AWARD NUMBER: W81XWH-15-1-0204

TITLE: Global Identification of Disease-Associated Genes in Fragile X Cells

PRINCIPAL INVESTIGATOR: Wenyi Feng

CONTRACTING ORGANIZATION: Research Foundation of State University
Syracuse, NY 13210

REPORT DATE: March 2017

TYPE OF REPORT: FINAL

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
March 2017	Final	1 Jul 2015 - 31 Dec 2016

4. TITLE AND SUBTITLE

Global Identification of Disease-Associated Genes in Fragile X Cells

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-15-1-0204

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Wenyi Feng

E-Mail: fengw@upstate.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

State University of New York 750 East Adams Street
Upstate Medical University Syracuse, NY 13210

Research Foundation of State University
Syracuse, NY 13210

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The aim of the proposed study is to test the hypothesis that the Fragile X mental retardation protein (FMRP) prevents/resolves R loop formation to maintain genome stability. Specifically we propose that stable R loop formation impedes replication fork progression, resulting in DNA double strand breaks (DSBs), and that FMRP functions to prevent such replication-trancription conflict. We have performed three biological replicate experiments to rigorously test if the Fragile X cell line produces more DSBs than the normal control. We also developed a yeast-based recombination assay to directly test the proposed function of FMRP in R loop prevention/resolution. Finally, we performed a ChIP-seq experiment to identify the chromatin binding sites of FMRP. We are working towards obtaining a short list of genes with overlapping DSBs, R loop forming sites and FMRP-binding sites. Potential disease-correlation of these genes will then be assessed.

15. SUBJECT TERMS

FMRP; Fragile X; Genome instability; DNA double strand breaks; R loops; Break-seq; ChIP-seq

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	18	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

1. Introduction

The proposed research project addresses the FY14 PRMRP Topic Area, Fragile X syndrome (FXS). FXS is responsible for the most common form of hereditary mental retardation in the world. Previous studies have established that in most cases FXS is caused by the absence of a protein named FMRP. It is generally believed that FMRP functions in regulating protein production of specific gene targets in the cytoplasm. However, we know that FMRP shuttles between the nucleus and the cytoplasm. To this date, the potential functions of FMRP in the nucleus have not been well examined. We had made a chance discovery that fragile X cells lacking FMRP show high level of chromosome breaks specifically at DNA sequences that are prone to form hybrid molecules of DNA and RNA during gene transcription. This finding led us to propose that FMRP binds to its substrates directly on the chromatin (in the nucleus) and ensures that during their transcription the DNA templates do not form stable DNA:RNA hybrids. The hybrid molecules are impediments to the DNA replication machinery and upon their collision chromosomes are more prone to break, ultimately affecting gene expression and/or protein production of genes near the chromosome breaks. We propose to directly test this hypothesis by querying the genome for FMRP binding sites. These experiments will allow us to identify those specific gene substrates of FMRP, particularly those expressed in the brain, that are implicated in FXS progression. Moreover, we use an innovative tool to directly map chromosome breaks on a genome-wide scale. These proposed experiments will yield a novel list of FXS-associated genes and provide a foundation for future research examining the mechanisms of FXS pathology.

2. Keywords

Fragile X syndrome (FXS)

Fragile X mental retardation protein (FMRP)

R loops (DNA:RNA hybrids)

DNA replication

Gene transcription

DNA double strand breaks (DSBs)

Genome instability

3. Accomplishments

- **What were the major goals of the project?**

  We proposed three major goals for this study listed as follows:

  1) Validate the co-localized R-loop formation and chromosome fragility in Fragile X cells, particularly at the brain-expressed genes, by ChIP-seq (detecting DNA:RNA hybrids) and Break-seq (detection DNA double strand breaks, DSBs), respectively.
  2) Determine the genomic binding sites of FMRP using ChIP-seq.
  3) Test the predicted function of FMRP in preventing/resolving R-loops during replication-transcription conflict in a yeast-based recombination assay.
  4) Examine potential changes of gene expression and protein levels of the putative FXS-associated genes in fragile X cells compared to normal cells.

- **What was accomplished under these goals?**

  Below I list the experiments and conclusions for each goal in the same order as listed above:

  1) We previously performed **three** biological replicate experiments of Break-seq in Fragile X cells compared to normal control cells to validate the observed correlation between DSBs and R-loop forming sites (RLFSs). In these experiments we induced DNA replication stress by varying concentrations of aphidicolin, a DNA polymerase inhibitor, to ascertain potentially different chromosome breakage profiles. All samples were treated with equal volumes of the vehicle, dimethylsulfoxide (DMSO). We also performed Break-seq on untreated samples (no aphidicolin and no vehicle) as controls. Finally, we compared cells harvested at the end of the 24-hour exposure to aphidicolin with or without metaphase arrest by colcemid (typically employed in cytological studies with metaphase chromosome spread preparations). The varying conditions of the experiments are summarized in Table 1 (no treatment controls are omitted in this table). These culture conditions represent the most comprehensive design for genome-wide mapping of chromosome breaks to the best of our knowledge). *In the second funding period (Sep 2016 through Dec 2016) we also initiated a fourth biological*

replicate experiment to reproduce the third experiment involving colcemid treatment. Results from this experiment are currently being analyzed.

Table 1. Summary of Break-seq experiments.

Biological replicate No.	Sample index	Aphidicolin concentration (μM)	Colcemid ("-", not added; "+", added)
1 (BH collection)	AMc121614	0, 0.03, 0.3	-
2 (AC collection)	AMc051915	0, 0.03, 0.3	-
3 (AC collection)	AMc050916	0, 0.03, 0.3, 0.6, 1.5	-/+
4 (AC collection)	AMc033117	0, 0.03, 0.3, 0.6	-/+

These experiments led to the following conclusions. **First**, our results confirmed that the Fragile X cells indeed accumulate more DSBs than normal control under replication stress. **Second**, it appeared that different concentrations of aphidicolin produced different spectra of DSBs (*we have confirmed this observation with results from the third biological replicate experiment since the last funding period*). We reasoned that the gene expression patterns are also different with increasing drug concentration, thus producing R loop formation at different sites in the genome. We are prepared to test this hypothesis with gene expression studies by RNA-seq as we collected additional samples from two of the three biological experiments. We have not yet performed RNA-seq on these samples. **Third**, our results indicated that DSBs are indeed correlated with the computationally predicted RLFSs from the first two biological replicate experiments (the third is being analyzed). We have not yet validated these findings with DRIP-seq experiments to directly identify R-loop formation based on DNA:RNA hybrid detection.

2) We have performed a ChIP-seq experiment (biological replicates are upcoming) in the Fragile X cell line compared to a normal control (Fig. 1). We first validated using Western blots that the Fragile X cell line indeed lacks detectable FMRP expression. Then, the ChIP-seq signals from the normal control cell line were first normalized to the whole chromatin control (no immunoprecipitation with anti-FMRP antibody), and then normalized against the ChIP-seq signals from the Fragile X cell line to eliminate background detection. We identified 5238 FMRP-binding sites with approximately equal

division between genic and non-genic sequences. Interestingly, only 148 FMRP-binding sites overlapped with an RLFS in a total of 191 genes. This result indicated that genes with co-localized FMRP binding sites and RLFSs were under-represented in the human genome ($p < 2.2E-16$ in a Fisher's exact test). Indeed, those FMRP-binding sites with overlapping R-loop

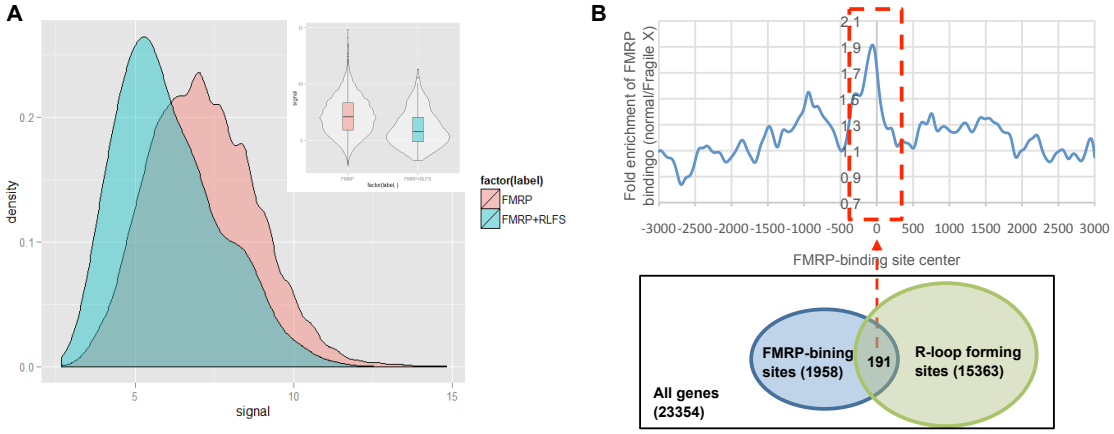

Figure 1. (**A**) Distributions of relative fractions of FMRP chromatin binding sites with (pink) or without (green) overlapping RLFSs. Inset shows the median levels of FMRP chromatin binding in the two categories. (**B**) Meta-analysis of RLFS distribution over a 6 kb window centering on the middle of all FMRP binding sites. RLFS density is relatively higher in a ± 300 bp window from the middle of a FMRP-binding site.

forming sites showed relatively lower ChIP-seq signals or FMRP-binding levels than the genomic average FMRP-binding level). However, further analysis indicated that FMRP-binding sites were enriched in RLFS-flanking regions ((±300 bp to 3 kb from RLFS center). Therefore, these results suggest that FMRP binding sites tend to be localized to regions adjacent to R loom forming sites.

We also added an experiment to test the functional role of FMRP in R loop regulation. We reasoned that mere identification of chromatin-binding sites of FMRP does not represent a functional test of the protein in R loop regulation. Therefore, we took advantage of our expertise in yeast genetics and implemented a plasmid-based recombination assay to directly test 1) the ability of human R loop forming sequences to induce DSBs and increase recombination frequency; and 2) the ability of FMRP to reduce R loop-induced DSB formation and recombination frequency. *We have completed this analysis and our data demonstrate that FMRP expression indeed reduces the R-loop-induced DSB and recombination frequency (**Fig. 2** & **Fig. 3**).*

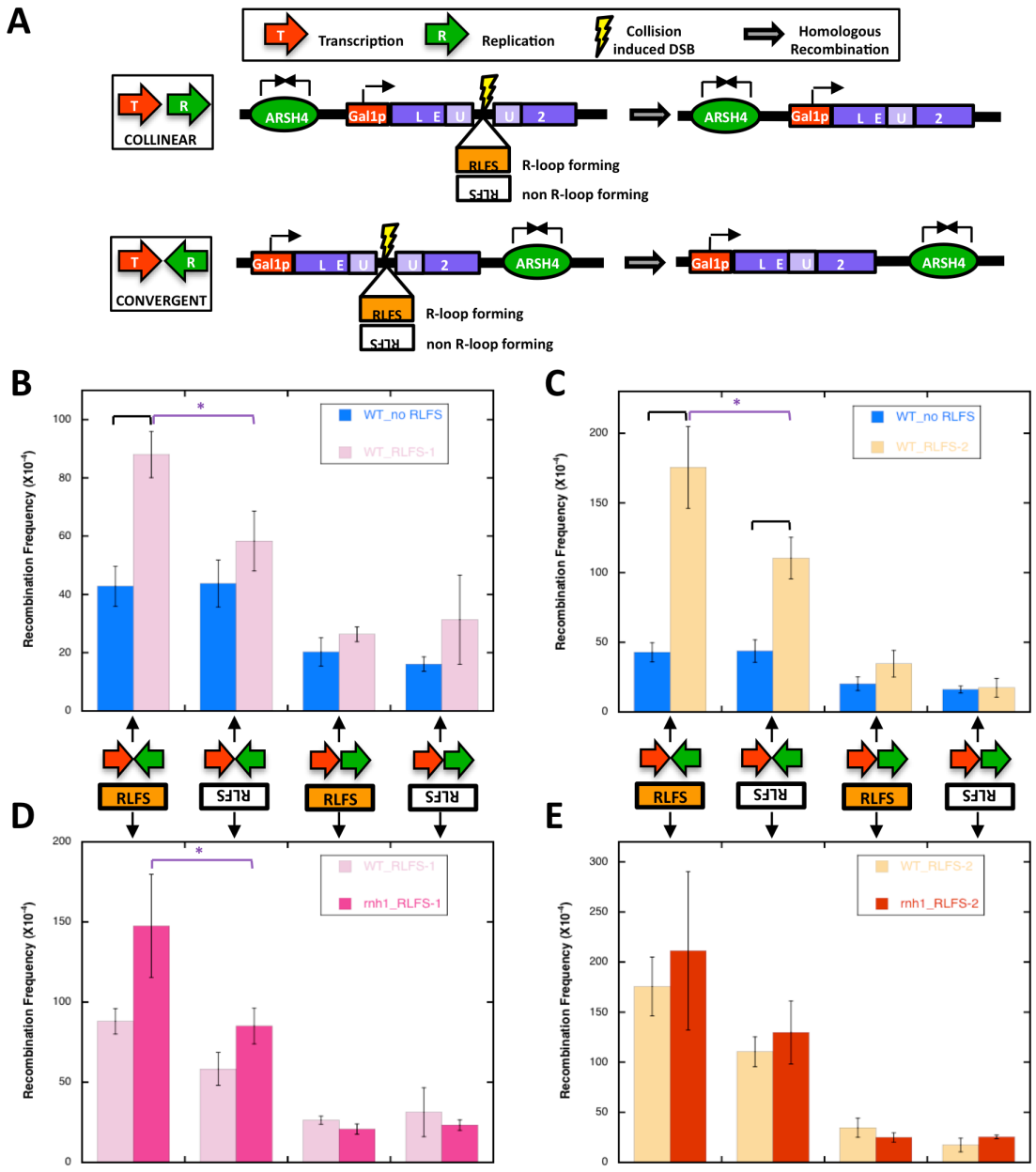

Figure 2. **Recombination assay measuring DSB frequency due to R-loop formation. (A)** Plasmid constructs containing an origin of replication (ARSH4) and a Gal1-induced non-functional LEU2 transcript interrrupted by extra sequences including two homologous sequences marked as "U". The nearer replication fork to *LEU2* is positioned in either collinear or convergent orientation relative to each other. Collisions between replication and transcription induce DSBs and homologous recombination, leading to the formation of a functional *LEU2*. RLFSs are inserted in the intervening sequences in *LEU2* in the "R-loop forming" or "non R-loop forming" orientation. (**B&C)** Recombination frequencies (RFs) of constructs containing RLFS in WT with transcription induction (galactose). (**D&E)** RFs of constructs increases in the *rnh1* mutant (lacking RNase H1, the enzyme responsible for degrading the RNA in R-loops). Statistically significant diffrences between control and RLFS ($p<0.02$) are shown by a black asterisk while those between 'R-loop forming' and 'non-R-loop forming' are shown by a purple asterisk ($p<0.05$).

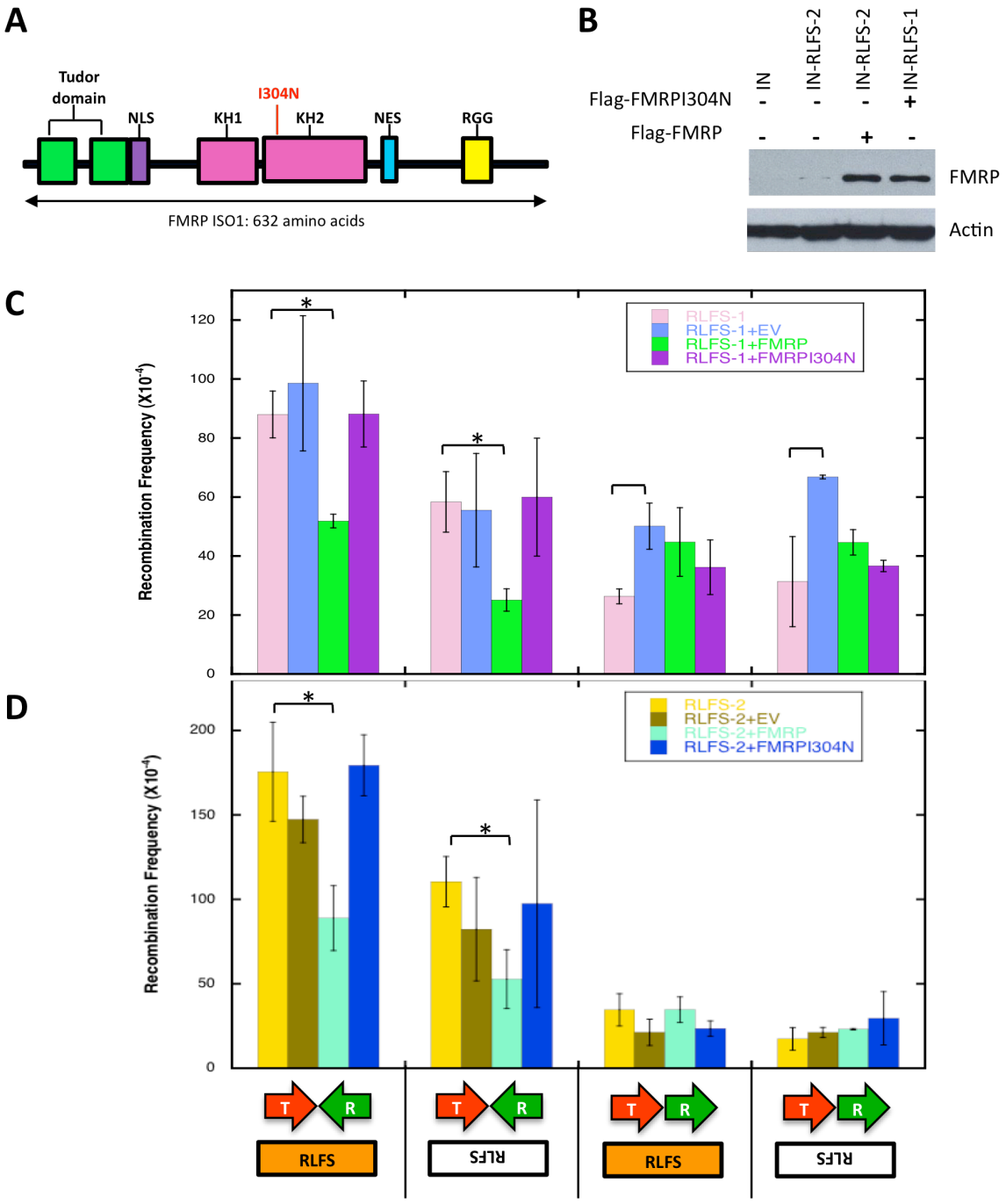

Figure 3. **Expression of FMRP reduces DSBs due to R-loop formation**. **A)** Domain structure of FMRP isoform 1 indicating the naturally occurring mutant FMRPI304N. **B)** Western blot showing the expression of either FMRP or FMRPI304N when co-transformed with IN-constructs. **C&D)** RF of constructs containing RLFS-1 and 2 when co-transformed with either empty vector(EV), FMRP or FMRPI304N. Only FMRP expression and not the EV or mutant reduces recombination in the convergent orientation. Statistically significant changes ($p<0.02$, $n=3$) is shown by an asterisk.

3) We have not yet analyzed gene expression of putative FXS-associated genes. Once the Break-seq analysis of all experiments is completed we will short list genes with overlapping DSBs, RLFSs and FMRP-binding sites as potential FXS-associated genes. We will then examine their potentially differential gene expression or protein levels in Fragile X vs. normal cell lines.

- **What opportunities for training and professional development has the project provided?**

  This project involves two personnel: a research technician (50% effort) and a graduate student (100%). *The research technician has become an expert in Break-seq library construction and is now the go-to person for Break-seq analysis for all lab staff and collaborators. The student has assisted me in writing a book chapter titled "Fragility Extraordinaire: Unsolved mysteries of chromosome fragile sites" in a special edition, "DNA replication: from old principle to new discovery", of the journal Advances in Experimental Medicine and Biology. This chapter is currently being reviewed. The student and I are also preparing a manuscript for publication.*

- **How were the results disseminated to communities of interest?**

  I have presented results at three international conferences (detailed in section 6 "Products") in the form of both oral and poster presentations.

- **What do you plan to do during the next reporting period to accomplish the goals?**

  N/A.

4. IMPACT

- **What was the impact on the development of the principal discipline(s) of the project?**

  We rationalize that there are two principal disciplines of the project: genomic analysis of chromosome breakage by Break-seq and FXS disease etiology. This project represents the first application of the Break-seq methodology in the human genome. As mentioned before, our experimental design is the most comprehensive compared to studies using either conventional cytological methods (Mrasek et al 2012) or other Next-gen sequencing-based DSB mapping methods (Crosetto et al 2013; Wilson et al 2015). Specifically we tested multiple concentrations of aphidicolin as well as compared cells treated with and without colcemid. These conditions are essential to our understanding of the mechanisms of chromosome fragile site formation. Moreover, our study is the first to identify fragile sites genome-wide in the Fragile X cells. The results from our study promise to shed new light on understanding the underlying cause for FXS.

  During the second funding period we have focused on a direct test of the functionality of FMRP in R-loop mediated DNA breakage in a yeast-based system. The results showed that FMRP expression is able to reduce R-loop-induced DSB frequency and, hence, recombination frequency. Moreover, this DSB-repressive function of the FMRP is dependent on its ability to interact with RNA substrate, and presumably also the R-loop substrate. These results are consistent with our hypothesis that FMRP prevents the formation and/or promotes the resolution of R-loop forming sequences. Our study is the first demonstration of this novel function of FMRP and has direct implication in the underlying cause of Fragile X syndrome. Future research to dissect the function of FMRP in genome maintenance will reveal novel targets for Fragile X therapeutics.

- **What was the impact on other disciplines?**

  Nothing to report.

- **What was the impact on technology transfer?**

Nothing to report.

- **What was the impact on society beyond science and technology?**
  We believe that the results generated from our study will contribute to our understanding of FXS and other neurological disorders, and ultimately medical interventions to combat these diseases.

5. CHANGES/PROBLEMS

- **Changes in approach and reasons for change**

  Nothing to report.
- **Actual or anticipated problems or delays and actions or plans to resolve them**

  Nothing to report.
- **Changes that had a significant impact on expenditures**

  Nothing to report.
- **Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents**

  Nothing to report.
- **Significant changes in use or care of human subjects**

  Not applicable.
- **Significant changes in use or care of vertebrate animals**

  Not applicable.
- **Significant changes in use of biohazards and/or select agents**

  Nothing to report.

6. PRODUCTS

- **Publications, conference papers, and presentations**
  - Journal publication

    Nothing to report, manuscript in preparation.
  - Books or other non-periodical, one-time publications

    Feng W. and Chakraborty A. Fragility Extraordinaire: Unsolved mysteries of chromosome fragile sites. ***Advs Exp. Medicine, Biology - DNA replication: from old principle to new discovery.*** Springer International Publishing Switzerland. Under review.
  - Other publications, conference papers, and presentations

    June 2016, Cold Spring Harbor Asia Conference on DNA damage, Metabolism and Diseases, Suzhou, China (Poster presentation)

    Title: Global chromosome fragile site mapping by Break-seq discovers novel function of the Fragile X mental retardation protein

    February 2016, NGS Data Analysis & Informatics Conference, San Diego, California (Poster presentation)

    Title: Global detection of chromosome breakage sites by Break-seq reveals novel functions of the fragile X mental retardation protein

    September 2015, Cold Spring Harbor Conference on Eukaryotic DNA Replication and Genome Maintenance, Cold Spring Harbor, New York (Platform presentation)

    Title: Novel Function of the Fragile X Mental Retardation Protein in Genome Stability
- **Website(s) or other Internet site(s)**

  Nothing to report.
- **Technologies or techniques**

  Nothing to report.
- **Inventions, patent applications, and/or licenses**

  Nothing to report.
- **Other Products**

  Nothing to report.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

- **What individuals have worked on the project?**

Name	Arijita Chakraborty	Andrew McCulley
Project Role	Graduate Student	Research Technician
Researcher Identifier	N.A.	N.A.
Nearest person month worked	10	4
Contribution to Project	Ms. Chakraborty performed cell culture preparations for Break-seq analysis. She also performed the ChIP-seq experiment.	Mr. McCulley assisted with Break-seq library preparations.
Funding Support	Institutional funds at Upstate Medical University	New York State Department of Health

- **Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?**

  The PI (W. Feng) has been rewarded an R01 grant from the National Institute of Health General Medical Sciences, which has initiated in September 2016.

- **What other organizations were involved as partners?**

  Nothing to report.

8. SPECIAL REPORTING REQUIREMENTS

- **Collaborative awards**

  Not applicable.

- **Quad charts**

  Nothing to report.

9. APPENDICES

Not applicable.