qPCR primers. Gradient PCR analysis also allowed us to visualize our PCR products to determine if our primers were amplifying a band of the correct size and determine if any non-specific products were being generated. A primer gradient was performed utilizing different concentrations of cDNA, as well as a gradient of temperatures (54.2, 55.2, 57.8, 59.5, 62.7, 64.1°C). The PCR products were run on a 2% agarose gel run for 50 minutes at 100 V. Bands did not appear for SOX5 or WWOX for 50 ng, 100 ng, and 200 ng. Faint bands were seen at 500 ng but not at 1 µg of cDNA at the tested temperatures signifying that these primers are not optimally amplifying their sequence targets or indicating that these genes were being expressed at very low levels (Fig. 9). The GAPDH primers, our positive control and housekeeping gene, showed amplification at each of the cDNA quantities as well as at all of the tested temperatures. We also saw amplification for the FANCD2 primers at 100 ng, 200 ng, 500 ng, and 1µg of cDNA at temperatures of 57.8, 59.5, 62.7, and 64.1 °C.

After determining that the WWOX and SOX5 primers tested were not optimally performing, new primer sets for both genes were ordered and tested. The new primer sets were tested at 200 ng and 500 ng of cDNA and temperatures of 59.4, 62.4, 64.3, 66.2, 68.1, and 71 °C. Bands were faintly seen at 500 ng for both SOX5 and WWOX at temperatures of 59.4, 62.4, 64.3, 66.2, 68.1°C (Fig. 10). GAPDH was again used as the positive control with bands at all temperatures except for 59.4 °C for 500 ng cDNA. The optimal temperature for all primer sets was determined to be 66 °C with
500 ng of cDNA as each primer set appeared to have the brightest band at this concentration and temperature (Fig. 10).
Figure 9. Primer gradients using different quantities of cDNA and melting temperatures. Different concentrations of cDNA (50 ng, 100 ng, 200 ng, 500 ng, 1µg), as well as increasing melting temperatures (54.2, 55.2, 57.8, 59.5, 62.7, 64.1 °C), were run using a gradient PCR to determine the optimal melting temperature for each primer set. Products were run on a 2% agarose gel for 50 minutes with a 100 bp ladder and imaged using BioRad imager. Optimal temperature and cDNA concentration determined by the brightness of the band.
**Figure 10. Optimal melting temperature determined for new primer sets.** New primer sets for *SOX5* and *WWOX* run on primer gradients using different quantities of cDNA and melting temperatures. Different concentrations of cDNA (200 ng, 500 ng), as well as increasing melting temperatures (59.4, 62.4, 64.3, 66.2, 68.1, 71.0 °C), were run using a gradient PCR to determine the optimal melting temperature for each primer set. Products were run on a 2% agarose gel for 50 minutes with a 100 bp ladder and imaged using BioRad imager. Optimal temperature and cDNA concentration determined by the brightness of the band.
Analysis of the protein expression of SOX5 following aphidicolin exposure

To study if APH inducing replication stress affected the protein expression of SOX5, we analyzed via immunoblotting whole cell lysates treated with APH for 24, 48, and a 24-hour release samples (Fig. 11a). It is known that APH induces the monoubiquitination of FANCD2 and FANCD2 was confirmed to be monoubiquitinated in this experiment as can be seen in Figure 9b showing the top band at the 24-, 48-, and 24-hour release samples. Monoubiquitination is confirmed by the top band in the FANCD2 western blot. L/S ratios which are the ratio between monoubiquitinated FANCD2 and unubiquitinated FANCD2 were calculated. There is an increase in the monoubiquitination after 48 hours of APH compared to the 24-hour APH sample and the release samples shows that monoubiquitination has decreased (Fig. 11b). We saw an increase in the protein expression in SOX5 following APH treatment in the immunoblots as can be seen by comparing the 24-, 48-, and 24-hour release samples to the no treatment sample (Fig. 11b). Tubulin was used a loading control. Protein bands were further analyzed using ImageJ to measure protein band intensities and band intensity was averaged across replicates (Fig. 11c). It can be seen in the fold change of protein expression graph that there was an increase in the band intensity further concluding that there was an increase in protein expression of SOX5 following the APH treatment (Fig. 11c). Standard deviations can also be seen on the fold change expression graph displaying replicates with similar expression levels. Bonferroni correction was applied by dividing 0.05 by the number of replicates tested (3).
**Figure 11. Effect of aphidicolin on SOX5 protein expression.** a. Cartoon of experimental design. U2OS 3xFLAG cells were incubated in the absence (No Treatment; NT) or presence of 0.4 µM APH for 24 hours, 48 hours, or 24 hours plus 24 hours of release. b. The abundance of SOX5 and FANC-D2 proteins present in whole-cell lysates was determined by immunoblot using anti-SOX5 and anti-FANC-D2 antibodies. Tubulin, detected with an anti-Tubulin antibody, serves as a loading control. c. Protein band intensities analyzed using ImageJ and statistical significance was determined using student’s two-tailed paired t-test. *P < 0.01 and Single Factor ANOVA. Bonferroni correction has been applied.
Analysis of protein expression of WWOX following aphidicolin exposure

We also analyzed the protein expression of WWOX following APH treatment and determined that there was no difference between the no treatment condition and the aphidicolin treated (Fig. 12b). The band intensity of the 24-, 48-, and 24-hour APH release samples look the same as the no treatment sample (Fig. 12b). Further analysis of the band intensity was again done by use of ImageJ quantification. It can be seen that there is minor increase in the protein expression in the treated samples (Fig. 12c). The standard errors are also shown on the fold change expression graph and display replicates with similar expression levels. Bonferroni correction was applied by dividing 0.05 by the number of replicates (3).
Figure 12. Effect of aphidicolin on WWOX protein expression. a. Cartoon of experimental design. U2OS 3xFLAG cells were plated at 1 million cells/mL and incubated in the absence (No Treatment; NT) or presence of 0.4 µM APH for 24 hours, 48 hours, or 24 hours plus 24 hours of release. b. The abundance of WWOX and FANCD2 proteins present in whole-cell lysates was determined by immunoblot using anti-WWOX and anti-FANCD2 antibodies. Tubulin, detected with an anti-Tubulin antibody, serves as a loading control. c. Protein band intensities analyzed using ImageJ. *P < 0.01 and Single Factor ANOVA. Bonferroni correction has been applied.
**Analysis of the impact/effect of loss of FANCD2 on SOX5 protein expression**

We then wanted to determine what the consequences of the loss of FANCD2 would have on SOX5 protein expression following aphidicolin treatment. We treated U2OS 3xFLAG cells with an siFANCD2 and an siControl for 24 hours followed by 0.4 µM APH treatment (Fig. 13a). In comparison to the siControl samples, there are no bands in the lanes treated with the siFANCD2 (Fig. 13a). Results of the siControl bands reveal that there is an increase in protein band intensity in the 48-hour and release samples in comparison to the no treatment sample (Fig. 13b). In samples treated with an siFANCD2, there is a minor decrease in protein levels in the 24-hour and release sample lanes in comparison to the siFANCD2 no treatment lane (Fig. 13b). There does not appear to be a change in expression levels when comparing the siFANCD2 48-hour sample to the siFANCD2 no treatment sample (Fig. 13b). Changes in expression levels were confirmed using ImageJ again to quantify the intensity of the protein bands showing the decrease in siFANCD2 samples and increase in siControl samples (Fig. 13c). The standard deviations can also be seen on the fold expression protein level graph and show that replicates had similar expression levels for each of the samples tested. Bonferroni correction was applied by dividing 0.05 by the number of replicates tested (3).
Figure 13. Effects on SOX5 after siRNA knockdown of FANCD2. a. Cartoon of siRNA experimental design. U2OS 3xFLAG cells were plated at 175,000 cells/mL and treated with either an siControl or siFANCD2 for 24 hours. On Day 4 U2OS were incubated in the absence (No Treatment; NT) or presence of 0.4 µM APH for 24 hours, 48 hours, or 24 hours plus 24 hours of release. b. The abundance of SOX5 and FANCD2 proteins present in whole-cell lysates was determined by immunoblot using anti-SOX5 and anti-FANCD2 antibodies. Tubulin, detected with an anti-Tubulin antibody, serves as a loading control. c. Protein band intensities analyzed using ImageJ with siControl shown in blue and siFANCD2 shown in orange. Protein band intensities analyzed using ImageJ. *P < 0.01 and Two-Factor ANOVA. Bonferroni correction has been applied.
Analysis of the impact/effect of loss of FANCD2 on WWOX protein expression

The same siRNA treatment was also performed to test WWOX using the same siFANCD2 and siControl and using whole cell extracts looking at protein expression via immunoblotting (Fig. 14a). In contrast to SOX5, western blots displaying WWOX do not show an increase or decrease in protein level expression in the siControl lanes or the siFANCD2 lanes except for the release sample in the siFANCD2 western lanes (Fig. 14b). There does seem to be an increase in expression level when comparing the siFANCD2 release sample to the siFANCD2 no treatment sample.

ImageJ analysis of band intensities display a slight reduction in the expression of WWOX protein in the absence of FANCD2 as well as in the siControl samples, but it is very minor (Fig. 14c). For both the siControl and siFANCD2 samples, they increase initially from the no treatment to the 24-hour APH treatment and then decrease back down to the same level as the no treatment sample. Standard deviations also show that replicates were similar (Fig. 14b). Bonferroni correction was applied by dividing 0.05 by the number of replicates tested (3).
Figure 14. Effects on WWOX after siRNA knockdown of FANCD2. a. Cartoon of siRNA experimental design. U2OS 3xFLAG cells were plated at 175,000 cells/mL and treated with either an siControl or siFANCD2 for 24 hours. On Day 4 U2OS were incubated in the absence (No Treatment; NT) or presence of 0.4 µM APH for 24 hours, 48 hours, or 24 hours plus 24 hours of release. b. The abundance of SOX5 and FANCD2 proteins present in whole-cell lysates was determined by immunoblot using anti-SOX5 and anti-FANCD2 antibodies. Tubulin, detected with an anti-Tubulin antibody, serves as a loading control. c. Protein band intensities analyzed using ImageJ with siControl shown in blue and siFANCD2 shown in orange. *P < 0.01 and Two-Factor ANOVA. Bonferroni correction has been applied.
Analysis of the impact/effect of loss of FANCD2 on SOX5 gene expression

Following the analysis of protein expression of SOX5 following an siFANCD2 treatment and APH treatment, we wanted to observe if there were transcript changes to SOX5. Samples were run in triplicate and three biological samples were tested and averaged with GAPDH used as the control housekeeping gene. In the absence of FANCD2, there is an increase initially in the no treatment sample compared to the siControl no treatment sample. After 24 hours of APH treatment, there is an increase in the transcript levels for both the siControl and siFANCD2 samples. In the 48-hour samples, there is a decrease in the siControl in comparison to the 24-hour sample and is at the same level as the no treatment sample. siFANCD2 at 48 hours also decreases compared to the 24-hour and again at the same level as the no treatment sample (Fig. 15). Both siControl and siFANCD2 displayed an increase in transcript levels in the 24-hour release samples compared to the no treatment (Fig. 15). Significant differences were seen between the siControl no treatment and siFANCD2 no treatment (Fig. 15). Bonferroni correction was applied by dividing 0.05 by the number of replicates tested (12).
Figure 15. Analysis of SOX5 gene expression in U2OS cells incubated in the absence or presence of aphidicolin (APH). U2OS cells were incubated in the absence (No Treatment) or presence of 0.4 µM APH for 24 hours, 48 hours, or 24 hours plus 24 hours of release. Cells were also treated with either an siControl (blue bars) or an siFANCD2 (orange bars). Error bars represent the standard errors of the mean calculated from twenty-seven technical replicate measurements from three independent biological replicate experiments. Statistical significance was determined using student’s two-tailed paired t-test. *, $P < 0.004$. Bonferroni correction has been applied.
Analysis of the impact/effect of loss of FANCD2 on WWOX gene expression

Again, following the analysis of protein expression of WWOX in the presence of an siFANCD2 treatment and APH treatment, we wanted to observe if there were transcript changes to WWOX. Samples were again run in triplicate and three biological samples were tested and averaged with GAPDH used as the control housekeeping gene. In the absence of FANCD2, there is an increase initially in the no treatment sample compared to the siControl no treatment sample. Post 24 hours of APH treatment, there is an increase in the transcript levels for the siControl sample, however the siFANCD2 sample has stayed level with the no treatment sample. In the 48-hour samples, there are decreases in the both the siControl and siFANCD2 samples compared to the 24-hour and no treatment samples (Fig. 16). Both siControl and siFANCD2 displayed an increase in transcript levels in the 24-hour release samples compared to the untreated samples (Fig. 16). A significant difference was seen between the siControl no treatment and the siFANCD2 no treatment. Bonferroni correction was applied by dividing 0.05 by the number of replicates tested replicates (12).
Figure 16. Analysis of *WWOX* gene expression in U2OS cells incubated in the absence or presence of aphidicolin (APH). U2OS cells were incubated in the absence (No Treatment) or presence of 0.4 µM APH for 24 hours, 48 hours, or 24 hours plus 24 hours of release. Cells were also treated with either an siControl (blue bars) or an siFANCD2 (orange bars). Error bars represent the standard errors of the mean calculated from twenty-seven technical replicate measurements from three independent biological replicate experiments. Statistical significance was determined using student’s two-tailed paired t-test. *, $P < 0.004$. Bonferroni correction has been applied.
RNA sequencing analysis of ACHT hTERT FANCG (*FANCD2*^-/-_) and ACHT hTERT FANCD2 reveal no levels of *SOX5* and low levels of *WWOX*

As previously stated, FANCD2 was found to bind to and traverse large actively transcribed genes in U2OS 3xFLAG cells under conditions of replication stress. However, we wanted to see the effects of treatment with aphidicolin on telomerase immortalized cells lacking FANCD2 as well as telomerase immortalized cells with FANCD2 complemented back into the cell. A different cell line was chosen as ACHT hTERT cells are human FA-D2 (*FANCD2*-/-) telomerase-immortalized skin fibroblast cells with one cell line having FANCD2 complemented into the cell. By utilizing a patient cell line, we are able to test and observe the effects that may be more realistically occurring within the cells of FA patients.

ACHT FANCG (*FANCD2*-/-) cells were treated with 0.4 µM aphidicolin for 24 and 48 hours, harvested and sent to Genewiz for RNA sequencing analysis at 150 bp paired end sequencing at 60 million reads per sample. The analysis looked at read counts which are the number of times a read maps to a specific gene. *SOX5* was not found to be expressed in the dataset and *WWOX* had similar read counts between cells with *FANCD2* and cells without *FANCD2* but the difference in read counts was not significant (Fig. 17).

Independent of *SOX5* and *WWOX*, the RNA-seq dataset allowed us to analyze the levels of expression of all FANCD2 targets identified by ChIP-seq. In the ChIP-seq dataset generated by Okamoto et al., FANCD2 was found to bind to approximately 120 genes under conditions of replication stress. Therefore, we asked the question of if there was a difference in the levels of expression of any of these genes in the absence
or presence of FANCD2 and in the absence or presence of replication stress. Using the differential analysis datasets for the no treatment samples, provided from Genewiz, it was found that 79 of the 120 ChIP-seq genes were expressed in the RNA-seq data. In the 24-hour APH samples 66 of the 120 ChIP-seq genes were expressed. Of the 79 genes found to overlap, 26 were found to be differentially expressed with 7 being downregulated and 19 being upregulated (Table 7). In this data, upregulated refers to an increase in ACHT hTERT FANCD2’s and downregulated refers to an increase in ACHT hTERT FANCG’s. In the 24-hour APH treated samples, 22 of the 66 overlapping genes were found to be differentially expressed with 7 being downregulated and 15 being upregulated (Table 8). Interestingly, 7 of the 26 genes found in the no treatment samples and 6 of the 22 genes found in the 24-hour APH treated samples are found to be related to autism, neurodevelopmental, psychiatric disorders.

The RNA-seq data also allowed us to observe if there were differences in gene ontology. Genewiz also provided a gene ontology analysis which categorized the gene ontology and enrichment of gene ontology. It was found that the 5th most significantly enriched gene sets in the 24-hour APH treated samples were nervous system development (Fig. 21).
Figure 17. RNA-seq analysis reveals no significant differential expression in **WWOX**. Differential expression analysis displays the number of reads found per gene and if there is a significant difference between the ACHT hTERT +FANCD2 cells and the ACHT hTERT +FANCG (*FANCD2* -/-) cells. Cells were incubated in the absence (no treatment) or presence of 0.4 µM APH. Normalized reads were averaged between the 4 replicates for both cell lines and the standard deviation was also calculated. In both conditions, *SOX5* did not exhibit differential expression.
| Gene Name | log2FoldChange | P-value |
|-----------|----------------|---------|
| TENM4*   | -2.36          | 2.34E-15|
| FAM155A  | -0.87          | 7.24E-07|
| NLGN1*   | -0.76          | 2.71E-10|
| PPP1CC   | -0.75          | 5.68E-04|
| SRSF3    | -0.65          | 4.56E-03|
| ASTN2*   | -0.61          | 3.84E-04|
| FARS2    | -0.60          | 1.73E-03|
| FRMD8    | 0.55           | 2.98E-05|
| GPC6     | 0.56           | 2.74E-04|
| PTPRM    | 0.63           | 1.86E-02|
| NEGR1*   | 0.69           | 2.98E-05|
| KIRREL3* | 0.72           | 2.00E-03|
| BTBD9    | 0.74           | 3.80E-06|
| FHIT     | 1.01           | 3.50E-02|
| SMYD3    | 1.17           | 4.37E-08|
| ANKS1B   | 1.41           | 2.50E-02|
| SPOCK1   | 1.43           | 8.51E-25|
| NTM      | 1.71           | 7.75E-11|
| FGF12    | 1.71           | 1.65E-03|
| LMCD1    | 1.79           | 2.46E-22|
| CDH4     | 2.42           | 3.62E-06|
| PLCB1*   | 2.53           | 4.38E-15|
| NCAM2*   | 2.94           | 5.40E-46|
| PAPPA2   | 4.52           | 4.69E-27|
| CTNND2   | 7.59           | 4.07E-12|
| PTPRD    | 9.44           | 2.24E-19|

Table 7. RNA sequencing heat map of no treatment samples. Of the top 120 genes from the Okamoto et al., 2018 ChIP-seq data, we observed that 27 genes from our RNA seq analysis using ACHT hTERT +FANCD2 and ACHT hTERT +FANCG were found to have log2fold changes either < -0.5 or > 0.5 and a P-value < 0.05. Genes with (*) are related with autism, neurodevelopmental or psychiatric disorders.
| Gene Name | log2FoldChange | P-value |
|-----------|----------------|---------|
| TENM4*    | -3.19          | 3.31E-56|
| FAM155A   | -0.98          | 8.66E-24|
| LARGE1    | -0.86          | 1.49E-26|
| FAT3      | -0.71          | 4.84E-08|
| MDGA1*    | -0.57          | 4.53E-08|
| MAGI2     | -0.55          | 2.91E-08|
| AFF2      | -0.53          | 7.74E-05|
| GPC6      | 0.68           | 3.18E-15|
| LSAMP*    | 0.76           | 1.80E-11|
| INPP4B    | 0.84           | 1.38E-17|
| THSD7B    | 0.89           | 6.53E-03|
| PRKN      | 0.97           | 1.57E-02|
| SPOCK1    | 0.99           | 1.54E-68|
| NEGR1*    | 1.03           | 1.73E-20|
| SMYD3     | 1.32           | 2.68E-23|
| CDH4      | 1.41           | 2.00E-05|
| ANKS1B    | 2.64           | 6.75E-07|
| NCAM2*    | 2.81           | 3.27E-95|
| PLCB1*    | 3.27           | 1.51E-31|
| FGF12     | 3.84           | 8.69E-07|
| PAPPA2    | 4.17           | 3.13E-11|
| PTPRD     | 7.99           | 4.41E-21|

**Table 8. RNA sequencing heat map of 24 h APH samples.** Of the top 120 genes from the Okamoto et al., 2018 ChIP-seq data, we observed that 17 genes from our RNA seq analysis using ACHT hTERT +FANCD2 and ACHT hTERT +FANCG were found to have log2foldchanges either <-0.5 or >0.5 and a P-value <0.05. Genes with (*) are related with autism, neurodevelopmental or psychiatric disorders.
CHAPTER 4

DISCUSSION

The protection and tight regulation of our genome is dependent upon the coordination of several DNA repair pathways, including the FA pathway, to ensure the timely repair of DNA damage. FANCD2, one of the key proteins in the FA pathway, has recently been seen to bind to and traverse large actively transcribed genes under conditions of replication stress (Okamoto et al., 2018). We sought to answer the question of why is FANCD2 binding to these genes and if by doing so it has another role in addition to the FA pathway as a potential regulator of transcription.

We used a candidate gene approach targeting SOX5 and WWOX which are found in the top 10 genes FANCD2 was seen to bind to in the Okamoto ChIP-seq dataset (Okamoto et al., 2018). SOX5 was chosen as it is a transcription factor that has been seen to be involved in breast cancer regulation by the transactivation of EZH2 (Sun et al., 2019). It is also involved in the regulation of embryonic development as well as the determination of cell fate (Pei et al., 2014). Choosing a gene involved in breast cancer regulation is relevant as the FA pathway primarily monitors and repairs sites of DNA damage that could otherwise lead to the development of bone marrow failure, leukemia, or premature cancers (Alter et al., 2018). WWOX was also chosen as it is a member of short-chain dehydrogenases and found to span the common fragile site FRA16D (Driouch et al., 2002). It has also been seen that WWOX may play a role in the DNA damage response as a modulator of the DNA damage checkpoint kinase.
ATM. Common fragile sites have an increased risk of being preferential targets of genomic instability such as chromosomal breaks in response to replicative stress (Abu-Odeh et al., 2014)

In this study, we observed that there is an increase in the protein levels of SOX5 under conditions of replication stress and that this increase is significant in the 24-hour APH treatment compared to the no treatment sample. We also did test to see if there was a difference in a 24-hour no treatment sample and 48-hour no treatment sample for both SOX5 and WWOX and did not see differences in the protein levels (Fig. 18). However, using qPCR to analyze the transcript levels of SOX5, we did not see a significant difference when comparing the APH treated samples to the no treatment control sample. There was also little change in WWOX both in protein and transcript levels while FANCD2 was present and under the same conditions of replication stress using APH. It was also found that to observe transcript levels, a large quantity of cDNA (500 ng/µL) was needed for qPCR to observe crossing points that were within the accepted analysis range. A possible reason for the limited changes in the protein levels and overall low abundance of gene expression could be that these two genes are just lowly expressed, and the proteins are very stable in U2OS 3xFLAG cells. It could also be that the proteins for these genes are very stable and therefore the mRNA is not always needed/necessary, so it is in low abundance. It should also be noted that there is the possibility that these genes are so big and as seen in the Okamoto et al., 2018 paper that it can take multiple cell cycles for these genes to be transcribed, potentially causing replication-transcription conflicts. It would be interesting to observe the half-life of the protein for both WWOX and SOX5 to
determine how long each of these can be seen in a cell. If they can be seen for a longer duration of time, it would support the idea that both proteins are stable and therefore less mRNA would be needed.

Further analysis using an siFANCD2 to knockdown FANCD2 expression in U2OS 3xFLAG cells as well as aphidicolin treatment, resulted in no difference in protein expression levels regardless of FANCD2 presence for WWOX. This experiment was performed to determine the impact the absence of FANCD2 could have had on both SOX5 and WWOX genes and proteins. Protein expression for SOX5 was seen to increase under APH conditions in the presence of FANCD2 while without FANCD2, there was slight decrease in the recovery sample, but the decrease was not significant. Interestingly, our experiments indicate that both SOX5 and WWOX transcript levels are increased in the absence of the FANCD2 protein. These results suggest that in the absence of FANCD2, WWOX and SOX5 transcript levels are increasing suggesting that FANCD2 may be monitoring their transcription and without it, the genes are less controlled. If the transcription of WWOX and SOX5 is less controlled, they could be transcribed even when they are not necessarily needed. As these genes are prone to CNVs and WWOX overlaps with FRA16D, the uncontrolled transcription of them could lead to an increase in genome instability and more susceptibility to cancer.

As the differences seen in our results are minor, with some of the differences not being significant, it brought us to the question of is FANCD2 regulating transcription and if not, what is it doing at these two gene sites? It could be that FANCD2 is regulating replication instead of transcription and helps in the stabilization
of the DNA at sites that are more prone to DNA damage. It has been previously seen that in the absence of FANCD2, replication forks stall within the AT-rich fragility core of the FRA16D CFS and this can lead to the dormant origin activation (Madireddy et al., 2016). It has also been previously seen that under conditions of replication stress, FANCD2 is able to promote DNA replication (Lossaint et al., 2013).

It has also been seen that FANCD2 may play a role in the regulation of R-loops. R-loops, which are DNA: RNA hybrids are hazards to the genome and can increase the probability of disruption of chromatin organization. The displaced ssDNA in R-loops also poses a threat to the stability of the genome as ssDNA is more prone to attracting nucleases (Sollier & Cimprich, 2015). However, R-loops can also be non-hazardous in the genome as they are required for immunoglobulin class switching (García-Muse & Aguilera, 2016). Previously, Okamoto et al., 2018 also saw that FANCD2 accumulates at large transcribed genes and that its accumulation may depend on the formation of R-loops which can be induced by transcription-replication collisions during mild replication stress. It would be interesting to determine if FANCD2 is proximal to where these events are occurring and comparing the same areas in the absence of FANCD2 under mild replication stress.

To study the role of FANCD2 in transcriptional regulation at a genome-wide level, we also performed RNA-seq analysis using ACHT hTERT FANCD2 and ACHT hTERT FANCG (FANCD2-/-) cells. We chose to use these cells because they are FA patient-derived skin fibroblasts unlike U2OS cells, which are derived from an osteosarcoma. ACHT cells therefore more closely resemble those of an FA patient. Using RNA sequencing we were able to analyze not only SOX5 and WWOX but also
the other 120 top gene hits from the Okamoto et al., paper. We observed that SOX5 was not expressed, and a small expression of WWOX was seen but with no significant difference with/without the presence of FANCD2 (Fig. 17).

Upon further analysis of our RNA sequencing data, we found that of the 120 genes in the Okamoto et al., 2018 data set, 79 of the ChIP-seq genes were expressed in the RNA-seq data in the no treatment sample and 66 genes in the 24-hour APH treated samples. Of those 79 genes found to overlap, 26 genes were found to have log2fold changes less than -0.05 or greater than 0.05 in the no treatment sample and 22 in the 24-hour APH treated samples (Tables 7 & 8). Of the 26 genes we also saw that 7 were downregulated meaning that there is an increase in the ACHT hTERT FANCG’s and 19 were upregulated meaning that there is an increase in the ACHT hTERT FANCD2 samples (Table 9). In the APH treated samples, it was found that 7 genes were being downregulated and 15 were upregulated (Table 10). Interestingly, 7 of the 26 in the no treatment samples and 6 of the 22 genes found in the APH treated samples are found to be related to autism, neurodevelopmental, psychiatric disorders. These findings and differences seen in the RNA sequencing data support the notion that FANCD2 could be playing a role in the transcriptional regulation of some genes.

Furthermore, it was found that a portion of the significantly differentially expressed genes are related to nervous system development or nervous system conditions, therefore suggesting that FANCD2 could also be contributing to the regulation of neuronal derived genes.

We also input the genes found to have significant differences in gene expression into the STRING database to determine their relationship to one another.
for both the no treatment samples and the APH treated samples (Figs. 19 & 20). If there is a line between nodes, this indicates that there is an interaction between the proteins. If there is no line, there is no interaction. The color of the line also represents different associations with blue being that it is a known interaction from a curated database, pink is experimentally determined, green is a gene neighborhood, red is gene fusions, blue is gene co-occurrence, yellow is textmining, black is co-expression, and purple is protein homology. Many of the proteins displaying an interaction are associated with neuronal development such as the link between CTNND2, PTPRD, ASTN2, and NLGN1 in the no treatment samples as seen in Fig. 18. All four of these proteins play a role in neuronal development. Looking at the 24-hour APH samples it can also be seen that there is a link between LSAMP, NEGR1, and MDGA1 which are also related to neuronal functions (Fig. 20).

This analysis using the STRING database and our RNA sequencing data reveals that there are proteins that have a significant difference in gene expression between ACHT hTERT +FANCD2 and ACHT hTERT +FANCG samples that are directly linked to neurological development functions. The STRING database is able to integrate all known and predicted associations between proteins including but physical and functional interactions. This is interesting because it further proposes the possibility that FANCD2 has another role distinct from ICL repair in the FA pathway and that the role could involve regulating neuronal development.

It was also observed that among the top differentially expressed genes clustered by gene ontology, nervous system development was seen as the 5th most differentially expressed under 24 hours of APH treatment (Fig. 21). The gene ontology
analysis aims to identify the biological processes, cellular locations, and molecular functions that have been impacted in the condition studied. This analysis is able to reduce the complexity and instead highlight the biological processes in the genome-wide expression studied. Identifying nervous system development as one of the systems affected in this study which utilized two cell lines with one lacking FANCD2 additionally supports a hypothesis of FANCD2 assisting with neuronal development and having another purpose apart from the FA pathway.

**Future Experiments**

We have made significant progress determining the consequences of replication stress in the presence and absence of FANCD2 for WWOX and SOX5 and if FANCD2 could be involved in the transcriptional regulation of these two genes. However, these were only two of the 120 genes that FANCD2 was found to bind to in the ChIP-seq data (Okamoto et al., 2018). Here, we will propose several experiments to further answer our question of why FANCD2 is binding and traversing large actively transcribed genes under replication stress conditions.

We would further like to determine whether FANCD2 is also involved in the regulation of R-loops as these as previously stated, also pose a threat to genomic stability. The S9.6 antibody can be used to detect the presence of R-loops and it has been previously seen that the members of the FA pathway (BRCA2/FANCD1) accumulate at R-loops (García-Rubio et al., 2015). Since the FA pathway is involved in the repair of interstrand crosslinks which block replication fork progression, this suggests that because R-loops can also block replication progression, that these R-
loops may be a contributor to genomic instability in FA cells (Schwab et al., 2015). Schwab et al., 2015 also found that in cells with downregulated FANCD2, there was a significant increase in asymmetric sister forks compared to their control cells which suggests that there is a deviation from the normal replication program. This also suggests that FANCD2 may play a role in the regulation of replication. A previous study also found that in cells deficient in FANCD2 and FANCA, there was an increase in the number of R-loops suggesting that these two proteins play a role in R-loop regulation (García-Rubio et al., 2015). It has also been previously found that the FANCD2: FANCI heterodimer (ID2) preferentially binds ssRNA but not RNA: DNA hybrids making it seem as though ID2 binds to the ssRNA that is displaced in R-loops (Liang et al., 2019).

We would also like to further investigate other genes, separate from the Okamoto dataset, that are found in our RNA sequencing data and have normalized read counts that are significantly different. In the no treatment conditions, we have examined the data and created a list of potential gene targets from different cell pathways and families such as retinaldehyde dehydrogenases. Of the genes found to have significant differences in normalized read counts in the RNA sequencing data, we will perform western blots and qPCR on a subset of these genes to further confirm the differences seen in the RNA sequencing data. We are also planning to perform ChIP-seq on the ACHT hTERT +FANCD2 and ACHT hTERT +FANCG under APH conditions to be able to compare to the U2OS 3xFLAG ChIP-seq data. This will be important to be able to tell how different the two cell lines and determine if the genes
that FANCD2 was found to bind to in the U2OS 3xFLAG cells are similar or different in ACHTs.

We would also like to use a neurological stem cell line and using our siFANCD2, determine the consequences of the absence of FANCD2 on a number of genes found in the RNA sequencing data. We found in our RNA sequencing data, that the development of neurological systems was found to be fifth in differentially expressed gene ontology and of the 22 genes found to overlap with the ChIP sequencing data and our RNA sequencing data under APH conditions, 6 of the 22 gene were found to be related with autism, neurodevelopmental, or psychiatric disorders (Table 8). Using the neurological stem cell line would help us understand more of whether FANCD2 is an integral part in maintaining neurological functions. We would use qPCR as well as western blotting to detect if there are differences with or without the presence of FANCD2 as well as in the presence or absence of APH.
Figure 18. U2OS 3xFLAG cells grown for either 24 hours or 48 hours with no treatment (NT) added. Abundance of SOX5 and WWOX proteins present in whole-cell lysates was determined by immunoblot using an anti-SOX5 and anti-WWOX antibodies. Tubulin, detected with an anti-Tubulin antibody, serves as a loading control.
**Figure 19. RNA sequencing reveals potential link between FANCD2 and neurological development in no treatment samples.** Cells were harvested, flash frozen using liquid nitrogen and sent to Genewiz for RNA sequencing analysis at 150 bp paired end sequencing at 60 million reads per sample. Using the differential analysis datasets for the no treatment samples provided from Genewiz, it was found that 26 of the 120 Okamoto et al., 2018 ChIP-seq genes FANCD2 was found to bind to overlap with the RNA seq data for the no treatment samples. Of the 26 genes, 7 of them are found to be related with autism, neurodevelopmental, or psychiatric disorders. The 26 genes were input to STRING database to see if they connect to one another. The color of the line also represents different associations with blue being that it is a known interaction from a curated database, pink is experimentally determined, green is a gene neighborhood, red is gene fusions, blue is gene co-occurrence, yellow is textmining, black is co-expression, and purple is protein homology.
Figure 20. RNA sequencing reveals potential link between FANCD2 and neurological development in aphidicolin treated samples. Cells were harvested, flash frozen using liquid nitrogen and sent to Genewiz for RNA sequencing analysis at 150 bp paired end sequencing at 60 million reads per sample. Using the differential analysis datasets for the 24-hour APH samples provided from Genewiz, it was found that 22 of the 120 Okamoto et al., 2018 ChIP-seq genes FANCD2 was found to bind to overlap with the RNA seq data for the no treatment samples. Of the 22 genes, 6 of them are found to be related with autism, neurodevelopmental, or psychiatric disorders. The 22 genes were input to STRING database to observe the relationship between one another. The color of the line also represents different associations with blue being that it is a known interaction from a curated database, pink is experimentally determined, green is a gene neighborhood, red is gene fusions, blue is gene co-occurrence, yellow is textmining, black is co-expression, and purple is protein homology.
Figure 21. RNA-seq data displaying 24-hour aphidicolin (APH) differentially expressed genes categorized by gene ontology. ACHT hTERT FANCG (FANCD2 -/-) cells and ACHT hTERT FANCD2 were incubated in the absence or presence of 0.4 µM aphidicolin (APH) for 24 and 48 hours in T-75 flasks. Cells were harvested, flash frozen using liquid nitrogen and sent to Genewiz for RNA sequencing analysis at 150 bp paired end sequencing at 600 million reads per sample. Genewiz analysis provided a differentially expressed genes by gene ontology table in which it was found that the 5th most differentially expressed was nervous system development in the 24-hour APH samples.
Abu-Odeh, M., Salah, Z., Herbel, C., Hofmann, T. G., & Aqeila, R. I. (2014). WWOX, the common fragile site FRA16D gene product, regulates ATM activation and the DNA damage response. *Proceedings of the National Academy of Sciences of the United States of America, 111*(44), E4716–E4725.
https://doi.org/10.1073/pnas.1409252111

Alter, B. P., Giri, N., Savage, S. A., & Rosenberg, P. S. (2018). Cancer in the national cancer institute inherited bone marrow failure syndrome cohort after fifteen years of follow-up. *Haematologica, 103*(1), 30–39.
https://doi.org/10.3324/haematol.2017.178111

Barlow, J. H., Faryabi, R. B., Callén, E., Wong, N., Malhowski, A., Chen, H. T., Gutierrez-Cruz, G., Sun, H. W., McKinnon, P., Wright, G., Casellas, R., Robbiani, D. F., Staudt, L., Fernandez-Capetillo, O., & Nussenzweig, A. (2013). Identification of early replicating fragile sites that contribute to genome instability. *Cell, 152*(3), 620–632.
https://doi.org/10.1016/j.cell.2013.01.006

Byun, T. S., Pacek, M., Yee, M. C., Walter, J. C., & Cimprich, K. A. (2005). Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. *Genes & development, 19*(9), 1040–1052.
https://doi.org/10.1101/gad.1301205

Chan, K. L., Palmai-Pallag, T., Ying, S., & Hickson, I. D. (2009). Replication stress induces sister-chromatid bridging at fragile site loci in mitosis. *Nature Cell Biology, 11*(6), 753–760. https://doi.org/10.1038/ncb1882
Debatisse, M., le Tallec, B., Letessier, A., Dutrillaux, B., & Brison, O. (2012). Common fragile sites: Mechanisms of instability revisited. In Trends in Genetics (Vol. 28, Issue 1, pp. 22–32). Trends Genet. https://doi.org/10.1016/j.tig.2011.10.003

Driouch, K., Prydz, H., Monese, R., Johansen, H., Lidereau, R., & Frengen, E. (2002). Alternative transcripts of the candidate tumor suppressor gene, WWOX, are expressed at high levels in human breast tumors. Oncogene, 21(12), 1832–1840. https://doi.org/10.1038/sj.onc.1205273

Feng, G., Yuan, Y., Li, Z., Wang, L., Zhang, B., Luo, J., Ji, J., & Kong, D. (2019). Replication fork stalling elicits chromatin compaction for the stability of stalling replication forks. Proceedings of the National Academy of Sciences of the United States of America, 116(29), 14563–14572. https://doi.org/10.1073/pnas.1821475116

Fernandes, P., Miotto, B., Saint-Ruf, C., Said, M., Barra, V., Nähse, V., Ravera, S., Cappelli, E., & Naim, V. (2021). FANCD2 modulates the mitochondrial stress response to prevent common fragile site instability. Communications biology, 4(1), 127. https://doi.org/10.1038/s42003-021-01647-8

Fiesco-Roa, M. O., Giri, N., McReynolds, L. J., Best, A. F., & Alter, B. P. (2019). Genotype-phenotype associations in Fanconi anemia: A literature review. In Blood Reviews (Vol. 37, p. 100589). Churchill Livingstone. https://doi.org/10.1016/j.blre.2019.100589

Garcia-Higuera, I., Taniguchi, T., Ganesan, S., Meyn, M. S., Timmers, C., Hejna, J., Grompe, M., & D’Andrea, A. D., (2001). Interaction of the Fanconi anemia proteins and BRX1 in a common pathway. Molecular cell, 7(2), 249-262. https://doi.org/10.1016/s1097-2765(01)00173-3
García-Muse, T., & Aguilera, A. (2016). Transcription-replication conflicts: How they occur and how they are resolved. In *Nature Reviews Molecular Cell Biology* (Vol. 17, Issue 9, pp. 553–563). Nature Publishing Group. https://doi.org/10.1038/nrm.2016.88

García-Rubio, M. L., Pérez-Calero, C., Barroso, S. I., Tumini, E., Herrera-Moyano, E., Rosado, I. v., & Aguilera, A. (2015). The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. *PLoS Genetics*, 11(11), 1005674. https://doi.org/10.1371/journal.pgen.1005674

Glover, T. W., Wilson, T. E., & Arlt, M. F. (2017). Fragile sites in cancer: More than meets the eye. In *Nature Reviews Cancer* (Vol. 17, Issue 8, pp. 489–501). Nature Publishing Group. https://doi.org/10.1038/nrc.2017.52

Hamperl, S., & Cimprich, K. A. (2016). Conflict Resolution in the Genome: How Transcription and Replication Make It Work. In *Cell* (Vol. 167, Issue 6, pp. 1455–1467). Cell Press. https://doi.org/10.1016/j.cell.2016.09.053

Hänsel-Hertsch, R., di Antonio, M., & Balasubramanian, S. (2017). DNA G-quadruplexes in the human genome: Detection, functions and therapeutic potential. *Nature Reviews Molecular Cell Biology*, 18(5), 279–284. https://doi.org/10.1038/nrm.2017.3

Hays, L., Frohnmayer, D., Frohnmayer, L., Larsen, K., & Owen, J. (2014). Fanconi Anemia: Guidelines for Diagnosis and Management. *Fanconi Anemia Research Fund, Inc.*, 1–391. http://www.ncbi.nlm.nih.gov/books/NBK1401/

Helmrich, A., Ballarino, M., & Tora, L. (2011). Collisions between Replication and Transcription Complexes Cause Common Fragile Site Instability at the Longest Human Genes. *Molecular Cell*, 44(6), 966–977. https://doi.org/10.1016/j.molcel.2011.10.013
Howlett, N. G., Taniguchi, T., Durkin, S. G., D’Andrea, A. D., & Glover, T. W. (2005). The Fanconi anemia pathway is required for the DNA replication stress response and for the regulation of common fragile site stability. Human Molecular Genetics, 14(5), 693–701. https://doi.org/10.1093/hmg/ddi065

Huang, Y., Leung, J. W. C., Lowery, M., Matsushita, N., Wang, Y., Shen, X., Huong, D., Takata, M., Chen, J., & Li, L. (2014). Modularized Functions of the Fanconi Anemia Core Complex. Cell Reports, 7(6), 1849–1857. https://doi.org/10.1016/j.celrep.2014.04.029

Joo, W., Xu, G., Persky, N. S., Smogorzewska, A., Rudge, D. G., Buzovetsky, O., Elledge, S. J., & Pavleitch, N. P. (2011). Structure of the FANCI-FANCD2 complex: Insights into the Fanconi anemia DNA repair pathway. Science, 333(6040), 312–316. https://doi.org/10.1126/science.1205805

Kalb, R., Neveling, K., Hoehn, H., Schneider, H., Linka, Y., Batish, S. D., Hunt, C., Berwick, M., Callen, E., Surralles, J., Casado, J. A., Bueren, J., Dasi, A., Soulier, J., Gluckman, E., Zwaan, C. M., van Spaendonk, R., Pals, G., de Winter, J. P., Joenje, H., Schindler, D. (2007). Hypomorphic mutations in the gene encoding a key Fanconi anemia protein, FANCD2, sustain a significant group of FA-D2 patients with severe phenotype. American journal of human genetics, 80(5), 895–910. https://doi.org/10.1086/517616

Kee, Y., & D’Andrea, A. D. (2012). Molecular pathogenesis and clinical management of Fanconi anemia. In Journal of Clinical Investigation (Vol. 122, Issue 11, pp. 3799–3806). American Society for Clinical Investigation. https://doi.org/10.1172/JCI58321
Kim, H., & D’Andrea, A. D. (2012). Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway. In Genes and Development (Vol. 26, Issue 13, pp. 1393–1408). Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/gad.195248.112

Kim, Y., Lach, F. P., Desetty, R., Hanenberg, H., Auerbach, A. D., & Smogorzewska, A. (2011). Mutations of the SLX4 gene in Fanconi anemia. In Nature Genetics (Vol. 43, Issue 2, pp. 142–146). Nature Publishing Group. https://doi.org/10.1038/ng.750

Le Tallec, B., Koundrioukoﬀ, S., Wilhelm, T., Letessier, A., Brison, O., & Debatisse, M. (2014). Updating the mechanisms of common fragile site instability: how to reconcile the different views? Cellular and molecular life sciences: CMLS, 71(23), 4489-4494. https://doi.org/10.1007/s00018-01401720-2

Liang, Z., Liang, F., Teng, Y., Chen, X., Liu, J., Longerich, S., Rao, T., Green, A. M., Collins, N. B., Xiong, Y., Lan, L., Sung, P., & Kupfer, G. M. (2019). Binding of FANCI-FANCD2 Complex to RNA and R-Loops Stimulates Robust FANCD2 Monoubiquitination. Cell Reports, 26(3), 564-572.e5. https://doi.org/10.1016/j.celrep.2018.12.084

Lossaint, G., Larroque, M., Ribeyre, C., Bec, N., Larroque, C., Décaillot, C., Gari, K., & Constantinou, A. (2013). FANCD2 binds MCM proteins and controls replisome function upon activation of s phase checkpoint signaling. Molecular cell, 51(5), 678–690. https://doi.org/10.1016/j.molcel.2013.07.023

Madireddy, A., Kosiyatrakul, S. T., Boisvert, R. A., Herrera-Moyano, E., García-Rubio, M. L., Gerhardt, J., Vuono, E. A., Owen, N., Yan, Z., Olson, S., Aguilera, A., Howlett, N. G., & Schildkraut, C. L. (2016). FANCD2 Facilitates Replication through Common
Fragile Sites. *Molecular Cell*, 64(2), 388–404. https://doi.org/10.1016/j.molcel.2016.09.017

Maffia, A., Ranise, C., & Sabbioneda, S. (2020). From R-loops to G-quadruplexes: Emerging new threats for the replication fork. In *International Journal of Molecular Sciences* (Vol. 21, Issue 4). MDPI AG. https://doi.org/10.3390/ijms21041506

Magdalou, I., Lopez, B. S., Pasero, P., & Lambert, S. A. (2014). The causes of replication stress and their consequences on genome stability and cell fate. *Seminars in cell & developmental biology*, 30, 154-164. https://doi.org/10.1016/j.semcdb.2014.04.035

Mankouri, H. W., Huttner, D., & Hickson, I. D. (2013). How unfinished business from S-phase affects mitosis and beyond. In *EMBO Journal* (Vol. 32, Issue 20, pp. 2661–2671). John Wiley & Sons, Ltd. https://doi.org/10.1038/emboj.2013.211

Milletti, G., Strocchio, L., Pagliara, D., Girardi, K., Carta, R., Mastronuzzi, A., Locatelli, F., & Nazio, F. (2020). Canonical and noncanonical roles of fanconi anemia proteins: Implications in cancer predisposition. In *Cancers* (Vol. 12, Issue 9, pp. 1–23). MDPI AG. https://doi.org/10.3390/cancers12092684

Niraj, J., Färkkilä, A., & D’Andrea, A. D. (2019). The fanconi anemia pathway in cancer. In *Annual Review of Cancer Biology* (Vol. 3, Issue 1, pp. 457–478). Annual Reviews Inc. https://doi.org/10.1146/annurev-cancerbio-030617-050422

Okamoto, Y., Hejna, J., & Takata, M. (2019). Regulation of R-loops and genome instability in Fanconi anemia. In *Journal of Biochemistry* (Vol. 165, Issue 6, pp. 465–470). Oxford University Press. https://doi.org/10.1093/jb/mvz019

Okamoto, Y., Iwasaki, W. M., Kugou, K., Takahashi, K. K., Oda, A., Sato, K., Kobayashi, W., Kawai, H., Sakasai, R., Takaori-Kondo, A., Yamamoto, T., Kanemaki, M. T.,
Taoka, M., Isobe, T., Kurumizaka, H., Innan, H., Ohta, K., Ishiai, M., & Takata, M. (2018). Replication stress induces accumulation of FANCD2 at central region of large fragile genes. *Nucleic Acids Research, 46*(6), 2932–2944. https://doi.org/10.1093/nar/gky058

Pei, X. H., Lv, X. Q., & Li, H. X. (2014). Sox5 induces epithelial to mesenchymal transition by transactivation of Twist1. *Biochemical and Biophysical Research Communications, 446*(1), 322–327. https://doi.org/10.1016/j.bbrc.2014.02.109

Scheckenbach, K., Morgan, M., Filger-Brillinger, J., Sandmann, M., Strimling, B., Scheuren, W., Schindler, D., Göbel, U., & Hanenberg, H. (2012). Treatment of the bone marrow failure in Fanconi anemia patients with danazol. *Blood cells, molecules & diseases, 48*(2), 128-131. https://doi.org/10.1016/j.bcmd.2011.11.006

Schlacher, K., Wu, H., & Jasin, M. (2012). A Distinct Replication Fork Protection Pathway Connects Fanconi Anemia Tumor Suppressors to RAD51-BRCA1/2. *Cancer Cell, 22*(1), 106–116. https://doi.org/10.1016/j.ccr.2012.05.015

Schwab, R. A., Nieminuszczzy, J., Shah, F., Langton, J., Lopez Martinez, D., Liang, C. C., Cohn, M. A., Gibbons, R. J., Deans, A. J., & Niedzwiedz, W. (2015). The Fanconi Anemia Pathway Maintains Genome Stability by Coordinating Replication and Transcription. *Molecular cell, 60*(3), 351-361. https://doi.org/10.1016/j.molcel.2015.09.012

Shima, N., & Pederson, K. D. (2017). Dormant origins as a built-in safeguard in eukaryotic DNA replication against genome instability and disease development. In *DNA Repair* (Vol. 56, pp. 166–173). Elsevier B.V. https://doi.org/10.1016/j.dnarep.2017.06.019
Sollier, J., & Cimprich, K. A. (2015). Breaking bad: R-loops and genome integrity. In *Trends in Cell Biology* (Vol. 25, Issue 9, pp. 514–522). Elsevier Ltd. https://doi.org/10.1016/j.tcb.2015.05.003

Stoepker, C., Hain, K., Schuster, B., Hilhorst-Hofstee, Y., Rooimans, M. A., Steltenpool, J., Oostra, A. B., Eirich, K., Korthof, E. T., Nieuwint, A. W. M., Jaspers, N. G. J., Bettecken, T., Joenje, H., Schindler, D., Rouse, J., & de Winter, J. P. (2011). SLX4, a coordinator of structure-specific endonucleases, is mutated in a new Fanconi anemia subtype. In *Nature Genetics* (Vol. 43, Issue 2, pp. 138–141). Nat Genet. https://doi.org/10.1038/ng.751

Sun, C., Ban, Y., Wang, K., Sun, Y., & Zhao, Z. (2019). SOX5 promotes breast cancer proliferation and invasion by transactivation of EZH2. Oncology Letters, 17, 2754-2762. https://doi.org/10.3892/ol.2019.9914

Tan, W., van Twest, S., Leis, A., Bythell-Douglas, R., Murphy, V. J., Sharp, M., Parker, M. W., Crismani, W., & Deans, A. J. (2020). Monoubiquitination by the human Fanconi anemia core complex clamps FANCI:FANCD2 on DNA in filamentous arrays. *eLife*, 9, e54128. https://doi.org/10.7554/eLife.54128

Taniguchi, T., Garcia-Higuera, I., Andreassen, P.R., Gregory, R. C., Grompe, M., & D’Andrea, A. D. (2002). S-phase-specific interaction of the Fanconi anemia protein, FANCD2, with BRCA1 and RAD51. *Blood*, 100(7), 2414-2420. https://doi.org/10.1182/blood-2002-01-0278

Wilson, T., E., Arlt, M. F., Park, S. H., Rajendran, S., Paulsen, M., Ljungman, M., & Glover, T. W. (2015). Large transcription units unify copy number variants and
common fragile sites arising under replication stress. *Genome research*, 25(2), 189-200. [https://doi.org/10.1101/gr.177121.114](https://doi.org/10.1101/gr.177121.114)

Zeman, M. K., & Cimprich, K. A. (2014). Causes and consequences of replication stress. *Nature cell biology*, 16(1), 2–9. [https://doi.org/10.1038/ncb2897](https://doi.org/10.1038/ncb2897)